



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2024

A new companion to the Romance languages

Glessgen, Martin

DOI: <https://doi.org/10.1515/9783111329338>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-265647>

Monograph

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Glessgen, Martin (2024). A new companion to the Romance languages. Berlin: DeGruyter.

DOI: <https://doi.org/10.1515/9783111329338>

Martin Glessgen

A new companion to the Romance languages

Martin Glessgen

A new companion to the Romance languages

DE GRUYTER

The open access publication of this book has been published with the support of the Swiss National Science Foundation.



ISBN 978-3-11-132924-6

e-ISBN (PDF) 978-3-11-132933-8

e-ISBN (EPUB) 978-3-11-132938-3

DOI <https://doi.org/10.1515/9783111329338>



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For details go to <https://creativecommons.org/licenses/by-nc-nd/4.0/>. Creative Commons license terms for re-use do not apply to any content (such as graphs, figures, photos, excerpts, etc.) not original to the Open Access publication and further permission may be required from the rights holder. The obligation to research and clear permission lies solely with the party re-using the material.

Library of Congress Control Number: 2024941227

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at <http://dnb.dnb.de>.

© 2024 with the author(s), published by Walter de Gruyter GmbH, Berlin/Boston.

This book is published with open access at www.degruyter.com.

Cover image: [Detail] Anna Pismenskova / iStock / Getty Images Plus

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Uxori carissimæ, filiæ mirabili

Table of Contents

Foreword — XVII

Acknowledgements — XIX

Figures — XXI

Symbols, abbreviations and punctuation — XXIII

Part 1: The study of Romance linguistics

1	The key characteristics of language and linguistics — 3
1.1	Romance linguistics and the structure of the <i>Companion</i> — 3
1.2	The concept of ‘language’ — 5
1.2.1	Three functions of language — 5
1.2.2	Language as a sign system — 8
1.2.3	<i>Langue</i> and <i>parole</i> — 10
1.2.4	The three organising principles of language: linearity, intelligibility and economy — 12
1.2.5	The variable and evolutionary character of language — 15
1.2.6	‘External’ factors relating to language use and evolution — 19
1.2.7	Speech and writing — 21
1.2.8	Summary — 21
1.3	Different ways of interpreting language — 22
2	Emergence and development of Romance linguistics — 24
2.1	The foundations of (Romance) linguistics before 1800 — 24
2.1.1	Romance linguistics within the general framework of linguistics — 24
2.1.2	The birth of linguistic studies and the rise of the professional ‘linguist’ — 25
2.2	The historical-comparative paradigm in Romance studies — 28
2.2.1	The foundations — 28
2.2.2	The preparatory phase (ca. 1800–1850) — 29
2.2.3	The formative phase (ca. 1850–1900) — 32
2.2.4	The fully-fledged historical-comparative paradigm (1900 to 1930/1950) — 34
2.3	The development of the present-day paradigms — 40
2.3.1	New methodologies in the Romance domain: sociolinguistics, pragmatics and applied linguistics — 40
2.3.2	Generative linguistics and the ‘cognitive turn’: creole studies, neuro- and psycholinguistics — 41
2.3.3	<i>Ex traditione innovatio</i> : typology and diachrony — 44

2.3.4	The current ‘centrifugal’ paradigm —	45
2.4	Reference works for the study of Romance linguistics —	46
2.4.1	Manuals and encyclopedic reference works —	46
2.4.2	Shorter manuals and textbooks —	48
2.4.3	Ongoing bibliographies —	50

Part 2: Romance languages and varieties today

3	Presentation of the Romance languages —	55
3.1	Terminology: languages, dialects and varieties in the <i>Romània</i> —	55
3.2	Romance languages in Europe and across the world —	58
3.3	Geolinguistic boundaries: Gallo- and Italo-Romance dialects —	62
3.4	External characteristics of the Romance languages —	67
3.4.1	French and the <i>langue d’oil</i> dialects —	68
3.4.2	Occitan or the <i>langue d’oc</i> dialects —	70
3.4.3	Gascon —	72
3.4.4	Francoprovençal —	73
3.4.5	Italian —	74
3.4.6	Sardinian —	77
3.4.7	Romansh —	78
3.4.8	Ladin and Friulian —	79
3.4.9	Spanish —	81
3.4.10	Catalan and Valencian —	84
3.4.11	Galician —	85
3.4.12	Portuguese —	86
3.4.13	Romanian —	87
3.5	Aspects of intra-Romance typology —	88
3.5.1	‘Languages by distance’ and ‘languages by development’ in the <i>Romània</i> —	88
3.5.2	Illustrating variation among Romance languages —	93
3.5.3	The classification of the Romance languages —	95
3.6	Language contact in the <i>Romània</i> —	101
3.6.1	Diglossia, bilingualism and language contact —	101
3.6.2	The importance of language contact in the <i>Romània continua</i> —	103
3.6.3	Romance-based creoles —	105
4	Variational linguistics and the Romance languages —	107
4.1	The impact of variation —	107
4.1.1	Linguistic variation – an empirical constant —	107
4.1.2	The limits of variation: intelligibility, grammaticality and acceptability —	109
4.1.3	The variational approach to language —	110
4.2	Geographical or diatopic variation —	112

4.2.1	Primary dialects — 112
4.2.1 no. 1	Status and characteristics of primary dialects — 112
4.2.1 no. 2	The language atlases of the <i>România</i> — 114
4.2.1 no. 3	The contribution of dialectometry: the example of <i>Gallo-Romance</i> — 118
4.2.1 no. 4	Dialectological descriptions and studies — 120
4.2.2	Language evolution and geographical space — 121
4.2.3	Secondary dialects — 124
4.2.4	Regiolects or tertiary regional variation — 126
4.3	Diastratic and diaphasic variation and their theorisation — 128
4.3.1	Foundations of variational theory — 128
4.3.2	Interdependence among the varieties of a language — 129
4.3.3	Defining diastratic and diaphasic variation — 131
4.3.4	Linguistic prestige — 132
4.3.5	Diastratic and diaphasic variation in the <i>România</i> — 135
4.4	Linguistic proximity and linguistic distance, speech and writing — 137
4.4.1	Oral sources in linguistic analysis — 137
4.4.2	The theorisation of linguistic immediacy and distance — 139
4.4.3	The particular characteristics of linguistic immediacy — 140
4.4.4	Immediacy and the diasystem — 142
4.5	Discourse traditions and pragmatics — 144
4.5.1	Textual genres, discourse traditions and the diasystem — 144
4.5.2	Issues in text linguistics — 147
4.5.3	Principles of pragmatics — 149
4.5.3 no. 1	Speech acts, the cooperative principle and implicature — 150
4.5.3 no. 2	Pragmatics, the diasystem and textual genres — 151
4.5.4	The significance of textual genres for linguistic analysis — 153
4.6	Open questions relating to variational theory — 156
4.7	Variational linguistics in the chapters that follow — 160

Part 3: The internal structures of the Romance languages and their history

5	Domains of language and historical periodisation — 163
5.1	The domains of language — 163
5.1.1	Traditional subdivisions — 163
5.1.2	Contributions from neuro-psychology — 166
5.1.3	The domains of language and grammaticalisation — 168
5.2	Periodisation of the Romance languages — 170
5.2.1	Periods established for external and internal linguistic analysis — 170
5.2.2	The concept of Proto-Romance — 173
5.3	Remarks on the following chapters — 174

6	Phonetics, phonology and graphemics — 175
6.1	Sounds, phonemes and sound change — 175
6.1.1	Sounds, phonemes and allophones — 175
6.1.2	Phonetics, phonology and sound change — 178
6.2	Sounds and phonemes in the Romance languages — 180
6.2.1	Place of articulation — 180
6.2.2	Manner of articulation — 183
6.2.3	Phonological systems in the <i>România</i> — 188
6.2.3 no. 1	French (of France) — 189
6.2.3 no. 2	Italian — 190
6.2.3 no. 3	Spanish (of Spain) — 190
6.2.3 no. 4	Portuguese (of Portugal) — 191
6.2.4	Phonological variation and change — 192
6.3	Syllables and prosody — 195
6.4	Writing, graphemics and orthography — 199
6.4.1	Alphabetic writing — 199
6.4.2	The relationship between writing and pronunciation — 201
6.4.3	Orthography and society — 203
6.4.4	The importance of graphemes for phonetic history — 204
6.5	Major trends of phonetic evolution in the <i>România</i> — 205
6.5.1	The Late Latin vowel systems and their evolution — 205
6.5.2	Stressed vowels in the Romance languages: diphthongisation, monophthongisation, ‘extreme’ vowels — 208
6.5.3	Unstressed vowels in the Romance languages — 212
6.5.4	Consonants in the Romance languages: intervocalic plosives, palatalisation, word-final consonants, consonant clusters — 214
6.5.5	Patterns of consonantal evolution from Latin to French — 219
6.5.6	General observations — 220
6.6	Latinisation and re-Latinisation — 221
6.7	General tendencies of sound change — 223
7	Inflectional morphology — 224
7.1	The subject matter of morphology — 224
7.1.1	Inflection vs. word formation — 224
7.1.2	Lexical and grammatical elements in morphology — 225
7.1.2 no. 1	Lexical vs. grammatical word classes — 225
7.1.2 no. 2	Stem vs. grammatical affix — 226
7.1.2 no. 3	Free vs. bound (lexemes and grammemes) — 227
7.2	Inflection in Latin and in Romance — 228
7.2.1	‘Morphosyntactic’ features — 228
7.2.2	Inflectional classes — 229
7.2.3	Inflection and agglutination from Latin to Romance — 230

7.2.3 no. 1	Allomorphy, fusion and suppletion —	230
7.2.3 no. 2	Cumulativeness —	232
7.2.3 no. 3	Syncretism —	232
7.3	The restructuring of the Latin nominal inflection system —	234
7.3.1	Inflectional classes —	235
7.3.2	Number —	236
7.3.3	Gender —	237
7.3.4	The reduction of the Latin case system —	238
7.4	Elements of adjectival inflection —	241
7.4.1	Comparison —	241
7.4.2	Adverb formation —	243
7.5	The pronoun system and the determiners ('D-system') —	243
7.5.1	Form and function: full pronouns, clitics and determiners —	243
7.5.2	Personal pronouns and determiners —	245
7.5.3	Personal pronouns and verbal politeness —	246
7.5.4	Possessive pronouns and determiners —	247
7.5.5	Demonstrative pronouns and determiners —	247
7.5.6	Relative pronouns —	248
7.5.7	Indefinite pronouns —	249
7.6	The restructuring of Latin verb paradigms —	249
7.6.1	Present indicative —	251
7.6.2	Morphemes —	252
7.6.3	Past imperfect indicative —	253
7.6.4	Future —	254

8 Syntax — 257

8.1	The study of syntax —	257
8.1.1	Major approaches to syntax —	257
8.1.2	Observations relating to sentence structure —	258
8.2	The noun phrase (NP) in the Romance languages —	260
8.2.1	Semantic and syntactic roles —	260
8.2.2	The noun phrase (NP) and its components —	261
8.2.3	Determination —	261
8.2.4	Modification —	263
8.3	The verb phrase (VP) in the Romance languages —	266
8.3.1	Components and structure of the VP —	266
8.3.2	Grammatical categories of the predicate: TAM —	267
8.3.2 no. 1	Tense —	267
8.3.2 no. 2	Aspect —	268
8.3.2 no. 3	Mood and modality —	269
8.3.2 no. 4	Verbal periphrases —	270
8.3.2 no. 5	Voice and diathesis —	272

8.3.2 no. 6	Lexical aspects of the VP: <i>Aktionsart</i> , verb valency, transitivity and semantic groups of verbs — 274
8.3.2 no. 7	Optional determiners: negation and adverbs — 279
8.3.3	The transformation of the verbal system from Latin to Romance — 280
8.4	The sentence in the Romance languages — 285
8.4.1	Denotative content and principles of linear marking — 285
8.4.2	Constituent order and principles of position — 288
8.4.3	Arguments and semantic roles — 291
8.4.3 no. 1	Overview — 291
8.4.3 no. 2	Semantic roles — 292
8.4.3 no. 3	Argument marking in Romance – the subject — 295
8.4.3 no. 4	Romance argument markers – 1 st object (O1) and 2 nd object (O2) — 296
8.4.3 no. 5	Arguments and circumstantial adjuncts — 299
8.4.3 no. 6	Latin, Romance and creole systems — 299
8.4.4	Information structure (topic-comment) — 300
8.4.5	Subordination — 302
8.5	Perspectives: the grammatical description of Romance — 305
8.5.1	Grammatical description of the Romance languages: observations — 305
8.5.2	General conclusions regarding the transformations in morphology and syntax between Latin and Romance — 308
9	Lexis — 310
9.1	The lexeme: general properties — 310
9.1.1	Basic properties of the lexeme — 310
9.1.2	Lexemes and proper nouns — 312
9.1.3	Number of words and frequency of use — 313
9.1.4	Variational connotation and communicative context — 315
9.2	Semantic and semiotic theory — 317
9.2.1	Lexical meaning, concept and referent — 317
9.2.2	Sign theory and the structure of the lexeme — 318
9.2.2 no. 1	The dichotomy of the sign — 319
9.2.2 no. 2	The semiotic triangle — 319
9.2.2 no. 3	The semiotic trapezium — 320
9.2.2 no. 4	The semiotic pentagon — 320
9.2.2 no. 5	The semiotic square — 321
9.2.2 no. 6	The enhanced semiotic pentagon — 322
9.2.2 no. 7	Signifieds, concepts and referents — 324
9.2.3	Onomasiology, structural semantics and cognitive semantics — 327
9.3	Semantic change and semantic relationships — 330
9.3.1	Types of semantic change and semantic relationships — 330
9.3.2	Taxonomic change — 332
9.3.2 no. 1	Generalisation (<i>species</i> → <i>genus</i>) — 332

9.3.2 no. 2	Specialisation (<i>genus</i> → <i>species</i>) —	333
9.3.2 no. 3	Co-hyponymy —	334
9.3.3	Metonymy —	334
9.3.4	Metaphor —	337
9.3.5	Changes by formal motivation —	339
9.3.6	General observations —	340
9.3.7	Historical onomasiology —	342
9.4	Word formation in the Romance languages —	347
9.4.1	Functions and general characteristics —	347
9.4.2	Derivation in the history of the Romance languages —	349
9.4.2 no. 1	Prefixation —	350
9.4.2 no. 2	Suffixation —	350
9.4.2 no. 3	Conversion and improper derivation —	354
9.4.2 no. 4	Change of gender —	355
9.4.2 no. 5	Parasynthesis —	356
9.4.3	Compounding in the Romance languages —	356
9.4.4	Other derivational mechanisms —	359
9.4.4 no. 1	Back-formation or regressive derivation —	359
9.4.4 no. 2	Affixoids —	360
9.4.4 no. 3	Reduplication —	360
9.4.4 no. 4	Derivatives without a base —	360
9.4.4 no. 5	Derivation within compounds —	361
9.4.5	Other mechanisms of word formation —	361
9.4.5 no. 1	Forms of word shortening: clipping, acronyms (and initialisms) —	361
9.4.5 no. 2	Delocutives —	361
9.4.5 no. 3	Blending —	362
9.4.5 no. 4	Intentional formal distortion —	362
9.4.6	The significance of diachrony in word formation —	363
9.5	Syntagmatic context: valency, collocation and phraseology —	366
9.5.1	Valency —	366
9.5.2	Collocations —	367
9.5.3	Phraseologisms and idioms —	367
9.5.4	Observations —	368
9.6	Lexical borrowing in the Romance languages —	369
9.6.1	Definition —	369
9.6.2	Contact languages and borrowing in the Romance languages —	370
9.6.2 no. 1	Languages in contact with Latin —	370
9.6.2 no. 2	Germanic languages —	371
9.6.2 no. 3	Arabic —	372
9.6.2 no. 4	Slavonic languages —	372
9.6.2 no. 5	English and other languages —	372
9.6.2 no. 6	Contact among different Romance languages —	373

9.6.3	Typology of lexical borrowing —	374
9.6.4	Quantitative importance of borrowing in the Romance languages —	377
9.7	Onomastics and deonomastics —	379
9.7.1	Proper nouns and common nominal lexemes —	379
9.7.2	The formation of personal names in the Romance languages —	381
9.7.3	The formation of place names in the <i>Romània</i> —	383
9.8	Historical lexicology and etymology in Romance studies —	385
9.9	Research methods: historical and etymological lexicography —	389
9.9.1	Historical and etymological dictionaries of French —	389
9.9.2	Historical and etymological dictionaries of the other Romance languages: Romance languages in general, Occitan, Italian, Spanish, Catalan, Portuguese, Romanian, Sardinian, Francoprovençal, Romansh, Ladin, Friulian —	391
9.9.3	Final remarks —	394

Part 4: **External history of the Romance languages and varieties**

10	External history of the Romance languages and varieties —	397
10.1	Establishing a framework for the study of external history —	397
10.1.1	The relationship between external and internal history —	397
10.1.2	The preoccupations of external history —	398
10.1.3	Establishing criteria for the study of external history —	400
10.1.3 no. 1	Geographical area and demographics —	401
10.1.3 no. 2	Language contact —	402
10.1.3 no. 3	Infrastructure and socio-cultural, political and economic organisation —	403
10.1.3 no. 4	Written culture, linguistic thought and the diasystem —	404
10.2	The Roman period (~ 7 th century BC – 5 th century AD) —	405
10.2.1	The shaping of the Roman Empire and Latinisation —	405
10.2.2	Language contact and the fragmentation of the <i>Romània</i> —	408
10.2.3	Variation and regionalisation of Latin —	411
10.3	The emergence of the <i>Romània</i> (5 th –10 th centuries) —	412
10.3.1	From Latin to Romance —	412
10.3.2	Geography: delimitation of the future Romance territories —	415
10.3.3	Migration and language contact: the Germanic and Arabic <i>superstrata</i> —	419
10.3.4	Political factors and infrastructure —	421
10.3.5	Written culture and linguistic thought —	423
10.3.5 no. 1	Written Romance —	423
10.3.5 no. 2	Pre-textual Romance elements in Latin texts —	426
10.3.5 no. 3	Normative Latin, Rustic Latin and Vernacular Romance —	427
10.3.5 no. 4	Metalinguistic evidence —	429
10.3.6	Conclusion: the transition from Latin to the Romance languages —	430

10.4	The Romance languages during the Low Middle Ages (11 th – 15 th centuries) — 431
10.4.1	Geographical area and migrations — 431
10.4.2	Infrastructure, socio-cultural development and hierarchical institutions — 436
10.4.3	Written culture and linguistic thought — 439
10.4.3 no. 1	Chronology of written production in Romance — 439
10.4.3 no. 2	Latin and the Romance languages through the centuries — 440
10.4.3 no. 3	The <i>scriptae</i> — 443
10.4.3 no. 4	Linguistic culture — 445
10.4.4	Minority languages and varieties — 447
10.4.5	The emergence of the national Romance languages — 448
10.5	The modern period (1500 – end of the 19 th century) — 451
10.5.1	Geography, geopolitics and demographics — 451
10.5.2	Infrastructure and socio-cultural evolution — 456
10.5.3	Written culture and language elaboration — 458
10.5.3 no. 1	The establishment of new ‘umbrella languages’ — 458
10.5.3 no. 2	Written culture — 459
10.5.3 no. 3	Linguistic thought: theory, grammar and lexicography — 461
10.5.3 no. 4	Language planning and language academies — 466
10.5.3 no. 5	Elite and population — 468
10.5.3 no. 6	Acrolectal spoken varieties and writing — 469
10.5.3 no. 7	The standard Romance languages — 470
10.5.4	Non-standard varieties — 471
10.6	The present-day period (1880 until today) — 473
10.6.1	Geographical space, geopolitical factors and varieties — 473
10.6.2	Demographic, infrastructural and socio-cultural factors — 477
10.6.3	Written culture and linguistics — 478
10.6.4	Evolutionary trends and present-day non-standard varieties — 481
10.7	Contributions of external history — 482

Part 5: Practical and ideological aspects of research and teaching in Romance linguistics

11	Philological approaches and corpus linguistics — 487
11.1	Inventories of medieval Romance textual genres — 487
11.1.1	Reference catalogues offering a systematic description of medieval texts — 488
11.1.2	Secular literature — 489
11.1.3	Religious literature — 490
11.1.4	Texts reflecting specialised knowledge — 491
11.1.5	Documentary writing — 494

11.2	‘Material encoding’ and manuscript culture —	496
11.2.1	Material encoding of texts in the history of the <i>Romània</i> —	496
11.2.2	Manuscript culture —	498
11.3	The problem of copies and textual adaptation (‘mouvance’) —	503
11.3.1	Copying and adaptation —	503
11.3.2	Textual ‘mouvance’ —	505
11.3.3	Aspects of textual transmission —	506
11.4	The theory and practice of editing —	507
11.5	Editing texts for linguistic analysis —	511
11.5.1	Methods of textual edition —	511
11.5.2	The contribution of textual genres to linguistic analysis —	513
11.5.3	The role of micro- and macroscopic description in historical linguistics —	514
11.6	Corpus linguistics and philology in Romance studies —	515
11.6.1	The general interest of corpus linguistics —	515
11.6.2	The processing of medieval textual data —	517
11.6.3	Programming tools and tools for linguistic analysis —	520
11.6.4	Principal databases —	521
11.6.5	Perspectives —	523

12 From theory to practical work — 524

12.1	Romance studies in theory, practice and teaching —	525
12.1.1	The development of linguistic theory —	525
12.1.2	Romance studies in academia (1): non-Romance-speaking countries —	526
12.1.3	Romance studies in academia (2): Romance-speaking countries —	528
12.1.4	Factors contributing to cohesion in Romance studies —	530
12.2	Linguistics, society and politics —	531
12.2.1	The intrinsic dangers of linguistics —	531
12.2.2	The influence of external factors on linguistics —	532
12.2.3	Ideological dependence and scientific (in)dependence —	533
12.2.4	Towards a new view of the discipline —	533

Epilogue: On the usefulness of (Romance) linguistics — 535

1	The place of Romance linguistics among the language sciences —	535
2	Linguistics and society —	535

References — 539

1	Acronyms and abbreviations —	539
2	General reference works, monographs and articles —	544

Subject index — 559

Name index — 591

Works cited — 599

Foreword

Habent sua fata libelli

Romance linguistics constitutes a field of research in its own right within the vast domain of the study of languages and cultures. Its boundaries are clearly defined and the potential it offers for both empirical analysis and methodological innovation is unique.

Within the *Romània*, a number of specific factors combine to provide ideal circumstances for the exemplary investigation of almost all questions relating to language variation and change, as well as the functioning of language itself. In the first place, its long tradition, which begins with Latin, the predecessor of the Romance languages, spans two and a half millennia; secondly, it embraces numerous dialects and individual languages; and thirdly, the use of Romance languages throughout Europe, America and Africa offers rich opportunities for cultural studies. Furthermore, the facilities and resources available to researchers in the field are of exceptional quality.

Towards the end of the 20th century, the rapidly expanding volume and the complexity both of the material available for study and of the knowledge acquired put the field at risk of fragmentation. Acknowledgement of this undesirable state of affairs has since led to the development of a number of major high-quality encyclopedias with the aim of providing an overview of the subject's many different facets. These are the *Lexikon der Romanistischen Linguistik* (consisting of 11 volumes), *Romanische Sprachgeschichte* (History of the Romance languages, in 3 volumes), the *Cambridge History of Romance Languages*, the *Oxford Guide to the Romance Languages*, the 60-volume series *Manuals of Romance Linguistics*, and, most recently, the comprehensive *Oxford Encyclopaedia of Romance Languages*, published online. Finding one's way around this multitude of extensive, composite and differently conceived works, however, which amount to a total of approximately 50,000 pages, can prove something of a challenge. For this reason, a detailed presentation of each one is provided in chapter 2.4.1.

The need for a shared canonical knowledge base underpinning the field of Romance linguistics, in order to counter the ever-increasing centrifugal tendencies currently pervading its development, if not that of linguistics in general, has long been a particular preoccupation of mine. In 2000, I analysed the 50 most authoritative manuals of Romance Linguistics written during the 19th and 20th centuries. The results of this investigation led me to undertake the development and publication of the encyclopedia *Romanische Sprachgeschichte* (2003-2008) with Gerhard Ernst, Wolfgang Schweickard and the late Christian Schmitt. Using this work as a foundation, I then wrote a shorter introduction to Romance linguistics for students (*Linguistique Romane*, published in 2007), which was replaced by a second version in 2012, with the same title but thor-

oughly revised content. Since 2004, my role as secretary of the *Société de Linguistique Romane* (Society for Romance Linguistics) has provided me with further opportunities for contributing to the cohesion of the discipline, chiefly through the organisation of the Society's triennial Congresses, but also as editor of the *Revue de linguistique romane* and its accompanying publication series, the latter now comprising one hundred volumes.

The present *Companion* thus represents the result of a quarter of a century of practice in the discipline. It attempts to capitalise on this experience by offering a complete and balanced overview of the core subjects which make up the field of Romance linguistics, placing particular emphasis on methodology. The purpose of the book is twofold. Firstly, it is conceived as a basic canon which provides a solid foundation in the discipline. Secondly, its systematic references to a broad spectrum of the latest encyclopedic literature, as well as to relevant studies on specific topics, provide the reader with easy and convenient access to highly specialised areas of research. The *Companion* is therefore intended both to be read from cover to cover as a stand-alone work, and to be used as necessary for its specific sections on individual subjects. Advanced students, early-career researchers, lecturers, specialists of other languages, socio-linguists, philologists and historians alike will all benefit from this accessible and up-to-date reference work, as it enables the reader to contextualise any detailed knowledge they may possess in the different areas.

Zurich, June 2024

Martin Glessgen

Acknowledgements

The various stages of development of this *Companion* are inextricable from the many and intensive contacts and friendships formed over recent decades. Deep sadness over the loss of cherished mentors and collaborators – first and foremost Max Pfister and David Trotter – is counterbalanced by the satisfaction of adding enthusiastic and competent new recruits to the field. Countless colleagues and students have contributed to the *Companion* through discussions, critical reading and suggestions.

I would like to express my gratitude to those who were instrumental in the book's development. The English of the first draft was the work of Laura Endress, revised by David Trotter and Myriam Bergeron. The English of the subsequent drafts resulting from the addition of new material, as well as of the final version, was revised by Marguerite Dallas with the help of John Barnes. Brenda Laca's revision of the chapter on syntax and the correction of the final version with accompanying remarks by Dominique Gerner, Adam Ledgeway and Martin Maiden were particularly valuable. From a stylistic point of view, the composite nature of the English used throughout this *Companion* is a faithful reflection of the different phases of its development and of the contributions of native and non native speakers of both American and British varieties. A systematic attempt to homogenise it would have erased all traces of the friends and colleagues with whom I have enjoyed mutually enriching relationships.

Finally, the formatting of the text is mainly the work of Dumitru Kihai and the preparation of the indexes that of Jessica Meierhofer. Our thanks go equally to Ulrike Krauss and Christine Henschel for their exemplary editorial advice.

The *Companion* is dedicated to my dear wife and our daughter, who has grown up along with this book, and whose zest for life fills us with fresh hope every day.

Figures

Fig. 1	Branches of the Indo-European language family — 2
Fig. 2	Bühler's Organon model — 6
Fig. 3	The dichotomy of the linguistic sign — 9
Fig. 4	Map of the <i>Romània continua</i> — 59
Fig. 5	Examples of isoglosses marking the linguistic dividing lines <i>La Spezia-Rimini</i> and <i>Roma-Ancona</i> — 65
Fig. 6	Geolinguistic structure of Italy according to the AIS — 76
Fig. 7	Geolinguistic transformations in the Iberian Peninsula after the <i>Reconquista</i> — 82
Fig. 8	Example of an article in Meyer-Lübke's REW (art. 3445) — 97
Fig. 9	Plural formation in the <i>Romània</i> — 98
Fig. 10	Terms for 'MARE' in the <i>Romània</i> — 99
Fig. 11	The formation of an interlanguage — 105
Fig. 12	Digitised map of the ALF for the verb 'TO SAW' — 115
Fig. 13	Dialectometrical map of linguistic zones of Gallo-Romance — 118
Fig. 14	Dialectometrical map of the <i>oil</i> area — 119
Fig. 15	Simplified map of the denominations of 'BEE' — 122
Fig. 16	Chain of variation — 130
Fig. 17	Terminology used to describe diastratic varieties — 133
Fig. 18	The continuum of linguistic distance and immediacy — 140
Fig. 19	Approximate classification of utterances according to textual genre — 146
Fig. 20	Interdependence between the four parameters of language production — 153
Fig. 21	Positioning of utterances in the diasystem — 158
Fig. 22	The model of Hickok/Poepfel — 167
Fig. 23	Articulatory organs and places of articulation — 181
Fig. 24	The 'vowel trapezium' — 188
Fig. 25	The phonotactic structure of It. <i>porta</i> — 195
Fig. 26	Visualisation of acoustic waves (oscillogram and prosodic curves) — 197
Fig. 27	The semiotic triangle — 319
Fig. 28	The semiotic pentagon — 321
Fig. 29	The semiotic square — 322
Fig. 30	The enhanced semiotic pentagon — 323
Fig. 31	The replacement of Latin by the Romance languages — 441
Fig. 32	The <i>Sequence of Saint Eulalia</i> — 501
Fig. 33	<i>Stemma</i> of the Latin manuscripts of <i>Moamin</i> — 509
Fig. 34	Facsimile of a charter from Lorraine, dated 1239 — 518

Symbols, abbreviations and punctuation

1 Symbols and punctuation

For the symbols used in the chapters on phonetics and phonology, the alphabet of the International Phonetic Association (IPA) should be referred to (cf. chapter 6.2).

□	phone
//	phoneme
{ }	morpheme
< >	grapheme
ˈ	accent falling on the following syllable
Ø	zero marking, disappearance of a sound or phoneme, absence
> or <	shows that a lexical form or a phonetic segment has evolved into a different form or segment, generally by way of phonetic evolution, e.g. MATRE > <i>mère</i>
→	suggestions for further reading
→ or ←	shows that a lexical form or a meaning has undergone a change by means of a process of derivation or semantic change, e.g. APE → <i>abeille</i> (from Lat. APICULA) or CAPUT “head” → “boss, director”
*	etymology: marks a form or meaning that is not attested in Latin (= Proto-Romance); grammar: marks a construction as ungrammatical
CAPITALS	Latin etymon
‘HOUSE’	a concept (as opposed to a definition, which appears between “...”), or a formalised lexical type
‘...’	terminology, central concept, occasional usage, quotation
“...”	definition

For Latin etyma, only long vowels are marked (by a macron); short vowels remain unmarked. Latin accusative forms appear without the word-final inflectional marker -M, as it disappears at a very early stage (cf. 3.5.2 and 6.5.4 no. 3, e.g. MATRE instead of MATREM, CENTU instead of CENTUM); first declension feminine forms in the accusative ending in -A (again, without word-final -M) are identical to the nominative forms (e.g. AMICA).

2 Abbreviations

For the abbreviations in the chapters on morphology and syntax the *Leipzig Glossing Rules* (cf. Comrie / Haspelmath / Bickel, 2008) should be consulted.

A	adjective, argument	Iber.	Ibero-Romance
acc.	accusative	<i>ibid.</i>	<i>ibidem</i>
AD	<i>anno domini</i>	<i>id.</i>	<i>idem</i>
adv.	adverb	impf.	imperfect
AGT	agent	ind.	indicative
al.	<i>alii</i>	Indo-Europ.	Indo-European
Am.	American	inf.	infinitive
Am. Sp.	American Spanish	It.	Italian
Ar.	Arabic	fig.	illustration
art.	article	Lad.	Ladin
Béarn.	Béarnais	Lat.	Latin
C	circumstantial (adjunct), comment (topic-comment), consonant	Logud.	Logudorese
ca.	<i>circa</i>	m. / masc.	masculine
cal.	Calabrese, Calabrian	Med.	Medieval (e.g. Latin)
Cat.	Catalan	M.	Middle (e.g. M.Fr.)
Celt.	Celtic	Mod.	Modern (e.g. Mod.Fr.)
centr.	central (e.g. Italian)	ms(s), MS	manuscript(s)
cf.	<i>confer</i>	N, n.	noun, northern
ch.	chapter	nom.	nominative
Cl.	Classical (Latin)	north. / North.	northern
Dalm.	Dalmatian	Nuor.	Nuorese
dat.	dative	O.	old (e.g. O.Fr.)
decl.	declension	O (1/2)	object (first / second)
dir.	director	Occ.	Occitan
ed.	editor	<i>op. cit.</i>	in the work cited
e.g.	for example	P	person, phrase (e.g. NP 'noun phrase'), predicate
Engad.	Engadine	p. / pp.	page(s)
Engl.	English	PAT	patient
EXP	experiencer	perf.	perfect
f. / fem.	feminine	Pg.	Portuguese
fam.	familiar (e.g. French)	Piedmont.	Piedmontese
fasc.	fascicule	pl.	plural
Fr.	French	pluperf.	pluperfect
Friul.	Friulian	pop.	popular (e.g. French)
Frpr.	Francoprovençal	pres.	present
fut.	future	Prov.	Provençal
Gal.	Galician	Rom.	Romansh
Gasc.	Gascon	Rom.n.	Romanian
Gaul.	Gaulish	S	subject
gen.	genitive	Sard.	Sardinian
Germ.	German, Germanic	sb.	somebody
Gr.	Greek	sg.	singular

Sic.	Sicilian	transl.	translated
South. / south.	southern	Tusc.	Tuscan
Sp.	Spanish	Umbr.	Umbrian
<i>sq(q).</i>	the following	V	verb, vowel
subj.	subjunctive	var.	various
sth.	something	Ven.	Venetan
Sursilv.	Surselvan	Vgl.	Vegliot
s.v.	<i>sub voce</i>	vol./vols.	volume(s)
T .	topic (topic-comment)	vs	versus

Part 1: **The study of Romance linguistics**

balto-slave	balte		ancien prussien			lituanien	
			letton				
	slave	slave de l'ouest			sorabe		
					polonais		
					slovaque		
					tchèque		
		slave du sud			slovène		
					serbo-croate		
					macédonien		
	slave de l'est			bulgare			
ukrainien							
biélorusse							
					russe		
germanique	nordique		ancien norrois	ancien islandais	moyen islandais	islandais	
						féroïen	
				ancien norvégien	moyen norvégien	norvégien	
			ancien suédois		moyen suédois	suédois	
			ancien danois		moyen danois	danois	
			ancien anglais		moyen anglais	anglais	
	westique		ancien frison	frison			
			ancien bas francique	moyen néerlandais		flamand	
						hollandais	
			ancien saxon	moyen bas allemand		afrikaans	
			ancien haut allemand	moyen haut allemand		bas allemand	
						haut allemand	
					yiddish		
ostique			gotique				
celtique	celtique insulaire	goidélique	ancien irlandais		moyen irlandais	gaélique irlandais	
		brittonique	ancien gallois			gaélique écossais	
	celtique continental	gaulois		ancien breton		gallois	
		celtibérien				breton	
		falisque					
italique	latino-falisque	latin	gallégo-portugais		galicien; portugais		
			ancien espagnol, asturien-léonais, aragonais, navarrais		dialectes hispaniques, espagnol		
			ancien catalan		catalan, valencien		
			ancien gascon		gascon		
			ancien occitan		occitan (méridional, septentrional)		
			ancien francoprovençal		francoprovençal		
			ancien français		dialectes d'oïl, français		
			ancien italoroman		italoroman, italien		
			ancien sarde		sarde (logudorais, campidanais)		
			ancien romanche, ladin, frioulan		romanche; ladin; frioulan		
			ancien dalmate		dalmate		
	ancien roumain		roumain; aroumain; méglenoroumain				
	sabellique	osque					
		ombrien					
thrace							
hellénique		grec ancien			grec byzantin	grec moderne	
arménien		arménien classique			arménien		
phrygien							
anatolien	hittite						
	louvite						
indo-iranien	iranien	iranien du sud-ouest	ancien persan		moyen persan		persan moderne
		iranien central				kurde	
						pachtoun	
						ossète	
					parthien		
		bactrien					
	indo-aryen	sanskrit védique		sanskrit classique			ourdou
						penjabi	
						sindhi	
						hindi	
						bengali	
tokharien							

Fig. 1: Branches of the Indo-European language family

1 The key characteristics of language and linguistics

1.1 Romance linguistics and the structure of the *Companion*

The family of Romance languages forms an easily recognisable group among the world's languages, since its representatives emerged from a common, relatively well-known base: spoken Late Latin, which corresponds to the regional varieties of Latin spoken in Late Antiquity (that is, from the 4th–6th centuries). Over the course of their history, spanning one and a half millennia, the Romance languages have evolved through gradual differentiation and divergence. Nevertheless, geographical proximity and a mutual link to written Latin as a cultural and religious language have led to numerous interactions and parallel developments among Romance varieties. Their partially shared history accounts for the high degree of cohesion that can be observed among members of the Romance language family.

The Romance group constitutes a tiny segment within the manifold diversity of the world's languages. It is a subgroup of one of the eleven branches of Indo-European languages – a vast family which also includes the Germanic, Slavonic, Celtic and Indo-Iranian languages, among others (*cf.* fig. 1) – and is spread across all continents today, its origins lying in a common language, Indo-European, which was spoken some 4,000 or 5,000 years ago, most likely in the Pontic area. The Indo-European languages, in turn, are merely one specific family among others, such as the Semitic or Sino-Tibetan languages, amid a plethora of non-classified, sometimes unclassifiable, languages.

Fig. 1 provides an overview of a large number of Indo-European languages, illustrating their genealogical structure and ramifications. Nevertheless, this list of languages is not exhaustive: to the Romance languages alone it would have been possible to add Ladin, Friulian or Valencian, not to mention numerous regional dialects. Moreover, certain aspects are difficult to convey in the form of a diagram. Thus the representation of Latin is slightly inaccurate, since it continued to exist in writing throughout the medieval and modern eras, when it had long disappeared as a spoken language. Furthermore, the evolution of given branches within language families can follow very different paths. Galician and Portuguese, closely related, evolved from a common base language: Medieval Gallego-Portuguese. Gascon and Occitan are also closely related; Gascon, however, diverged from both regional Latin and the other varieties of early Romance independently of and much earlier than Occitan. Southern Occitan and Catalan differed less from one another during the Middle Ages than they do today, and in order to study their evolution, one must also consider Aragonese and Gascon. These points will be discussed in more detail below (*cf.* 3.4.11).

The Romance languages share one characteristic which gives them a particularly interesting status among other well-known language families: they have a common ancestor, Latin, on which a considerable amount of detailed information is available. By comparison, the source languages of almost all today's languages are unknown or very poorly documented. Moreover, written sources from the whole of the second millennium provide in-depth documentation on the evolution of several individual

Romance languages. The only period of their development that remains largely inaccessible is the transitional phase between spoken Late Latin (4th–6th/7th centuries) and the early Romance languages (7th/8th–10th centuries). Nevertheless, significant source material exists which provides evidence even for this period.

Fewer than a tenth of the world's languages have been satisfactorily described by linguistic studies, and researchers can rely on historical evidence for even fewer of these, not to mention knowledge of their source language. Remarkably, however, since the many Romance languages and the evolution of Latin have been extensively documented and described thanks to almost two centuries of linguistic scholarship, they offer a unique field of observation. As a consequence, the discipline of Romance studies has developed particular strengths in terms of historical, comparative and variational approaches. In these three research areas, Romance studies contribute significantly to the understanding of linguistic activity in general, as well as to the development of methodology in the field of linguistics.

Apart from being of interest to theoretical linguistics, knowledge of the Romance languages and their history is useful for the understanding of their heritage – both European and extra-European. French, Spanish, Portuguese, Italian and Romanian as national languages cover a sixth of Europe. In addition, Spanish and Portuguese are dominant throughout most of America, while French, Portuguese and Spanish are also present as primary or secondary languages in otherwise Anglophone North America and, in some cases, Africa (*cf.* 3).

Considering the factors outlined above, a Romance perspective is a useful starting point even for the study of an individual Romance language: an awareness of the underlying Romance framework leads to a deeper understanding of grammatical, lexical or sociolinguistic phenomena. In view of these considerations, this manual intends to raise 'Romance awareness' among students of French, Spanish or Italian. Even without an in-depth knowledge of all Neo-Latin languages, it is possible to familiarise oneself with their general history as well as with some of their characteristics.

This manual aims to present the basic elements of the Romance paradigm outlined above, encompassing the following topics:

- the fundamental theoretical aspects of research in Romance studies, including the history of the discipline itself (ch. 2);
- a succinct description of the present-day Romance languages (ch. 3) and their study from a variational perspective (ch. 4);
- the four main domains of language – phonetics and phonology (ch. 6), morphology (ch. 7), syntax (ch. 8), and lexis (ch. 9) – which, after a short methodological introduction (ch. 5) will be examined from the point of view of language change in the Romance languages;
- the history of the *Romània* from a sociolinguistic point of view, or the 'external history' of the Romance languages (ch. 10), including a presentation of their textual

foundations and the methods of the philological approach (ch. 10.7). The latter are of particular importance, since they determine the nature of the linguistic data and therefore the results of any analysis.

The final chapter (ch. 11) focuses on more practical aspects of research and teaching, describing the place of linguistics and Romance studies in society, in different countries and in universities.

The Romance languages and their concrete use in communication share a number of general, 'universal' linguistic features with all other languages of the world: language is a constant of a universal and anthropological nature, common to all human beings, and, at the same time, each specific language is imbued with individuality and historicity. In this manual, emphasis will be given to characteristics typical of the Romance languages; however, many of the observations made here are equally valid for many other languages. Before presenting the history of the discipline and the characteristics of the various Romance languages and varieties, we will consider some basic definitions of 'language' which are essential to avoid any misunderstandings which might otherwise arise from the chapters which follow.

1.2 The concept of 'language'

Understanding the complex phenomenon of language requires several different and complementary approaches. One of the aims of this manual as a whole is to arrive at an interpretation of language through the lens of Romance studies. The basic features of language requiring definition in this introductory section are the following:

- its three fundamental functions: social, cognitive and creative (no. 1.2.1)
- its nature as a system of signs (no. 1.2.2)
- the opposition between abstract language competence and the concrete production of utterances (no. 1.2.3)
- the basic principles of language's organisation: linearity, intelligibility and economy (no. 1.2.4)
- its variable and evolutionary nature (no. 1.2.5)
- the relationship between language-internal and external factors (no. 1.2.6)
- its two principal modes of realisation: speech and writing (no. 1.2.7)

1.2.1 Three functions of language

Language use is always motivated by non-linguistic, so-called 'external' factors which also catalyse language variation and change (*cf.* 1.3.2). Such external stimuli can be linked to three basic functions of language – a social function, a cognitive function

(linked to human intellectual faculties), and an expressive function. While these three functions complement one another in some ways, they are contradictory in others. They play a role in the evolution of the human language capacity as well as in that of the individual languages within their various present-day contexts of use.

The three functions of language first appear in the communication model established by the German psychologist Karl Bühler (1879–1963) in 1934 (*Sprachtheorie*). His ‘semiotic’ model (cf. fig. 2 below), known as the ‘Organon model’, specifies three distinct elements within the act of communication:

1. the sender
2. the receiver (referred to as *Empfänger* in the original German model)
3. the referential information (*Gegenstände und Sachverhalte*, cf. 8.1).

The expressive function (*Ausdruck*) is assigned to the sender, the social or ‘conative’ function (*Appell* [~ ‘appeal’]) to the receiver, and the cognitive function (*Darstellung* [~ ‘representation’]) to the referential information:

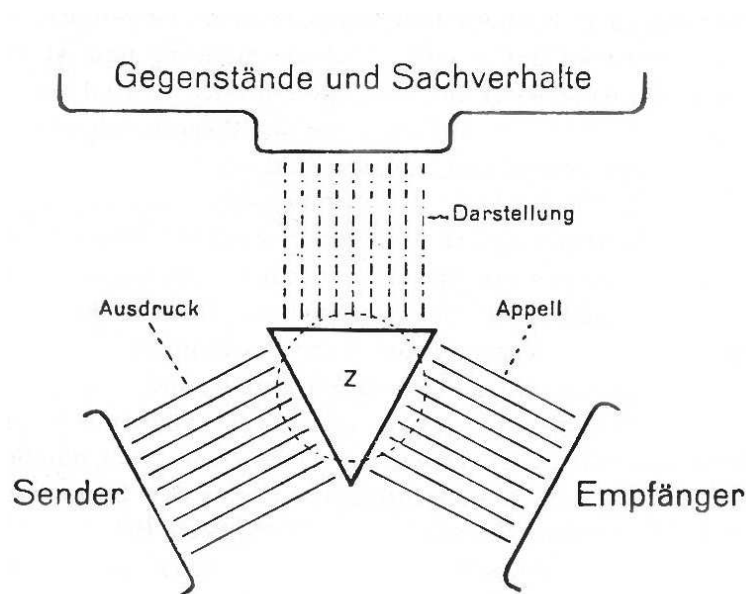


Fig. 2: Bühler's Organon model (source: Bühler 1934: 28)

The social functions of language primarily concern its use as a means of exerting influence on others and of creating distinctions between different socio-cultural groups. The capacity of language to influence people is an everyday phenomenon: we constantly use language to make requests or to convince others to act in a specific manner. In doing so, we may tell the truth, lie, or even take advantage of linguistic misunderstandings to

influence a situation; language lends itself to all these uses. In extreme cases, words can even be transformed into actions. Expressing greetings, excusing oneself, threatening someone or making promises belong to this special category of so-called 'performative illocutionary acts' (cf. 4.5.3 no. 1).

The phenomenon of group differentiation relies on the same innate and divisive human reflex to reject persons who "do not speak as we do" (cf. the etymology of Greek *bárbaros*, "barbarian", originally meaning "one who says *brrr-brrr*", i.e. who speaks an incomprehensible language). It is difficult to determine how great a role this reflex plays in the constitution of geographically defined groups that speak different languages. However, it is certain that once formed, divergent languages tend to separate groups more than they unite them. In complex societies, linguistic differences have a more direct impact on the shaping of social groups than on geographical groups. Social groups are both generated and characterised by variation in linguistic usage within a single space; in this way, language contributes to creating socio-cultural differences.

Finally, at the level of the individual, language is a powerful tool for asserting social position. In many cases, the purpose of speaking and writing is less to transmit information than to position oneself in relation to others (consider the classic example of election speeches).

The cognitive functions of language arise from its capacity to name individual concepts and to establish a relationship between them. By means of language it is possible to describe and interpret the universe and human existence; it thus leads to abstraction, and enables complex, self-referential thought. An awareness of one's own identity would not be possible without language. It also enables cultural information to be transmitted from one generation to the next, and allows for this information to be accumulated, elaborated and interpreted.

The emergence of human thought would be unimaginable without language, even though the relationship between the latter and other cognitive faculties is extremely difficult to assess. Although 'thought' extends beyond the elements of language – especially those of a specific language – its development is inseparable from the human language capacity (cf. 2.3.2 no. 2 and, especially, 9.2). In this respect, the cognitive functions of language are arguably the most characteristic and specific of the three.

The expressive functions make use of the creativity inherent in language, also fulfilling imaginative and emotive purposes. Since language is available to the individual speaker as a means of externalising personal, inner thoughts, the expressive function corresponds best to the elements of individuality present in language. This creative force is directly correlated with the variable nature of language, which, in turn, represents a quality no less necessary for the fulfilment of its social and cognitive functions.

1.2.2 Language as a sign system

Languages are sign systems, like other natural communication systems (such as bird-song) or artificial communication systems (such as traffic signs). By definition, the function of a sign is to stand for something else (*aliquid stat pro aliquo* “something stands for something else”, as affirmed by medieval scholars). At a cognitive level, linguistic signs can refer via ‘concepts’ to the external world and the relationships of interdependence that one perceives in it (cf. ‘referential information’ above). To be more exact, such is the case for lexical signs (or ‘lexemes’, such as nouns or verbs), whereas grammatical signs (or ‘function words’, such as prepositions or conjunctions) refer to relationships, which they themselves establish between other linguistic signs (cf. 2.2.1 no. 1).

According to the sign theory (known as ‘semiotics’ or ‘semiology’) of American philosopher Charles S. Peirce (1839–1914), three types of sign can be distinguished: indexical, iconic and symbolic. Indexical signs indicate relationships of ‘contiguity’, relying on spatial, temporal or causal proximity to their referents (cf. 9.3, nos. 9.3.1 and 9.3.3 in particular). Thus, ‘smoke’ as a symptomatic sign refers back to ‘fire’, as the cause of its existence. Iconic signs rely on a relationship of acoustic or, especially, visual resemblance to their referents. Characteristic examples of iconic signs are city maps or certain road signs (which in some countries display the silhouette of a person). Finally, symbolic signs are defined by convention (for example, the road sign for ‘no entry’). The linguistic sign chiefly belongs to this third category, although it can also involve elements of an iconic nature (cf. below).

Owing to their conventional nature, the use of specific linguistic signs at any given time in history (i.e. in synchrony) retains an element of arbitrariness. More precisely, the arbitrariness of the linguistic sign lies in the nature of the relationship between a given linguistic form and a specific meaning: there is no apparent reason why the same animal should be named *dog*, *chien*, *perro* or (in Brazil) *cachorro*. Linguistic signs are not, however, completely arbitrary, even when considered from a synchronic perspective. They may be motivated by iconic factors in particular, as in the case of onomatopoeia, whose phonological form expressively resembles the sound referred to (consider verbs like *blabbering* or *roaring* or the equivalents in French *papoter*, *beugler* or Italian *chiacchierare*, *muggire*). Some degree of iconicity can also be observed at the level of morphology (plural forms, for instance, are longer than their singular counterparts: in written French, *un chien* – *trois chiens*; likewise in English, *a dog* – *three dogs*, cf. 2.2.3 no. 2). Moreover, words formed by derivation or compounding necessarily display at least partial morphological motivation (a ‘lecturer’ is a person who gives lectures; likewise, the French word *sonnette* (“bell”) refers to an instrument used for the act *sonner* [“to ring”]).

The idea of an essentially arbitrary combination of a linguistic form with a meaning led to the concept of the ‘dichotomy of the sign’. This fundamental idea was introduced

by the Swiss linguist and specialist in Indo-European languages Ferdinand de Saussure (1857–1913) in his famous work *Course in General Linguistics* (*Cours de linguistique générale*, published posthumously in 1916). The linguistic sign associates a specific form – the ‘signifier’ (Fr. *signifiant*) – with an element of content – the ‘signified’ (Fr. *signifié*) or lexical meaning. At the moment of utterance, the two elements are joined to form a new unit, which, in Saussure’s classic formulation, is as inseparable as two sides of a sheet of paper.

The conventional aspect of this combination becomes apparent when one considers the fact that the form of a word can change while its meaning remains stable, or vice versa: when Lat. AQUA /'akwa/ becomes Fr. *eau* /'o/, the new form still means “water”, and when the form *mouse* (Fr. *souris*) is used in the context of computers, it no longer has anything to do with a small animal (cf. the different types of semantic change presented in 9.3). Similarly, when a bilingual person hears the French form /'arbr/ or Spanish /'arbol/, they will think of the same concept: the mental representation of a tree (i.e. the concept linked to the signified). The concept does not change, despite the fact that the term used to designate it (the signifier or phonological form) does. On the other hand, a single linguistic form can have multiple meanings, as in the example *mouse* (this phenomenon is known as ‘polysemy’, cf. 9.2.2 no. 7 (8)).

The various meanings of a ‘polysemic’ word may however be semantically motivated to some extent, since polysemy is the result of semantic change (= change in the meaning of word forms over time). When a specific form acquires a new meaning, the existing meanings of the latter are always taken into account. Thus, the use of the term *mouse* in English to designate a computer accessory is metaphorically motivated, because the object shares certain obvious features of similarity with the rodent, such as its size, shape, ‘tail’ and the ability to move quickly. Hence, today the form *mouse* (Fr. *souris*, Sp. *ratón*, Pg. *rato*, Cat. *ratolí*, Germ. *Maus*) has two meanings that are semantically linked.

To put it more precisely, in Saussure’s words: ‘the linguistic sign unites not a thing and a name, but a concept and a sound-image’ (“le signe linguistique unit non une chose et un nom, mais un concept et une image acoustique”, CLG, part 1, ch. 1, §1, cf. fig. 3):

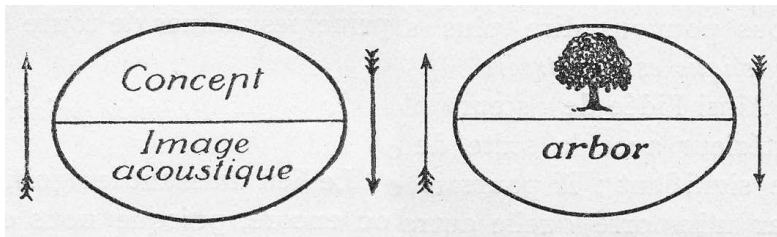


Fig. 3: The dichotomy of the linguistic sign

The specific nature of the ‘signifier’ and the ‘signified’, as well as that of the relationship established between the two will be discussed in more detail below (cf. 6.1.1 and 9.2.2). In both cases, Saussure’s initial idea remains fundamental.

Language is the most complex system of communication used by human beings, but not the only one. It is accompanied by other systems, including body language and gestures, mimicry and facial expression. These systems may even replace language under certain circumstances. In some situations, gestures are used instead of words (consider non-verbal signs signifying “yes”, or “shush”); moreover, language can be replaced by another system, for instance by sign language (cf. 2.3.2 no. 2). Language also co-operates with cultural sign systems, such as social conventions (factors in this category include the physical distance between conversation partners, table seating arrangements or the interpretation of punctuality), fashion trends or architectural styles. Focusing on certain aspects of language requires a consideration of other systems of human communication for an accurate linguistic interpretation to be achieved. Language nonetheless remains the only system of communication that can be used to refer back to itself (i.e. a self-referential system), and the only system that enables the creation of other, artificial systems.

Artificial sign systems are indeed far less complex than the natural and cultural systems previously mentioned, as one can deduce from road signs or even computer languages. The frequently mentioned Morse alphabet, on the other hand, is not an autonomous sign system, but a transposition of the Latin alphabet into acoustic or visual signs.

1.2.3 *Langue and parole*

The phenomenon of ‘language’ is not tangible *per se*. It can only be perceived in the form of one of the individual languages of the world (= ‘particular languages’, after Coseriu). Even an individual language cannot be perceived as such directly; it is mirrored only in the actual utterances produced by speakers, in spoken or written discourse. Behind these concrete acts of speech, however, there exists an abstract language competence shared by all speakers of a specific language, which, in turn, belongs to a general (or ‘universal’) language competence, common to all mankind. Language competence thus pertains to each individual, while at the same time it is shared by all human beings. It is both a personal and an inter-subjective phenomenon, with an intrinsic anthropological dimension.

The idea of an essential distinction between abstract system (= ‘*langue*’) and concrete production (= ‘*parole*’) has been an integral part of linguistic theory since Saussure at the latest (1916). ‘*Langue*’ is governed by both general and language-specific competences and is situated in the human brain. Utterances or acts of speech, on the other hand, are products of this abstract competence, and are therefore the only phenomena

available for immediate linguistic observation (the same contrast exists between phonemes and sounds, cf. 6.1.1). Linguistic introspection, which allows acceptability judgments to be made (cf. 4.1.2), bridges the two dimensions, and relies on an individual's knowledge of usual utterances.

The Saussurean dichotomy of *langue* and *parole*:

langue (= system of signs) is the system of a language that is available to a community of speakers and can be acquired by individuals, who understand the system and how to deploy its phonological, lexical and grammatical components as well as the rules for combining them. *Langue* is abstract and is shared by an entire speech community;

parole (= actual use of the system of signs in question) is the concrete realisation of *langue* by means of individual (spoken or written) acts of linguistic expression (i.e. utterances).

This Saussurean dichotomy points, moreover, to a third and complementary parameter termed '*langage*', which refers to the human capacity of expressing oneself through the use of language.

Before Saussure, the German philosopher Wilhelm von Humboldt (1767–1835) had emphasised the individual and creative use of language, which he called *énérgeia*, distinguishing it from language as a static product, which he referred to as *érgon* (cf. 2.2.2 no. 1). These two Humboldtian concepts are thus close to the Saussurean dichotomy of *langue* and *parole*, but they emphasise the highly dynamic aspect of *parole*.

In the 1950s and 1960s, Noam Chomsky (*1928) reinterpreted the '*langue*' – '*parole*' dichotomy as a contrast between linguistic 'competence' and 'performance'. He emphasises the fact that language competence is a source of unconscious and implicit knowledge that enables the individual to generate and understand an unlimited number of utterances in a given language. The main focus of the generative grammar movement founded by Chomsky is the language faculty, which is thought to determine language acquisition during the early childhood of native speakers, or 'ideal speakers', of a language. With regard to performance, the actual use of language in the form of utterances in concrete situations, Chomsky, like Humboldt, emphasises the regular presence of variation, which in extreme cases can violate linguistic norms, together with the frequent occurrence of phenomena marking hesitation and incomplete constructions. In contrast to Humboldt, however, Chomsky excludes these phenomena from his linguistic analyses, as he does not regard them as belonging to linguistic competence proper (cf. Chomsky 1957, 1965). In comparison to Chomskian terminology, the Saussurean concepts are more inclusive, and thus more appropriate for interpreting the human language capacity.

Drawing on the work of the Danish linguist Louis Hjelmslev, Eugenio Coseriu (1921–2002) expanded the *langue-parole* dichotomy by introducing a third, intermediate level, which he called the 'norm' (or 'usage norm'). According to Coseriu, while *langue* (or 'system' in Coseriu 1952) displays functional invariability, the 'norm' consists of commonly used, and thus variable, realisation patterns. Concretely, Coseriu's usage norm concerns:

- phenomena of variation in the discrete domains of language: e.g. free phonetic variants of a given word, such as the French pronunciation [baRbəkju] vs. [baRbəkʏ] for the English loan-word *barbecue* (cf. 6.2.1) or derivational variants such as It. *chiarezza* “clarity”, as opposed to the obsolete and/or poetic form *chiarità*;
- the domain of pragmatics, where recurrent usage patterns (such as forms of greeting or textual models) are treated as a category. Such conventionalised models place certain restrictions on the free use of *parole* (or ‘discourse’ in Coseriu’s terms); however, they do not constitute inviolable rules (cf. 4.5.3 no. 2).

As will be shown later, Coseriu’s idea of ‘norm’ can be defined more adequately and in more detail if replaced by two other concepts: ‘textual genres’ and ‘diasystem’ (cf. 4.5.3 no. 2).

Notwithstanding terminological nuances, it is accepted that language is organised in the brain of the individual, and that this ‘organisation’ takes place largely at an unconscious level, like human instincts. The linguistic unconscious, i.e. the individual’s language competence, is essentially beyond the reach of conscious thought. Consequently, linguistic description and interpretation are inevitably based entirely on observation of the ‘symptoms’ of *competence* or *langue*, i.e. the phenomena of *performance* or *parole*. By observing concrete spoken or written linguistic production and by comparing discourses or texts, linguists attempt to deduce the principles governing the functioning of ‘language’ in the unconscious mind. Like other sciences, linguistics then works with models through which the observed phenomena may be understood. In an ideal case, these models reflect the organisation of language in the human brain. Since the 1980s in particular, studies in neurolinguistics and psycholinguistics have made important contributions in this area (cf. 2.3.2 no. 2 and 5.2).

1.2.4 The three organising principles of language: linearity, intelligibility and economy

Processes of linguistic encoding follow three organisational principles which are essential for the functioning of language and ubiquitous in the act of speech. Linguistic expression (1) is primarily linear, (2) must be intelligible and (3) should reach its goals with minimal effort (= principle of economy).

1 Linearity of speech

The linearity of speech as a general principle was first outlined by Saussure:

‘The signifier, being auditory, is unfolded solely in time from which it gets the following characteristics: (a) it represents a span, and (b) the span is measurable in a single dimension: it is a line.’ (*Course in General Linguistics*, 70)

[“Le signifiant, étant de nature auditive, se déroule dans le temps seul et a les caractères qu’il emprunte au temps: a) il représente une étendue, et b) cette étendue est mesurable dans une seule dimension: c’est une ligne.” (*Cours de linguistique générale*, 103)]

Linearity of speech is not absolute as it is always accompanied by gestures and facial expressions, while prosody adds a further layer. It is, however, immediately perceptible and of particular importance for syntactic construction, by means of which the constituents of a sentence, organised hierarchically in the brain, are transformed into a linear sequence there (cf. 8.1.1). According to Saussure, the linearity principle of linguistic expression corresponds to the so-called ‘syntagmatic axis’, along which linguistic signs are connected according to the chain of speech (elements *in praesentia*, i.e. present in speech). The ‘paradigmatic axis’, in contrast, concerns all interchangeable elements that can be introduced in place of other elements (‘commutation’). These elements, considered as elements *in absentia* (i.e. not present in speech), fulfil identical syntactic functions, although they may differ in meaning:

Marie	↔	mange	↔	une	↔	pomme.	Paradigmatic axis
Yves	↔	veut	↔	cette	↔	pomme.	
Elle	↔	mange	↔	la	↔	tarte.	
La grand-mère	↔	achète	↔	une	↔	pomme.	
Syntagmatic axis							

2 Intelligibility

Intelligibility is arguably the most compelling principle conditioning language. In order for an act of communication to be successful, the ‘receiver’ must understand the ‘sender’ (cf. also 4.1.2). Considered from this angle, the concept of the linguistic ‘mistake’ is relativised: violating a particular linguistic rule or norm is less critical than an unsuccessful attempt at communicating. This same perspective allows us to account for the large number of apparent ‘mistakes’ in oral language (cf. 4.3.4).

Still, the condition of intelligibility is only valid within a restricted community of speakers and does not apply to individuals outside that group. One of the (secondary) functions of language is precisely to enable the formation of groups of individuals who speak a language that is unintelligible to others (cf. 1.2.1). This phenomenon is particularly pertinent with regard to writing, especially in cultural communities where the latter is mastered by only few individuals and is thus perceived as an element of power, in some cases even magic. Languages with a sacred status, in particular, are often unintelligible for the majority of the population, for example Latin in Europe from the medieval period until the present, or Sanskrit in modern India. The potentially magic status of these lan-

guages originates in the idea of their unintelligibility, in other words in an inversion of the fundamental principle of communication, respected in normal situations¹.

The multiple encoding of elements in acts of speech reduces the danger of their mis- or non-comprehension. On the language-internal level, this concerns the repeated marking of a given parameter. In the French sentence *tu as vu les deux belles maisons blanches?*, for instance, five plural markers can be identified, at least in writing. Redundancy is also apparent on a language-external level, through gestures, mimicry, and body language in general. The effect of redundancy is to underline the importance of intelligibility in communication processes, especially since it is in direct contrast with the third organising principle, that of economy.

The necessity for intelligibility also has an impact on language change. It appears to have a slowing-down effect, since each new developmental stage of a language needs to be compatible with the previous one. The coexistence of old and new linguistic forms (be they of a phonetic, lexical or grammatical nature), which may persist for several generations, even centuries, ensures that inter-comprehensibility is maintained.

As a concrete example from phonetics, in Old French, the form *françois* (meaning “French” or “related to France”) was pronounced /frān'tswɛ/. This form, containing a stressed word-final diphthong, is recorded until around 1800. From 1500 onwards, however, concurrent forms such as /frā'swa/ or /frā'se/ appeared and continued to exist alongside the old form. The polymorphy (cf. 6.1.2; 6.5.2 no. 2) has now been resolved in this specific case, so that /frā'swa/ applies only to the forename *François*, and /frā'se/ to the standard language and inhabitants of France. This example also shows that language change can occur over a period as long as three centuries (or more).

3 Economy

The third principle organising language, the factor of economy, is less immediately perceptible and has given rise to numerous controversies. But with due caution one can state that it complies with the modes of general brain functioning as well as with the requirements for successful communication. First, the brain systematically follows the path of least resistance. As it displays ‘laziness’ to some extent and since handling language is highly complex, all processes need to be simplified as much as possible. Second, in order for an act of communication to be successful, it should focus on what is essential; this favours a maximal concentration of information within a short time span.

These two forms of economy relate to different levels of language: one is linked to the functioning of the brain, the other to the organisation of utterances. Nevertheless, they act together in the case of certain linguistic phenomena, such as phonetic

¹ As an enlightening etymological anecdote, the English term *hocus-pocus* is said to be derived from the Latin or pseudo-Latin phrase *hoc est corpus (meum)*, which was no longer understandable by the common people, nor possibly even by priests who uttered it while consecrating the host and declaring its transubstantiation – an act that is in itself eminently ‘magical’.

change, where lexical or grammatical forms tend to be shortened (*cf.* the example cited in 4.2.2 of Lat. *APEM* “bee” > O.Fr. *ef*, *es*, *ee*, or Lat. *AQUA* /'akwa/ “water” > Fr. *eau* /o/, Lat. *-EBAM* /'eba(m)/ > Fr. *-ait* /ε/). In general, languages favour short lexemes (= words): forms consisting of one or two syllables are among the most frequent. In these cases, the two types of economy work together, preventing lexemes and markers from becoming too long, too bulky or redundant.

As pointed out previously, the principle of economy stands in opposition to that of intelligibility, as the latter is only fulfilled when there is a sufficient amount of contrast between linguistic forms. To return to phonetic shortening, when a word undergoes severe reduction (take, for instance, the Old French form *ef*, *ee* “bee” cited above), new formal elements may be introduced, in a process that counter-balances the principle of economy. New forms often display increased expressiveness (thus, O.Fr. *ef*, *es*, *ee* is replaced by the diminutive form *abeille* < Lat. *APICULA*). High-frequency words (such as Fr. *eau* /o/), however, may also survive in radically reduced form.

Generally speaking, the true purpose of using language is to achieve maximum expressivity with a minimal amount of effort, as is the case with other social factors.

1.2.5 The variable and evolutionary character of language

Variability is yet another essential characteristic of language. All language use involves variation of some sort. This is particularly noticeable in the case of pronunciation. Phonetic variation allows us to identify the voices of different persons (when answering the telephone, for example), from the first few words they utter. Moreover, individual variation is present in speech in the form of grammatical and lexical forms that deviate from the norm (and are therefore perceived as ‘incorrect’), and in writing (for instance in sports and fashion magazines or in political satire) in the form of unconventional metaphors or newly borrowed and derived words. Variational phenomena of this type do not destabilise the coherence of a linguistic ‘system’ in place at a specific moment; however, they are factors that initiate its transformation over time. Therefore, it is important that they be taken into consideration by linguistic theory (*cf.* 4.1.1).

These observations lead to a conclusion with huge implications: linguistic variation at a given moment in time is the indispensable, immediate conditioning force underpinning language change. Even though variation does not always lead to actual language change (a process that requires competing linguistic forms to have reached a certain level of diffusion throughout a speech community), the latter is an omnipresent feature of everyday language use. Specifically, language change can be observed in the form of neologisms (consider the numerous forms in use in various present-day Romance languages that were borrowed from English), as well as archaisms (frequent in certain textual genres, including lyric poetry or legal texts) and even in variation linked to generational differences (*cf.* 9.1.4, ‘chronological connotations’). In some cases, orthographic reforms and certain particularly frequent language ‘errors’ are indicators

of the ongoing transformation of a language. Linguistic change is also revealed in a more artificial and drastic manner in the reading of texts from past centuries, since difficulties in comprehension usually increase in proportion to the length of time between composition and reading.

All these phenomena of variation and evolution add up to a basic property of language; at the same time, they allow for a better understanding of the manner in which the latter functions.

Language evolution can be partly explained as an effect of the repeated use of linguistic forms, which tend to be shortened in accordance with the principle of economy (*cf.* 2.4.4 no. 1). Above all, however, it is a consequence of the inherent functions of language itself (i.e. cognitive, social and expressive). Language needs to be able to adapt and change, in order to react to transformations taking place in the world and society and to maintain its capacity for communication and expressivity (*cf.* 1.2.6 below).

The most striking effect of the evolutionary drive of language is the extreme diversification of individual languages, which almost certainly originated in a much smaller number of shared speech forms. Languages must have displayed a certain degree of variation from their very beginnings, due to the existence of different groups of speakers. Nevertheless, when one considers that the emergence of language is intimately linked with the emergence of the human species (= phylogeny), and that a much smaller number of people and groups of people existed in earlier periods, linguistic diversity must necessarily have been less pronounced than today.

Theories on the origin of modern humans (*Homo sapiens*) underwent considerable refinement in the second decade of the 21st century owing to advances in genome analysis. It is still assumed that man as a species has a single common origin (= monogenesis), going back to the earlier hominid *Homo erectus* who appeared at the latest around 2 million years ago and who was distinct from the former, still ape-like *Australopithecus*. It is also still assumed that *Homo sapiens* emerged in Africa, in parallel with an evolutionary enlargement in cranial capacity that favoured processes of memorisation, essential for the functioning of language.

At the same time, it has been shown that a greater diversity of early forms of humanity existed, and that genetic crossbreeding took place between *Homo sapiens* and Neanderthals in northern Europe, as well as between the former and *Homo denisova* in Asia. Equally, the origins of language are now thought to go back at least 500,000 years, while the genetic emergence of *Homo sapiens* probably dates only from around 250,000 to 200,000 years ago, as already stated by Ernst Mayr, for example (*What evolution is*, 2001, ch. 11). Language thus reached its current level of complexity well before the time of Cro-Magnon man (around 35,000 BC).

As humankind and language evolved in tandem, the latter plays an essential role in the functioning of the brain. Language is defining for man as a species – and therein resides the complexity, both psychological and cultural, of linguistic phenomena.

In the same way that all anthroponyms ultimately derive from the lexicon (*cf.* 9.1.2, 9.7.2), we can suppose that the lexicon itself, in its earliest stages, goes back to onomatopoeia, i.e. signs of iconic nature that in some way resemble the extra-linguistic reality to which they refer (consider *brrr-*, *pl-*, etc.). Once accepted by a community, however, such signs must soon have acquired an independent status, enabling the creation of new, symbolic signs defined by convention. The birth of individual signs – 'words' – was probably determined by a number of regularities as Georgescu (*La regularidad en el cambio semántico*, 2021), has shown, taking as a starting point Hilmer's highly accomplished although almost completely unknown work on Old English (*The origin and growth of language*, 1918). It is therefore logical to suppose that the production of individual words was followed by the development of syntactic and morphological features, as recent research on animal sounds has been able to substantiate to some degree.

Linguistic variation was increased by migration and the spread of human settlements. Since groups generally tend to differentiate themselves from their immediate neighbours by their manner of speaking, it is not surprising that groups separated by considerable geographical distance tend to remain isolated from innovations taking place elsewhere. The combined factors of variation and geographical separation are thus responsible for the vast number of individual languages in existence today, which are estimated to be between 3,000 and 6,000.

The website of the *World Atlas of Language Structures* <wals.info> lists some 2,700 living languages (although about ten of the 24 Romance 'languages' cannot be classed as such). The site <ethnologue.com> lists more than 7,000 'languages' (here, the categorisation is even less reliable as the 43 Romance varieties mentioned do not distinguish between regional dialects and actual languages: French is accorded the same status as its dialects, such as Picard or Walloon).

Despite the great diversity of modern languages, we should not forget that they all share the same fundamental principles of linguistic functioning, and that language serves more to unite than to divide, as suggested by Wilhelm von Humboldt (*cf.* 2.2.2 no. 1). The fundamental similarity of all languages becomes apparent when one considers the ease with which they influence one another; so-called linguistic 'interference' can take place between even the most diverse languages (*cf.* 3.6).

The exact number of languages in the world cannot be established for various reasons:

1. Many speech varieties are not sufficiently well described to allow their degree of autonomy as 'languages' to be assessed.
2. In the case of non-standardised varieties (i.e. which exist only in a dialectal, non-elaborated form) it is very difficult to distinguish those which form a dialectal group within a language from those which are to be regarded as individual languages in their own right.
3. Distinguishing one language from another is in itself not always clear cut. These points will be discussed in more detail with regard to Romance languages (*cf.* 3.1 and 3.4).

It is nevertheless apparent that the great tendency of language towards differentiation and divergence, which has been present since the dawn of humankind, has been reversed over the past 50 years. Today, a new process can be observed whereby national languages and languages used for cross-national communication gain more and more foothold in the world, while languages used for communication within restricted groups tend to disappear. Consequently, ‘small languages’ and regional dialects are either no longer spoken or gradually converge with neighbouring standard languages. It is likely that by the end of the 21st century, about half the languages still spoken today will be extinct, and that many others will no longer be used by younger generations of speakers. The process of ‘language death’ (sometimes termed ‘glottophagy’) is a worldwide process which can be compared to the extinction of living species, since both result in a decrease in biological or cultural variety. The fact that there are several endangered languages within the Romance family reflects these general tendencies (cf. 3.4).

To some extent, however, the movement towards uniformity is counterbalanced by centrifugal phenomena, including the development of ‘sociolects’ (language varieties tied to a specific socio-cultural group such as those specific to urban peripheries) and the ‘pluricentric codification’ of major languages of communication (as illustrated by the multiple forms of elaborated standard English in former colonies of the British Empire) (cf. 10.6).

The analogies between the development of languages and that of animal or plant species are very striking. Language change and the evolution of species thus share important characteristics (cf. the synthesis by Ernst Mayr, 2001), while at the same time displaying some essential differences. Shared features are:

- like living animal species, languages diverge owing to geographical or spatial isolation. A language needs to be spoken in two places with no close link to one another in order to split into two new languages;
- the evolution of species can be accelerated in small populations: if a species is geographically divided into one larger and one smaller group, the smaller group may evolve at a faster rate. The same rule seems to apply to languages and would explain the large number of languages in New Guinea or in Amazonia, spoken by small groups who have very little contact with one another. With regard to the *România*, it would account for linguistic diversification in the Alpine valleys or in the Pyrenees, for example.

Differences are:

- in contrast to the evolution of species, language change does not involve genetic reproduction; therefore, it does not take place from one generation to the next, but in a continuous manner. The closest parallel to language change in biological evolution is possibly found in the evolution not of animals or plants but of prokaryotes (e.g. bacteria), which rely on *asexual* reproduction, allowing a high level of interference between different types (comparable to language contact);
- in the evolution of species, *natural selection*, where better adapted individuals have a greater chance of survival and thereby pass on more of their genes to the general population, plays

a paramount role. In contrast, languages are unique to a single species – humans – and they *are intrinsically incapable of improvement*. It may be possible to enlarge the vocabulary of a language and to develop a more complex syntax; however, all languages share the same communicative qualities and are capable of expressing any intended content, even though in some cases this may involve considerable effort. Linguistic change is thus *not* induced by natural selection.

Finally, the variable and evolutionary nature of language is reflected in the (internal and external) history of individual languages (*cf.* 1.2.6). Each language or other variety has its own unique history, and as far as written languages are concerned, this historical and evolutionary pathway plays a significant role in their use. The large number of borrowings from Latin present in the Romance languages are particularly illustrative in this respect (*cf.* 6.6 and 9.6.4).

As a result, two essential approaches within the field of linguistic studies can be distinguished: one concentrates on the change that a language undergoes over time (= 'diachronic linguistics'), whereas the other considers the state of a language at a particular moment or within a defined timespan (= 'synchronic linguistics'). It is important to note that a synchronic lens can also be applied to the study of a historical stage of a language. It is possible, then, to study Latin of the 1st century or Spanish of the 8th century AD from a synchronic perspective. Moreover, an accurate synchronic description requires the consideration of linguistic variation present during the relevant historical period (*cf.* the epilogue, ch. 12). The opposition between 'synchrony' and 'diachrony' constitutes the third of the main dichotomies in language introduced by Saussure, in addition to those between *signifiant* and *signifié* and *langue* and *parole*.

→ Loporcaro, *Teoria e principi del mutamento linguistico*, RSG 3, art. 221

1.2.6 'External' factors relating to language use and evolution

In order to describe a language adequately, not only its 'internal' (*i.e.* phonetic, grammatical and lexical) features, but also its 'external' characteristics, must be taken into consideration. The latter define the status of a given language within a society and include issues such as its geographical extension, number of speakers, coexisting languages, the infrastructural, political and cultural organisation of the society in which it is used and the modes of its concrete use (e.g. whether it is purely oral, or written as well, which domains it is used in, whether it is used by the media, etc.).

Though there is no direct connection between geographical, social and political factors and concrete phonetic or grammatical features, the external factors mentioned do have an impact on linguistic usage and on the internal evolution of a language. Thus, language change is triggered by transformations taking place in the outside world or in the mind of an individual, and effected through acts of communication driven by the cognitive, social and creative functions of language (*cf.* 1.2.1). The desire of speakers (1) to account for changes in their environment and (2) to position themselves in relation

to others, (3) acting in conjunction with the expressive potential of language, induces variation and, as a consequence, linguistic change.

The concept of indirect interdependence is neatly illustrated by the following (non-linguistic) example: a pathway across the lawn of a university campus is spontaneously shaped by the coincidental concentration of several individuals' intentions to reach their destination by the shortest route. Rudi Keller includes this model in a theory of the 'invisible hand', inspired by economist Adam Smith (1776, cf. Keller, *Sprachwandel. Von der unsichtbaren Hand in der Sprache*, 1994). This theory does not in itself explain concrete linguistic changes; however, it helps towards a better understanding of the indirect nature of the relationships between patterns and effects in language. Thus, the realisation of a cognitive, social or expressive aim entails the use of a given linguistic form, which, in turn, is subject to the internal regularities of language functioning.

Major historical changes within a language are always the result of 'external' events or developments that accelerate and influence the process of language change (cf. 5.2.1, for the problem of historical 'periodisation'):

- geographical separation causes new languages to develop (cf. 1.2.5 above);
- language contact is a cause of linguistic interference and borrowing (cf. 3.6 and 9.6);
- following the establishment of political states, languages may be standardised and put into writing (cf. 10.5.3);
- social differences cause the creation of new linguistic varieties (cf. 6.4);
- technological development leads to specialised vocabulary and specific syntactic forms (cf. 6.4.3);
- changing concepts of politeness may bring about the introduction of particular forms of linguistic expression (such as formal and informal modes of address, e.g. Fr. *vous* vs. *tu*; cf. 7.5.3).

In order to interpret changes at the internal level of a language, initial motivations in the external environment need to be considered. At the same time, however, internal transformations follow principles which have a functional logic of their own, and which also require description and interpretation. The double nature of every change complicates in-depth explanations of concrete phenomena of variation and transformation.

To provide a concrete phonological example, when the Romance languages emerged, the set of voiceless plosives *p*, *t*, and *k* in intervocalic positions became voiced in the Western regions (Latin *RIPA* became Sp. *riba*, Fr. *rive*, etc., whereas the plosives were conserved in O.It. *ripa*, Romanian *râpă*; cf. 6.5.4 no. 1). Around the same period, identical phonemes underwent affrication in certain Germanic languages (the Germanic form *appul* evolved to German *apfel* ['second Germanic consonant shift'], but remained stable in English *apple*). Divergences such as these between phonetic transformations testify to the unidirectional nature of changes in language. Conversely, the fact that whole series of phonemes undergo a similar evolutionary process calls for other explanations, which are tied to the general principles of communication and largely transcend the level of isolated phenomena.

1.2.7 Speech and writing

Speech is the original and most natural form of linguistic expression. Human language originated, evolved and lived on for hundreds of thousands of years in oral form. Only some 5,000 years ago did writing appear, and several more millennia were required for it to reach a substantial percentage of different human communities and their members.

Although writing appeared much later than spoken language, and is secondary to speech in an ontological sense, the two have become partners of equal status for literate individuals. Thus, the brain easily incorporates written data, and it does not take it any longer to access a given concept via a written form than by way of a spoken one or an image. Moreover, in societies where it is used, writing has an unparalleled impact on language use and evolution.

Initially, writing allows communication over distance and ensures the preservation of texts over time. The processes of textual condensation and memorisation and the diversification of textual genres then allow writing to considerably increase the syntactic complexity and the vocabulary of a language. These factors form a foundation enabling the development of areas of specialised knowledge, such as the sciences, which in their elaborated form would be unthinkable without writing. Furthermore, writing promotes the formation of standard varieties of languages, which are indispensable for the governing of modern societies.

Writing, then, enlarges the spectrum of the possible uses of language while intensifying the influence of each of its three essential functions: cognitive activities (memorisation, etc.), social interaction (long-distance communication and administration) and creative potential (diversification of the uses of language). Writing transforms language, broadening and intensifying its overall impact. Therefore, it is not merely another external factor of language like geographical or political context, but an actual mode of language use.

1.2.8 Summary

Our attempt at defining the nature of language has assembled a set of somewhat diverse aspects:

1. its three basic functions (cognitive, social and creative),
2. its nature as a system of signs,
3. the dichotomy between its essence as a system (*langue*) and its concrete appearance (*parole*),
4. its three main principles of organisation (linearity, intelligibility and economy),
5. its variable and evolutionary properties,

6. the dichotomy between its internal organisation and the constellation of external factors on which it feeds,
7. its two modes of speech and writing.

These features cannot be placed in any sort of hierarchy; rather, they can all be considered starting points that lead to a better understanding of language. The complexity and the significance of language are often underestimated, even by scholars who have dedicated their entire life to linguistic issues (*cf.* the epilogue, ch. 12). To grasp the importance of the role played by language, it suffices to attempt to imagine existence without it. Such an attempt is bound to fail; abandoning language would mean giving up the essence of our identity.

1.3 Different ways of interpreting language

The areas of interest within linguistics are as varied as language itself. The discipline focuses both on inherent general aspects and on individual historical ones, which are reflected in two major research perspectives, one theoretical, the other variational. These perspectives are complementary to one another and are mutually enriching. Understood in this way, linguistics, like most branches of modern science, developed at the beginning of the 19th century (with pioneers such as Wilhelm von Humboldt). But reflection on language existed as early as Ancient Greece (for instance in the work of Aristotle), or in India during the first millennium AD, in Western Europe during the Middle Ages (consider Dante Alighieri), and at the beginning of the modern era (e.g. Henri Estienne and Bernardo Aldrete).

The theoretical linguistic approach focuses primarily on what appears to be general or constant in language: principles of communication and cognition, semantic and pragmatic constants as well as phonological, morphological, syntactic, and lexical configurations. Important bases for this type of linguistic study are European Structuralism (represented by Saussure, the Prague school, Tesnière and Coseriu) and American Structuralism (Sapir, Bloomfield, Boas and Hockett). It is present today in the various branches of generative linguistics (Chomsky, Jackendoff), of functionalism (Martinet, Dik and Halliday) and of construction grammar (Fillmore and Kay), enriched by the more recent fields of cognitive linguistics, neurolinguistics and psycholinguistics.

Variational linguistics focuses more on relations between language and society and on the phenomena that arise from this interface. These include regional dialects, varieties linked to socio-cultural groups, specialised and technical languages, linguistic differences between various textual genres and the relationship between writing and speech. Variational linguistics is closely linked to historical approaches as every change presupposes earlier variation within a specific context of language usage.

Both approaches can be applied to either an individual language, or to a set of different languages, which, in the second case, are compared to one another with the aim of identifying ‘typological’ similarities and differences. It is important to note that the two approaches intersect and are complementary. The variationist study of a given language or variety relies on the principles of general linguistics, to which its results contribute. At the same time, typological research is restricted to the study of the multiple varieties of individual languages, comprehensive knowledge of which is necessary for coherent interpretation.

Finally, as shown above, linguistic questions can focus on a specific period in history or on the evolution of language over time (*cf.* 1.2.5). In the first case, a given synchronic stage of a language is analysed; this may be a present-day language or texts from a historical period. The second approach addresses diachronic language change, both as regards internal features (such as phonetic and lexical transformations or change in phrasal structure) and external aspects (concerning language use, its political role or the development of written or standardised languages). As in the case of the research perspectives outlined above, the two axes, synchrony and diachrony – as defined by Saussure – also mutually complement each other, since the one cannot exist without the other.

In research, the four parameters discussed above can be combined flexibly: (i) theoretical vs. variational, (ii) one individual language or variety vs. several different languages, (iii) synchronic vs. diachronic, and (iv) present-day or past language. The terms ‘synchronic’ and ‘diachronic’ tend, however, to be used improperly in connection with linguistics, the former being identified specifically with (i) the theoretical analysis of (ii) a given language – usually of its standard variety – (iv) in its current state. But a synchronic approach can also be applied (iv) to a medieval period, (i) within a variational framework and (iii) from a comparative standpoint. As for ‘diachronic linguistics’, it is often taken to refer to the analysis (iv) of the historical stage of (ii) a given language, which actually equates it with synchronic analysis applied to a past stage of a language (= ‘historical linguistics’).

2 Emergence and development of Romance linguistics

2.1 The foundations of (Romance) linguistics before 1800

2.1.1 Romance linguistics within the general framework of linguistics

Over the course of its history, reflection on language has followed a path that has taken numerous twists and turns, many of which are at present without major importance for an understanding of linguistic phenomena. Often, it is of greater interest for a more general historiography of the sciences than for that of linguistics in itself. Nevertheless, the history of linguistics has a strong epistemological dimension, since the terminological inventory used today and the concrete working methods and questions pursued are the result of gradual elaboration. The course of their development also allows an understanding of the diversity of modern scientific traditions, which are often complementary though heterogeneous.

The history of linguistics in general, and even that of Romance studies in particular, has benefitted from a vast amount of research, which cannot be presented here. The objectives of this brief section are:

- to outline the general tendencies of the evolution of (Romance) linguistics over the course of the past two centuries, thereby bringing the researchers mentioned throughout this manual into context;
- to illustrate the extent to which this discipline has evolved since its emergence and the intellectual achievements that have contributed to its continuous development, to which we owe the clarity of the concepts and methods in use today;
- to provide, by means of the above, an epistemological dimension which allows a better understanding of the aims and methods of linguistics;
- to highlight a certain number of relationships between linguistic research and the historical, political and cultural context in which it is situated.

We will briefly present the emergence of the modern tradition since Antiquity (2.1.2). Linguistic thought of the Middle Ages and early modernity will be discussed in greater detail in the corresponding sections on external history, since its development was essential for standardisation (cf. 10.5.3 no. 3 and 4). We will then focus on the emergence of the historical-comparative paradigm in the 19th century (2.2.1) and on its further development until the middle of the 20th century (2.2.2–2.2.4). This paradigm remains fundamental for Romance linguistics today, both in terms of the objects studied and the lines of research pursued, and even, to a certain extent, its theoretical and methodological approaches. Finally, we will present the changes to which the established paradigms were subject during the 1960s/70s, with a particular focus on the methodological innovations which have their roots in that period (2.3).

Since our aim is to highlight the characteristics of Romance linguistics, other fields of research such as generative linguistics, creole studies or neuro-linguistics will only be mentioned briefly, despite their undoubted importance. Furthermore, the years from the 1980s onwards will not be covered in this chapter, since the literature of this period forms the basis of the following chapters and will be dealt with there. The present handbook as a whole is thus intended to document the current state of Romance linguistics in all its diversity.

2.1.2 The birth of linguistic studies and the rise of the professional ‘linguist’

Linguistics as a science is fundamentally based on the auto-reflexivity inherent in all language use. As we have seen, language is the only sign system that allows and even requires self-reflection. At every possible level, language production and perception involves human judgement and evaluation – in other words, operations relating to a metalinguistic level – albeit in a very intuitive manner which may remain unconscious. In synchrony, this type of evaluation leads to the intuitive and immediate reconstruction of missing sentence parts or to the identification of a speaker’s social position and origins during oral interaction. Diachronically, for example, it provokes the alignment of verb or noun paradigms on the basis of analogy, as well as the phenomenon of folk etymology or the reorganisation of word families (*cf.* e.g. 7.6 or 9.3.5). Humans do not use language as a ready-made tool, but control and shape it on the basis of partially pre-conceived elements. Reflection on language is an integral part of linguistic competence.

‘Linguistic’ thought in the modern sense began when the act of thinking about language, which is innate to human beings, was consciously put into words in a systematic and abstract manner. The earliest palpable phenomenon in this respect was the creation of writing, which is an act involving constant reflection. Writing was generally followed by the emergence of thought on language, as in the case of Chinese, Sanskrit or the Semitic languages, illustrated, for example, by the work of the important ancient Indian grammarian Pāṇini, who is thought to have lived at some time between the 6th and 4th centuries BC.

More strictly metalinguistic evidence appeared with the elaboration of the Ancient Greek script as the first ‘true’ alphabetic writing system (using separate symbols to indicate both vowels and consonants). Many of the great philosophers from Ancient Greece (including Heraclitus, Plato, Aristotle and stoics such as Zeno and Chrysippus) pondered the nature and functioning of language. They raised the issue of the linguistic sign and its relationship to the outside world, and established the basis for syntactic analysis and the identification of word classes. By doing so, they laid the foundation for Western linguistic thought, and therefore also for Romance studies.

The strikingly ‘modern’ character of some of these age-old lines of reasoning (formulated practically *ex nihilo*) emphasises the extent to which linguistic reflection is simply the expression of an awareness of language that exists independently of science.

Although observations and experiments within specific scientific traditions help improve coherence when interpreting linguistic phenomena, linguistics essentially draws from questions relating inherently to human nature, and is thus based on intuitive judgements that precede its development as a scientific discipline.

For the intellectuals of Graeco-Latin Antiquity, thought on language was only one issue amongst many. The distinction of different parts of speech by Aristotle or the definition of classes of signs by Saint Augustine (*De magistro*, 389) were both part of a much broader body of philosophical and theological work. This is less true in the case of Isidore of Seville (ca. 560–636), who is often considered to be the last author of Antiquity (cf. 10.3.1) and who is best known today for the etymological explanations he provides in his encyclopedic work *Etymologiae*, which are revolutionary despite their sometimes incongruous nature (cf. Schweickard, '*Etymologia est origo vocabulorum*', 1985). One would not, however, consider this thinker to be a 'linguist'.

The only intellectuals whose profession was linked to questions of language were teachers of rhetoric or foreign languages (particularly those who taught Greek in Rome and Latin in Greece). Their teachings led to the first descriptive grammars of Latin and Greek, the most significant of which are the treatises by Donatus and the pseudo-Probus in the 4th century, Priscianus in the 5th century and Boethius at the beginning of the 6th century. At this early point in history, however, these branches only intersected to a limited extent with the theory of language.

This situation did not change significantly throughout most of the medieval period, as the reception of the linguistic heritage of Antiquity within the network of monastic and cathedral schools remained limited. It was not until the 13th century, in particular, that the thinkers of Antiquity found imitators, translators and late successors. Intellectuals such as Thomas Aquinas (1225–1274) or Albertus Magnus (1205–1280) revitalised Aristotelian thought. Scholasticism produced important thinkers who investigated the nature of signs. Dante Alighieri considered the aspect of geographical variation in the Romance languages (in his work *De vulgari eloquentia*). In the 14th and 15th centuries, a flourishing tradition of glossography and certain other forms of reflection on language (such as grammatical description or language learning guides) played a role in the elaboration of written Romance languages, though these were not the endeavours of 'linguists' in the modern sense of the word (cf. 10.4.3 no. 4).

Boundaries began to blur in the 16th and 17th centuries, when linguistic thought became an active component of great importance in the elaboration of modern standard languages. The aim of linguistic discussions of this period was less to understand the abstract functioning of language than to create explicit prescriptive norms for a given language. Language nevertheless remained a subject which prompted intense reflection. The intellectual achievements of scholars such as Robert Estienne (the author of the first *Dictionnaire françois-latin*, 1539), Vaugelas (who developed his theory of correct usage in his *Remarques sur la langue française*, 1647) or Aldrete (who provided an ade-

quate description of phonetic change, 1606) were succeeded in the 17th and 18th centuries by those of the authors of the *Grammaire de Port-Royal* (Arnauld and Lancelot, 1660) and of intellectuals such as Leibniz (1646–1716), who produced works on the origin of language and the importance of Sanskrit. These scholars prepared the way for the establishment of linguistics as a scientific discipline in the 19th century (cf. 10.5.3 no. 3).

Finally the intensification of university teaching transformed linguistics into a profession: whereas from the 16th to the beginning of the 19th century, only teachers of language and rhetoric were able to rely on their linguistic competence for survival (and with some hardship), the theorisation of language has since become a reasonably well-established, socially acceptable profession.

The term ‘linguist’ itself mirrors this transformation. It is attested sporadically during the modern period with the meaning of “a person who studies languages, especially old languages” (Neo-Lat. *linguista* is documented for the first time in 1593, but it had almost certainly been borrowed into other languages earlier than this: cf. Engl. *linguist*, 1582; German *Linguist*, 1593; Fr. *linguiste*, 1632; It. *linguista*, before 1700). It received its current meaning of “specialist of the (scientific) study of languages” around 1800 (used for the first time in Germany in 1794, with calques appearing in Italian and French from 1812 onwards). The dates of the appearance of the term *linguistics*, derived from the name of the profession, are even more significant: German *Linguistik* appears in 1777 (alongside *Sprachenkunde* and *Sprachwissenschaft*), followed by the French calque *linguistique* in 1809 (generalised in 1826) and by Italian *linguistica* in 1848 (cf. DELI, TLF, OED and, in particular, the TLF-Étym article *linguiste*, *linguistique*, revised by F. Rainer).

Linguistics thus became a subject which would eventually be taught at university level. Its influence on society, although often indirect, increased at the same rate as that of higher education in general. Moreover, the growth of the discipline led to an ever greater diversification of the subject matter.

In addition to shaping linguistics as an academic discipline, the expansion of written culture also brought about the development of popular thought on language. Intellectuals, teachers, writers and others without an academic background in linguistics began to take part in discussions on the nature and use of language. Such trends, termed ‘folk linguistics’ or ‘lay linguistics’ (German *Laienlinguistik*) found expression in popular media such as newspaper columns and readers’ letters and in discussions on subjects such as minority languages, orthography reforms or the use of Anglicisms in other European languages (cf. 10.6.3).

Historically speaking, folk linguistics has some aspects in common with the ideas of early intellectuals such as Vaugelas, although the latter was writing in the absence of a linguistic discipline *per se*, instead paving the way for its development. The most direct precursor of lay linguistics should perhaps be sought in metalinguistic humour such as the imitation of accents found in Molière’s plays. The intuitive component of these

trends ultimately leads back to the origins of linguistics in Antiquity, emphasising once again that reflection on language is a preoccupation inherent to humankind.

2.2 The historical-comparative paradigm in Romance studies

2.2.1 The foundations

Several trends in the study of language, each of which underwent marked developments at the beginning of the 19th century, merged to form the field of Romance studies. Scholars in the field of Indo-European were the first to shed light on relationships of dependence between various languages, enabling them to be grouped by branches and families and thereby providing a basis for the objective identification of the Romance languages.

In parallel, a Romanticist interest developed for European cultures, which were considered to be authentic, traditional and ‘natural’, and whose traces were sought in literary texts from the medieval period, as well as in folklore and modern dialects. This interpretation was not free from misunderstandings: medieval literary production was in no way ‘natural’ but reflected finely elaborated aesthetic ideals that were linked to courtly culture. As for dialects, these are a result of centuries of language evolution and do not reflect a supposedly unchanged state which can be traced back to distant ancestors.

Nonetheless, the three-way comparison of present-day standard forms with their medieval and dialectal counterparts within a given historical language provided a solid and significant basis for linguistic studies. It was in the fields of the Romance and Germanic languages in particular, with their wealth of old texts and their marked dialectal varieties, that this revolutionary approach produced exceptional results. Paradoxically, then, the distorted interpretations of the scholars of the Romantic era contributed to the creation of a stable foundation for the modern, scientific research paradigms in linguistics.

Important innovations in linguistic theory as well as the continuation of grammatical and lexicographical practices, which had already been developed to a great extent during the 18th century, were involved in shaping linguistics as a modern discipline in Europe. In their beginnings, Romance studies were only an indistinct area within this new field of linguistics, which comprised not only the medieval and modern Romance and Germanic languages, including English as a ‘bridge language’ between the two families, but also the classical languages Latin and Greek.

Among the Romance languages, scholars were particularly interested in French, Occitan and Italian, as a great deal of medieval evidence is available for these languages and they exhibit abundant dialectal variation. Languages such as Spanish, Portuguese and Catalan or even Sardinian and Romanian remained of secondary importance, and Romansh, Ladin, Galician and Francoprovençal were barely even recognised as lan-

guages in their own right. Interest in the *România nova* and the *România creolica* did not appear until the 20th century, above all in the second half (cf. also 3.5.1 no. 5 for the number of Romance languages identified by the authors of various manuals).

The development of the historical-comparative paradigm can be divided into three distinct phases: a preparatory phase that lasted for approximately the first half of the 19th century (cf. 2.2.2 below); a formative phase during which Romance philology began to gain recognition as a distinct academic discipline and which saw the development of the first theories (throughout the second half of the 19th century; 2.2.3), and finally, a phase of elaboration and diversification during the first half of the 20th century (2.2.4).

→ Barbato, Diez, Meyer-Lübke and Co. *The founding of Romance linguistics*, OxfEnc

2.2.2 The preparatory phase (ca. 1800–1850)

The preparatory phase is marked by the influence of an older, established tradition of grammatical and lexicographical practice (and theory) combined with the more recent factors mentioned above: innovations in linguistic theory (no. 1 below), the discovery of a genetic relationship between languages by means of comparison (no. 2), and the study of old texts and dialects (no. 3).

1 The linguistic thought of Wilhelm von Humboldt

Among the new language theorists, the philosopher and diplomat Wilhelm von Humboldt (1767–1835) stood out due to his remarkably modern ideas. He was interested both in the differences between languages and in the features they share (cf. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*, 1836²). According to Humboldt, every specific language reflects the linguistic competence of its speakers, and all human beings therefore have more in common through their languages – even though these differ – than divides them. This last intuition is an extremely powerful one for all reflection of an anthropological nature.

Moreover, Humboldt formulated the concept of the double nature of language, i.e. language belongs both to the individual and to a community of speakers (it is both individual and shared), arguing, in parallel, that language is not a fixed product (*érgon*), but an ever-renewing activity (*enérgeia*), which he describes as the “ever repeated *mental*

² This fundamental work was translated into English 150 years later and given the title *The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*. It was translated by Peter Heath and published by Cambridge University Press in 1988.

labour of making the articulated sound capable of *expressing thought*" (English trans. 1988: 49).

Thus, Humboldt's ideas form the basis of all methods that focus on the production of speech, or *parole*, in order to interpret language, instead of seeing it – as structuralism and generativism tend to do – as an abstract system, or *langue*.

Finally, Humboldt addressed the issue of the relationship between language and thought. Although he overestimated the influence language has on thought at a given moment (i.e. in synchrony) and underestimated its importance for the evolution of thought (i.e. in diachrony), his views almost attain the level of complexity which characterises recent research in cognitive linguistics.

Humboldt had no immediate impact on linguistics, and the discipline as a whole did not absorb the principles he had established until 150 years after his death. Nevertheless, his theories heralded the beginning of modern linguistics.

2 The genetic and comparatist approach

The comparatist approach, brought to the fore by Humboldt, involved the idea of genetic relationships between languages. This concept was already present in discussions on the origins of European languages initiated in the 16th century. Although the etymological relationship between Latin and French suggested by Gilles Ménage in his *Origines de la langue françoise* (1650) includes many curiosities (such as the hypothesis that *Fr. savoir* comes from its Latin synonym *scire*, hence the graphic form *sçavoir*), modern science has nevertheless confirmed approximately 70% of his propositions, showing that these are in fact true reflections of the changes that occurred. The laws of phonetic change between Latin and Spanish formulated by Bernardo Aldrete at the beginning of the 17th century (*Del origen y principio de la lengua castellana*, 1606) attain an even higher degree of accuracy, despite being based on intuition (cf. 10.5.3 no. 3).

The foundations for a coherent theory on the relationships between languages were provided by the comparison of different Indo-European languages. The first to take this approach was the British scholar and judge Sir William Jones (1746–1794), stationed in India, who identified the relationships between Sanskrit and Greek and Latin (cf. *The Sanscrit Language*, 1786, in which he proposed that the three languages have a common origin, which, moreover, could be linked to Gothic, the Celtic languages and Persian).

Jones was followed by Franz Bopp (1791–1867), a close friend of Humboldt's, who found marked similarities in the verbal morphology of geographically separate languages such as Sanskrit, Persian, Greek, Latin and the Germanic languages (1816, *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprachen*³). His discovery established not

3 "On the conjugation system of the Sanskrit language in comparison with that of the Greek, Latin, Persian and Germanic Languages".

only the principle of genetic relationships between languages, but also the nature of the regularities that link the languages of a single family: these share morphological and syntactic structures and have words that not only resemble one another in their form and meaning but also display regular patterns of correspondence in their phonetic evolution (cf. also Rasmus Kristian Rask and his *Grammaire étymologique du danois*, written in 1810).

Research by Bopp and some of his contemporaries, such as the brothers Friedrich Schlegel (*Über Sprache und Weisheit der Inder*, 1808) and August Wilhelm Schlegel (*Indische Bibliothek*, 1820–30), marked the birth of Indo-European studies. This discipline, in turn, immediately led to the emergence of Romance studies, since the comparative principles established earlier enabled scholars of the time to easily identify the Romance languages as a distinct entity. The awareness of Romance identity was already apparent in the first textbook of Romance linguistics, which was published in 1831 and consisted in a brief general overview of the phonetic and morphological characteristics of the six Romance languages that were to become dominant national languages during the course of the 19th century (Lorenz Diefenbach, *Ueber die jetzigen romanischen Schriftsprachen, die spanische, portugiesische, rhätoromanische (in der Schweiz), französische, italiänische und dakoromanische (in mehreren Ländern des östlichen Europa's [...])⁴*).

3 Philology and studies on modern dialects

From the very beginning, the comparatist approach to language was intimately linked to the study of old texts and dialects. The German philologist Friedrich Diez, for whom the first chair of Romance linguistics was created (i.e. the chair of 'Romanische Philologie' in Bonn) began by translating Old Spanish romances into German and editing troubadour poetry (1818, *Altspanische Romanzen*; 1826, *Die Poesie der Troubadours*). He then wrote his historical and comparative grammar of the Romance languages (1836–1843, *Grammatik der romanischen Sprachen*, 3 vols) and, finally, his major work, the etymological dictionary of the Romance languages, which also takes into account various dialects, above all those of Italian (1854, *Etymologisches Wörterbuch der Romanischen Sprachen*). The tradition of modern etymological research in Romance studies is entirely founded on this dictionary.

A similar development took place in the field of Germanic Studies: the brothers Grimm, who undoubtedly influenced Diez, themselves edited medieval texts, collected popular tales, studied dialects and wrote the historical dictionary of German that remains an authoritative reference today.

Occitan received particular attention at this time, since it was sometimes considered to be the most faithful reflection of the common basis of the various Romance

4 "On the current Romance written languages, Spanish, Portuguese, Romansh (in Switzerland), French, Italian and Daco-Romance (in several countries of Eastern Europe ...)".

languages, owing to certain conservative phonetic features: Diez's edition of troubadour poetry (1826) was preceded by the Sanskrit specialist August Wilhelm Schlegel's *Observations sur la langue et la littérature provençales* (1818) and followed by François Raynouard's *Lexique roman ou Dictionnaire de la langue des Troubadours* (1838–1844), the first volume of which contained an anthology of troubadour poetry.

The examples of Diez, Schlegel and the brothers Grimm show that the various traditions that took shape in the first half of the 19th century, setting themselves apart from the thought of the three preceding centuries, cannot be separated from one another. The discovery of genetic relationships among the Romance languages, the first philological studies and the interest in modern dialects were parallel developments. The historical-comparative description of Romance grammar and vocabulary, which chiefly began with Diez and Raynouard, was thus closely associated with this early philological and Romanticist movement.

2.2.3 The formative phase (ca. 1850–1900)

The intellectual trends that led to the emergence of Romance studies in the first half of the 19th century were concentrated in German-speaking Europe, as can be inferred from the names of the scholars mentioned. During the second half of the century, the concept of Romance studies diversified and became more international. French- and Italian-speaking philologists developed editorial philology and dialectology, as well as geolinguistics, which was conceived as a historical science. Methodological progress nevertheless remained the work of a limited number of intellectuals within academic settings. Personal contacts and written correspondence crossed the boundaries of the new national states that were still in the process of becoming established.

During this period, an important contribution made by research in the sphere of Germanic studies was the development of systematic thought on sound change, conceived as the result of 'sound laws' (German *Lautgesetze*, cf. Hermann Paul, *Prinzipien der Sprachgeschichte*, 1880). The idea of almost mechanical correspondences that allowed the projection of the sounds of one language onto those of another completed and reinforced the reasoning that had taken shape at the beginning of the century. The key figures of the linguistic movement that had developed these theories received the name '*Junggrammatiker*' (Neogrammarians or 'Young Grammarians'), a term which was first coined by their opponents but was subsequently proudly adopted by the proponents themselves.

The desire to assimilate the 'principles' of linguistic and literary study to those of the natural sciences was part of the spirit of this positivist and resolutely 'scientific' period, which at the same time saw the appearance of Charles Darwin's *The Origin of Species* (1859), Ernest Renan's *Life of Jesus* (1863) and Friedrich Nietzsche's *Zarathustra* (1883–1885). Even though phonetic change actually follows highly complex patterns, as

will be shown, the exaggerated postulation of sound ‘laws’ had to suffice as the backbone of Romance studies in the early stages of their development.

The theory and practice of editorial philology found their most determining inspiration in France, thanks to scholars such as Gaston Paris (1839–1903, professor at the École Pratique des Hautes Études, still a prestigious postgraduate institution today), Paul Meyer (1840–1917, professor at the École des Chartes, the para-university college for archive curators), his student Antoine Thomas (1857–1935, professor at Sorbonne University) and Joseph Bédier (1864–1938, who occupied the chair of ‘*Langue et littérature française du Moyen Âge*’ at the *Collège de France*, another para-university institution aimed at a broader audience).

The foundation of the *Société des anciens textes français* in 1875 and the journal *Romania* in 1872 established new standards for philology and editorial practice. In 1890, the new series published by the *Société* welcomed Bédier’s first edition of the *Lai de l’Ombre*, accompanied by an editorial commentary which elucidated the importance of material witnesses and their specific (linguistic) form for the history of texts and language. An entire century would pass before Bédier’s convictions became established in common practice (cf. 3.7.3), but the founding principles had been laid. Paul Meyer’s semi-diplomatic edition of the *Documents linguistiques du Midi de la France* (containing material dating from the 12th to the 16th century) created a model that was equally as essential for the editing of documentary texts (1909, cf. the partial edition in *Romania* 27, 1898, 337–341).

Two other great philologists of this period who should be mentioned here are Adolf Tobler, who prepared the material for the *Altfranzösisches Wörterbuch*, which was continued after his death by his student Erhard Lommatzsch, and the Austrian scholar Adolf Mussafia, whose editions of Old Italian texts are among the most reliable of the 19th century.

Around almost the same time, the work of the linguists Graziadio Isaia Ascoli (Italian) and Jules Gilliéron (Swiss) gave new momentum to dialectology. Ascoli identified the geolinguistic areas of Ladin (cf. *Saggi ladini*, 1873) and Francoprovençal, thereby securing their place within Romance studies. Gilliéron began by writing the *Atlas linguistique du Valais* (1880), the first linguistic atlas of a Romance language, after which he began to collect the material for the monumental *Atlas linguistique de la France* (ALF, cf. 4.2.1). With the founding of the *Revue des dialectes galloromans* (1887) by Gilliéron and his teaching of dialectology at the École Pratique des Hautes Études (1883), followed by the establishment of the chair of *Dialectologie de la Gaule romane* (1894), created especially for him, the discipline of dialectology became firmly anchored within the field of linguistics.

Thus, in 19th-century Paris both general and Romance linguistics witnessed a soaring development, attracting eminent scholars such as Gaston Paris, his protégé Jules Gilliéron, Paul Meyer, Antoine Thomas and Joseph Bédier, as well as the Indo-Eu-

ropean specialist Michel Bréal (1832–1915), considered to be the inventor of modern semantics (*Essai de sémantique*, 1897), his student Antoine Meillet (1866–1936), who was also an Indo-European scholar, and even – from 1881 to 1891 – Ferdinand de Saussure (1857–1913, *cf.* below 2.4.4 no. 3).

The 19th century also witnessed the appearance of important lexicographical works. Synchronic studies, in France, included Émile Littré's *Dictionnaire de la langue française* (typically referred to as '*le Littré*') and Adolphe Hatzfeld's *Dictionnaire Général* (1890–1900). The diachronic dimension was served in Italy by the *Vocabolario universale* (1829–1840), published by the Tramater society, and by Tommaseo-Bellini's *Dizionario* (1865–1879), while in Columbia Rufino José Cuervo inaugurated the *Diccionario de construcción y régimen*, which was completed only in 1998 (*cf.* 9.9.2 no. 4). Grammatical description and thought on grammar were focused on contemporaneous language. The majority of contributions were of minor importance, though some remain of interest – the writings of Andrés Bello (1781–1865) on Spanish, for instance. Moreover, the first histories of national languages began to appear (*cf.* 2.2.4 no. 2 below).

Not all these linguistic currents followed the line of 'Romance' tradition *per se*; their aim was not the pursuit of interpretative research based on the comparison of languages, but rather the intensive elaboration of the languages of the new nations. Their protagonists nevertheless used the same interpretative principles as general and Romance linguists.

The state of the Romance paradigm at the end of the 19th century is illustrated in a remarkably balanced way by the *Grundriss der romanischen Philologie* (volume 1 on linguistics, 1888 [followed by a second – and more elaborated – edition between 1904 and 1906]; volume 2 on literary studies, 1902). Its editor, Gustav Gröber (1844–1911), in his capacity as professor at the (German) university of Strasbourg, also founded the *Zeitschrift für romanische Philologie* (1877), which remains an important point of reference for Romance linguistics today. Even though most of the authors included in both the *Grundriss* and the *Zeitschrift* were scholars of German-speaking origin, these works provide an account of the research carried out in various Romance-speaking countries as well as an overview of the already impressive achievements in Romance studies, which had at the time almost reached their centenary.

2.2.4 The fully-fledged historical-comparative paradigm (1900 to 1930/1950)

The first third of the 20th century brought about major expansions to the historical-comparative paradigm of Romance studies. Fields of study that had already been established remained in place: textual edition and linguistic studies on old literature, as well as descriptive work on the dialectal varieties of individual villages of France, Italy or Spain all flourished. At the same time, notable methodological innovations were accom-

plished in linguistic geography, lexicography (especially etymological) and grammatical description (mainly historical). The paradigm opened up to include contemporary popular varieties, new interpretative approaches such as ‘idealism’, as well as the external history of language. The emergence of structuralism, which would oust the historical-comparative approach in the second half of the 20th century, could at this early stage still be considered a methodological enrichment for the latter.

1 The development of established domains: geolinguistics, etymology and grammar

The beginning of the 20th century was marked by the appearance of the first language atlas to be produced for a national Romance language: the ALF by Jules Gilliéron and Edmond Edmont (1902–1912), mentioned above. This project was imitated, and greatly surpassed in terms of methodological precision, by Karl Jaberg and Jakob Jud’s Italo-Romance *Atlante linguistico ed etnografico dell’Italia e della Svizzera meridionale* (AIS, cf. 4.2.1 no. 2). The work carried out by scholars such as Jakob Jud and Matteo Bartoli in the domain of geolinguistics established a new method for the analysis of the evolution of language across geographical space.

The onomasiological trend, initiated in 1895 (cf. Ernst Tappolet, *Die romanischen Verwandtschaftsnamen*), was closely linked to dialectology. The two subdisciplines merged within the *Wörter und Sachen* (“words and things”) movement, whose object was the semantic motivation of terms relating to rural and material culture (cf. 4.2.1 no. 4 and 9.3.7). It led to the appearance of countless monographs on lexis that used dialect material as a basis for onomasiological classifications relating to subjects such as the names for ‘visual disabilities’, ‘bee’, etc. in a particular Gallo-Romance dialect, or in a particular geographical location (such as ‘the Italian valley of ...’). In accordance with the ideology characteristic of the 19th century, the same types of study were carried out on the vocabulary of medieval literary texts (e.g. on the names of ‘weapons’, ‘greeting formulas’, etc. in the ‘chansons de geste’ or other genres).

Achievements in etymological lexicography were just as noteworthy. A remarkable successor to Friedrich Diez was the Swiss philologist Wilhelm Meyer-Lübke (1861–1936), who reorganised Diez’s dictionary by etyma, considerably enlarging the number of headwords and systematising the treatment of vocabulary. The *Romanisches etymologisches Wörterbuch* (REW, 1935), the work of a Romance scholar with the spirit of an Indo-Europeanist, remains to this day an indispensable source for Ibero-Romance and, to a certain extent, also for Italo-Romance. It is still the only pan-Romance dictionary with a broad scope. In the ensuing years, research concentrated on the main linguistic areas within the *România*. The REW thus benefitted from significant additions for Italo-Romance as a result of etymological research undertaken by Carlo Salvioni (collected by Paolo Faré), while the contributions made by García de Diego for Ibero-Romance were less innovative (cf. 9.9.2 no. 4).

The principal work undertaken in the field of Romance studies during this period was dedicated to Gallo-Romance. The *Französisches Etymologisches Wörterbuch* (1922–2002) by the Swiss lexicographer Walther von Wartburg (1888–1971) provided a synthesis between the diachronic axis of written language – the history of French, Occitan, Gascon and Francoprovençal from their origins until the 20th century – and the synchronic axis, represented by the modern dialects of these four languages.

According to its third successive director, Jean-Pierre Chambon, the FEW, which is an extensive series of individual studies rather than an actual dictionary, became, and remains to this day, an essential and unsurpassed model for lexicography as well as for historical and etymological lexicology. It replaced a form of etymology whose sole objective was to specify the origin of words with a form that was concerned with their ‘etymological history’, and which thus consisted in the attempt to retrace the evolution of a word across time and space.

Grammatical description had not been perfected to the same extent as lexicography, as the tools for grammatical interpretation were not yet as advanced: the quality of Meyer-Lübke’s *Grammatik der romanischen Sprachen* (1890–1902), a revised version of Diez’s work, did not entirely match that of his parallel revision of Diez’s dictionary. Nevertheless, the grammars of Nyrop (1903–1930) and of Damourette and Pichon (*Essai de grammaire de la langue française*, 1911–1949) or the numerous writings of Gustave Guillaume (e.g. *Temps et verbe*, 1929) developed an entirely new basis for grammatical reflection, at least for French. Guillaume’s ‘psychomechanics’ had considerable influence on many French linguists until the mid-20th century, a time, moreover, that was dominated by structuralism *per se* (cf. no. 3 below).

2 More recent issues: diasystematic variation, idealism and external history

In addition to dialects, diastratically marked varieties began to interest some linguists, once again in Paris first of all. Examples of such scholars are Lazare Sainéan (*Les Sources de l’Argot ancien*, 1912), Albert Dauzat (*L’Argot de la guerre*, 1918) and Gaston Esnault (*L’Imagination populaire: métaphores occidentales*, 1925), who were followed by the Swiss Henri Frei (*La Grammaire des fautes*, 1926). Leo Spitzer’s analysis of *Kriegsgefangenenbriefe* (i.e. letters written by prisoners of war), published in 1921, introduced sources written by semiliterate individuals into research on popular language (cf. Carles/Glessgen, *Les écrits des Poilus*, 2020).

Extra-European varieties of Romance languages as well as creole languages began to receive sporadic attention for the first time, mainly from Hugo Schuchardt (*Das Lateinamerikanische Spanisch und der Vokalismus des Vulgärlateins*, 1866–1868; *Beiträge zur Kenntnis des kreolischen Romanisch I. Allgemeineres über das Negerportugiesische*, 1888). These studies, still scarce at the time, were the precursors of research that would become fundamental in the second half of the 20th century (cf. 2.3.2 no. 2 below).

By contrast, a movement that was much more fashionable at the time was idealism, which was opposed to the idea of the infallibility of language change championed by Neogrammarians and instead introduced the history of ideas into historical linguistics (see e.g. Spitzer, *Stilstudien*, 1928; Curtius, *Europäische Literatur und lateinisches Mittelalter*, 1948; Auerbach, *Mimesis*, 1968). The objective of idealism was to identify possible links between the ‘spirit of a language’ and the ‘spirit of a people’ at a given time in history; in other words, language served as a witness to structures underlying thought and culture. Idealism only partially survived during the Fascist period, despite the efforts of its proponents, who were often persecuted by the Nazi regime.

Idealism reflects the difficulties which beset most linguistic currents of that period: although the movement raised pertinent questions, its proponents did not yet possess the scientific tools necessary to provide coherent answers. Although the links between language, thought and society are very close, they do not allow certain shortcuts to be made: the idea of assimilating – as has been done – the use of the future tense in French to a cultural tendency towards precision of expression (in contrast to German, where the future can be expressed by the present tense) is far removed from modern interpretations. Once it had been refounded on the basis of present-day cognitive sciences, however, idealism witnessed the same renewal as onomasiology and historical semantics.

Idealism can be viewed as a particular form of study of the external history of language, which appeared in a more generalised manner at the beginning of the 20th century. Though the term ‘external history’ (as opposed to the ‘internal history’ of language) was coined by Georg von der Gabelentz in 1891 (cf. 10.1.2), the earliest synthesis of linguistic history that treated internal and external aspects of language in parallel did not appear until the beginning of the 20th century, in the form of Ferdinand Brunot’s *Histoire de la langue française* (1905–1939, published posthumously and consisting of 13 tomes in 23 separate volumes). Particularly from the 4th volume – on ‘classical language’ (1660–1715) – onwards, Brunot describes phonetic, grammatical and lexical developments from the point of view of their general evolution, considering language within its various contexts of use and distinguishing its diaphasic varieties (e.g. in vol. III/1, ch. IV: ‘Les mots bas’ (vulgar words) and ch. VI: ‘Les mots de métier’ (words relating to professions)).

The simultaneous consideration of linguistic, literary, political and socio-cultural aspects also characterises the principal work of Ramón Menéndez Pidal (*Orígenes*, 1926; *La España del Cid*, 3 vols., 1929). His impact on the development of linguistic research in Spain was largely superior to that of Brunot in France; throughout the 20th century Spain maintained a philological tradition that was both linguistic and literary, and that partially converged with the tendencies followed by international research.

Numerous studies on the emergence of modern national languages during the medieval period cited, more often than not erroneously, external influences as catalysts for language evolution:

- Leonardo Olschki saw the geopolitical importance of Paris as a cause of the centralisation of the French language (cf. *Der ideale Mittelpunkt Frankreichs im Mittelalter, in Wirklichkeit und Dichtung*, 1913). Though the idea is not entirely incorrect, it is founded on significant inaccuracies and even anachronisms, as Olschki assumes the existence of trends in the 13th century which only appeared one or two centuries later (cf. 10.4.5 (2));
- Walther von Wartburg considered the Germanic *superstrata* as determining factors for the fragmentation of the *România* (cf. *Die Entstehung der romanischen Sprachen*, 1934). The fragility of this hypothesis has since been pointed out (cf. 10.3.3);
- Heinrich Morf emphasised the importance of diocese boundaries for the dialectal fragmentation of the *langue d'oïl* (cf. *Zur sprachlichen Gliederung Frankreichs*, 1911; see also 10.4.2), thereby exaggerating their significance (cf. the later, more prudent study by Jakob Jud, *Sur l'histoire de la terminologie ecclésiastique de la France et de l'Italie*, 1934 [in 1973]; cf. also Jakob Wüest, *La dialectalisation de la Gallo-Romania*, 1979).

The roots of current research on external language history lie in these studies, albeit following a period of around half a century during which methodologies were once again exploited to serve nationalist ideologies.

3 At the periphery of Romance studies: structuralism

The early 20th century also witnessed the appearance of structuralism, which was to provide the backbone for new developments in linguistics throughout the century. The emergence of this important movement was marked by the posthumous publication in 1916 of Ferdinand de Saussure's lectures in general linguistics (between 1906 and 1911) by his two students Charles Bally and Albert Sechehaye (cf. the critical edition by Rudolf Engler, 1967–1968). Saussure introduced the three major dichotomies of *langue-parole*, signifier-signified and synchrony-diachrony into linguistics (cf. 1.2.2, 1.2.3 and 1.2.5). The 20th century would then further clarify and elaborate on Saussure's ideas, which, nevertheless, have lost none of their revolutionary vigour.

Modern structuralism has its roots in the 1930s, though it only fully developed after World War II, in the 1950s and 1960s. The first major innovation after Saussure was the development of phonology by the *Prague linguistic circle* or *Prague school* of linguists, namely by Roman Jakobson (1939, 1949) and Prince Nikolai Trubetzkoy (1939), whose ideas were taken up and elaborated on by André Martinet (1939, 1953, 1970). In parallel, structural semantics was elaborated by Jost Trier (1931), Charles Bally, mentioned above, (1944) and Leo Weisgerber (1949–1950, [1953–1959]), whose works also contributed significantly to later Romance studies (cf. 9.2.3).

It was towards the end of the same period that Lucien Tesnière (1893–1954) wrote his pioneering work *Éléments de syntaxe structurale* (published posthumously in 1959), based on the idea of a dependency grammar according to which the elements of the

phrase are governed by the verb. This important synthesis on syntax thus marked the conclusion of European structuralism's transformation of the linguistic research paradigms in the three major domains of language.

The most important representative of structuralism specialising in Romance linguistics was Eugenio Coseriu (1921–2002), who was of Romanian origin and who was first active in Uruguay before founding his own linguistic school in Tübingen. Coseriu brought more transparency to structuralist analysis, and to linguistic analysis in general, thereby contributing to a general consolidation of linguistic thought during the second half of the 20th century (cf. Coseriu 1956, 1966, 1981, 1988, etc.).

Other contemporary structuralist tendencies played a less determining role in the development of Romance linguistics. These included:

- ‘Glossematics’ and the Copenhagen linguistic circle (1931; cf. Louis Hjelmslev, *Omkring sprogteoriens grundlæggelse*, 1943 [Engl. *Prolegomena to a theory of language*, 1953; Fr. *Prolégomènes à une théorie du langage*, 1966]). The Copenhagen school of linguistics included several influential Romance linguists in the decades that followed (especially Knud Togeby, *Structure immanente de la langue française*, 1965 and Emilio Alarcos Llorach, *Gramática estructural*, 1969).
- American Structuralism (initiated by Edward Sapir, 1921 and Leonard Bloomfield, 1933), often referred to as ‘distributionalism’ (cf. also Franz Boas and Charles Hockett). This movement was detached from European structuralism, mainly owing to its particular conception of semantics, as well as to its concentration on Native American languages without a written tradition.
- British Structuralism (cf. e.g. John R. Firth, [1934], 1951). The work of Michael A. K. Halliday was particularly important for variational linguistics (1956, 1961, 1966, 1970).

One of the principal strengths of structuralism was that it rigorously considered language from an overall perspective, and aimed to understand and describe what it assumed to be a coherent system regulating all individual linguistic phenomena. Later studies, especially in the field of cognitive sciences, would show that the cohesion of language as a whole is so complex that a conception of it as a ‘system’ designed according to clear principles can convey only a simplistic image. Nevertheless, the fact that this fundamental current focused on individual elements as part of language viewed as a global framework paved the way for an adequate assessment of the autonomy exhibited by the different linguistic modules, as well as of the numerous interdependencies and interactions between them in speech.

The only major criticism which the present-day linguist could level at structuralism is that it shifted the focus of linguistic study to language in its synchronic and contemporaneous state while largely neglecting its historical dimension and its high degree of intrinsic variance. Thus, the synchrony-diachrony dichotomy – a concept that is in itself neutral and judicious – was reinterpreted in the vein of a linguistic ideology that

would ultimately sever the dialogue between (traditionalist) historians and (innovative) ‘synchronists’. Generativist trends continued to follow this line, which would only be interrupted towards the end of the century by variationist and cognitive currents, the proponents of which managed to rescue the valuable concept of diachrony at the last minute, before its interpretative methods were irretrievably lost.

The ‘synchronic turn’ and the accompanying syntactocentrism were particularly damaging for the discipline of Romance studies, the very definition of which relies on a consideration of the history and the dialectal variance of its languages. Whereas in the first third of the 20th century Romance studies actively contributed to important innovations taking place in linguistics, in the ensuing years the discipline was subject to strong centrifugal forces (*cf.* below 2.3.4)

4 Romance studies in the first half of the 20th century: a summary

The evolution that took place in linguistics between 1880 and 1930 bears witness to a phase of rich intellectual activity. The fact that we have chosen to focus on the most significant traditions means that the multiple interactions among the central figures involved and the chronological overlap between different developments tend to be overlooked; the same applies to the immense gulf that existed between the highly intellectual activity of a number of erudite scholars and the much more limited scope of university education that students received at the time (*cf.* 12.1).

The years between 1930 and 1950 were a period of stagnation, less in terms of actual production (*cf.* the bibliographical overview by Kuhn, *Romanische Philologie*, 1951) than in terms of the intellectual development of the traditions that had been established within Romance studies. The Fascist regimes and their repercussions in Europe, the Spanish Civil War and the mass bloodshed caused by the Holocaust and World War II led to a break in these traditions, and the post-war era opened on a changed world – for linguistics, too.

2.3 The development of the present-day paradigms

2.3.1 New methodologies in the Romance domain: sociolinguistics, pragmatics and applied linguistics

In the 1950s and 1960s, Romance linguistics broadened and underwent significant internal consolidation. At the same time, new lines of tradition developed, with an intense focus on general linguistics that led to a regrettable divide between the research and teaching paradigms.

Of the new methods, sociolinguistics proved to be closer to the central preoccupations of Romance studies than the systemic approaches to language. This field corresponded to the former study of diastatic variation. It experienced significant growth

in America as an empirical form of contact linguistics (cf. Uriel Weinreich, *Languages in contact*, 1953), which drew on observations pertaining to the linguistic diasystem (cf. William Labov, *The social stratification of English in New York City*, 1966; Basil Bernstein, *Class, codes and control I*, 1971). As a result of these new tendencies, the attention previously given to language as a system structured by functional oppositions was refocused on its variable character and on factors linked to the individual user.

As a consequence, variational linguistics saw a particularly intensive phase of development within Romance studies, first as part of a synchronic approach (cf. 4); it subsequently also came to be used in language history (cf. 10, 10.7). Many large-scale methodological innovations in this area came from researchers in Romance studies (these include the distance-immediacy continuum, the study of diaphasic varieties and technical language, lay linguistics and linguistic culture, and the evolution of external linguistic history and variationist philology).

The discipline of pragmatics had already been firmly established in the first half of the 20th century by the works of Charles S. Peirce (*The 1903 Harvard lectures on pragmatism*, 1903) and Charles Morris (*Foundations of the theory of signs*, 1938). In the 1960s and 1970s, however, decisive theoretical contributions came from the philosophers of language Ludwig Wittgenstein (1958) and H. Paul Grice (1968, 1975, 1991). The formulation of Grice's cooperative principle in the act of communication led to particularly significant innovations in the field of text linguistics, as well as in grammatical analysis and the study of language acts (cf. 4.3). Romance scholars have remained active in all these domains, and the potential they hold for philological and diachronic study has expanded over the course of recent decades.

The development of 'applied' linguistics is further removed from the research topics that are of immediate relevance to Romance linguistics. Nevertheless, numerous Romance scholars have made important contributions to the fields of language learning and language acquisition, translation studies and digital humanities.

2.3.2 Generative linguistics and the 'cognitive turn': creole studies, neuro- and psycholinguistics

1 Generative linguistics

The structuralist renewal, which had reached its peak in the 1950s and 1960s was subsequently left behind by the proponents of generativism. The desire of a generation marked by the Second World War to break with the past was reflected in the soaring success of Noam Chomsky and his theory of generativist-transformationalist grammar (cf. *Syntactic Structures*, 1957; *Aspects of the Theory of Syntax*, 1965).

This approach regards grammar as the central and essential part of language – Chomsky equates it with the 'faculty of language in the narrow sense'. Chomsky's gen-

erativism takes the form of various models of grammar which constitute revisions of his earlier theories, in particular that of *Government and Binding* (GB) 1981 (based on the *Principles and Parameters* model of language) and the *Minimalist Program* (MP), established in 1993, which he still considers to be valid. Examples of concurrent developments which took place in the late 1970s are *Generalized Phrase Structure Grammar* (GPSG) and *Lexical-Functional Grammar* (LFG).

After some initial reluctance, generativism became prevalent in studies focusing on Romance languages during the 1990s, at least in the field of grammar. The strong Romance traditions in lexicology (including onomastics and lexicography), variational linguistics (as regards both contemporaneous and diachronic approaches), sociolinguistics, external history and philology, by their very nature, remain beyond the scope of this approach. Although the complementary nature of Romance and generative linguistics is unquestionable and should not pose any problems, the latter tends to perceive and present all other approaches as peripheral or outside the scope of linguistics. The notion that one single theory could suffice to explain every aspect of language and its use has therefore led to a strong and regrettable syntactocentrism that was already inherent in modern structuralism and which will hopefully be overcome in the course of the following decades.

→ Pescarini, *The reception of generativism in Romance linguistics*, OxfEnc
Koch, *Es gibt keine Konstruktionsbedeutung ohne Bedeutungswandel*, 2012

2 Creoles and sign languages

Finally, one of the most striking innovations in the history of linguistics since its emergence at the beginning of the 19th century was the development of creole studies in the 1970s. The discovery of ‘approximative systems’ of communication and the concept of ‘interlanguages’ provided a new foundation for research on language universals. As well as contributing to a better understanding of phenomena resulting from language contact, creole studies paved the way for the elaboration of a functional form of cognitive linguistics.

Unlike structuralism and generativism, which have a mathematical-logical basis, cognitive studies are rooted in biology and anthropology. They therefore allow a type of synthetic analysis that is compatible with the Romance historical-comparative paradigm, and that is currently in the process of becoming established (cf. 6).

Creole languages are the product of extreme cases of language contact (cf. 3.6.3 for Romance-based creoles). They emerged in the context of European colonisation, as a result of attempts to establish communication in situations characterised by extreme levels of domination and linguistic distance. In each case, a first generation of non-native speakers developed a pidgin language, which was followed by the emergence of the creole itself amongst the native speakers of the subsequent generation. In con-

trast to pidgins, creoles are fully-fledged linguistic systems. From a linguistic point of view, the most remarkable aspect of such languages, generated almost *ex nihilo*, is that while their vocabulary is largely borrowed from a given dominating language (English, French, Arabic, etc.), their grammatical features are quite homogeneous: creoles are all strikingly similar as far as their grammar is concerned, even though they often arose independently of one another, in the context of different dominating and dominated languages. This suggests that creoles are based on certain principles that are shared by all human beings and that are consistent with the anthropological conditions which govern the language-producing faculty.

One of the most noticeable elements of creole languages is their verb marking, whereby invariable verb forms are preceded by three distinct markers for tense, aspect and mood (= TAM). These markers generally occur in the same order, though they may also be omitted (= 'zero' marking [cf. 8.3.2]). This particular structure consisting of TAM markers followed by a verb highlights the importance of the predicate within the proposition (cf. 8.1) and allows a better understanding of the essential factors that anchor the verbal element in its context. Other characteristics of creoles include a rigid constituent order (cf. 8.4.3), a reduced number of derivational morphemes and lexemes (consequently often polysemic), and the absence of a verb expressing the meaning 'to be'.

Creoles display several characteristics that strongly resemble those of sign languages. Both these communication systems develop under atypical circumstances and each one provides a different – and therefore complementary – type of information relating to human cognition and linguistic principles. The existence of creoles and sign languages illustrates the innate capacity of individuals to create a language (90% of hearing-impaired individuals have hearing parents and do not acquire an established form of sign language from them). This capacity is only lost in the total absence of other humans during the first years of a person's life (as in the case of 'feral children')⁵.

In sign languages, gestures (three-dimensional in nature) allow the expression of parallel or even hierarchical linguistic information. Thus, a forward hand movement incorporating a turn simultaneously encodes two different types of information. Nevertheless, these gestures should probably be classified as surface phenomena; in their basic principles, sign languages adhere to the constraint of linearity characteristic of typical linguistic representation (cf. 1.2.4).

⁵ The moving account provided by Susan Schaller in *A man without words* (1991) tells the true story of a deaf man who forges a system of linguistic expression with his brother, who is also deaf. In doing so, he learns the basic principles of communication, but at the same time creates an almost insurmountable barrier to other systems of communication, including conventional forms of sign language.

3 Neuro- and psycholinguistics

Psychological studies focusing on brain functioning provide linguistics with a further important tool for analysis. Empirical psychology and neurology employ various methodologies and have contributed a number of important insights into linguistics. The disciplines are highly developed and some of the issues they deal with overlap partially, and sometimes even fully, with questions that are of relevance to linguistics. Research focusing on neurological or physiological activity of the brain emerged as early as the second half of the 19th century. The first area of study that had a strong impact on linguistics involved the analysis of brain injuries. The syndromes of ‘aphasia’ and ‘dysphasia’ involve serious linguistic deficiencies caused by injuries located in the brain, which do not necessarily impede the injured individual’s other cognitive competences. Neuro- and psycholinguistic research has, however, broadened significantly since the 1980s, mainly owing to the development of information technology and to advances in analytical tools such as electroencephalography (EEG) and magnetic resonance imaging (MRI). Though contributions from this field may be sporadic, in certain areas they prove fundamental.

Research in the field of neurolinguistics is based directly on physiological analyses of brain activity, which have allowed a more accurate localisation of the fundamental domains of language – phonology, morphology, syntax and lexis – within the brain (*cf.* 5.1.2).

Psycholinguistic studies have contributed additional descriptive and interpretative data. A large number of recognition and reaction tests have allowed various aspects of the brain’s handling of language to be specified.

Within this context, research on language acquisition constitutes a further specific field of study founded on psychology and highly relevant for linguistic analysis.

The integration of results from the fields of neuro- and psycholinguistic research within linguistics is far from trivial – as the author has experienced firsthand for the purposes of a study on the status of the ‘word’ in the brain and in language (Glessgen, *Le statut épistémologique du lexème*, 2011) –, not least because psychological research only makes very partial use of work in linguistics. The interface between the two disciplines therefore still holds unexplored potential for innovation.

→ Bambini/Canal, *Neurolinguistic research on the Romance languages*, OxfEnc

Marzi/Pirrelli, *Psycholinguistic research on inflectional morphology in the Romance languages*, OxfEnc

2.3.3 *Ex traditione innovatio*: typology and diachrony

The comparison of languages – both closely and distantly related ones – remains to this day the primary basis of all interpretative linguistic research. The traditional comparative approaches of the 19th and 20th centuries are largely congruent with the

more modern concept of linguistic ‘typology’ (which, however, was first introduced by von der Gabelentz in 1891). Typology tends to favour the study of highly divergent languages, since parallels drawn among them easily reveal general features of language, as opposed to historically dependent aspects. Nevertheless, the study of a single language family remains an equally useful part of language comparison and typology, as it yields complementary results. In Romance linguistics, the typological approach is by its very nature not only supported but greatly enriched by data stemming from diachronic and dialectological investigations.

The diachronic approach in and of itself represents a second and no less important method of analysis when it comes to achieving a better understanding of the functioning of language. As far as the languages of the *Romània* are concerned, studies focusing on a diachronic perspective can prove even more rewarding than research undertaken in a purely synchronic perspective. More than a millennium of dialectal variance within a whole range of languages with a vast amount of available historical documentation creates a major advantage for observation and evaluation. It allows researchers to observe the progression of linguistic changes in detail, and thus to study the variability of language. Considered in a more general context, the study of this documentation illuminates the textual and historical heritage of a large part of Western Europe and the New World. The tiny segment of the world covered by the Neo-Latin languages therefore provides infinite research opportunities.

The typological and diachronic approaches are essential for research in linguistics, and this explains why the historical-comparative Romance paradigm has lost none of its importance. Even if one is only interested in a specific question concerning morphology in standard Peninsular Spanish, it is useful to take into account other Hispanic varieties, other Romance languages and different historical stages since Late Latin. Unaffected by the recent focus on the systemic analysis of grammar mentioned above, research over the past few decades has yielded considerable methodological innovations in all areas of the classic paradigm of Romance studies, which thus has a much greater reach today. Much of the future potential of Romance studies therefore lies in the innovative reinterpretation of its earlier traditions and achievements.

2.3.4 The current ‘centrifugal’ paradigm

As has been shown, linguistics has experienced fundamental innovations in recent decades. It is not easy to find one’s way through the heterogeneous universe that the discipline has become. Despite the restricted applicability of linguistics, this issue affects the most diverse areas of academics. Linguistics as a domain has become impossible for a single individual to master, including professionals with an academic background and many years of study behind them. But even at a more abstract level, each subdiscipline of linguistics is faced with the great difficulty of coping with the ‘centrifugal’ trends currently dominating the field. These pose an even greater challenge for students and

young academics: depending on their teachers and their country of education, students often find themselves confronted with forms of linguistics that seem to belong to distant and impenetrable worlds.

It is therefore essential to reflect on ways in which cohesion can be maintained within a discipline such as Romance studies, and, in doing so, to consider its relationship with other fields of linguistics. In the following pages, we will attempt to illustrate the particular characteristics and strengths of Romance studies, which lie in its multiple focus on variational linguistics (*cf.* part 4 of this manual), internal language change (part 5), and external history and philology (part 10). Ours is a holistic approach that takes into account the interdependencies between the individual topics while at the same time attempting to explain the significance of each within the overall architecture of the discipline.

2.4 Reference works for the study of Romance linguistics

2.4.1 Manuals and encyclopedic reference works

Several important encyclopedic works relating to both general and Romance linguistics have appeared in recent years.

The best general overview of Romance linguistics remains that provided by the *Lexikon der Romanistischen Linguistik* (LRL), published in Tübingen by Niemeyer, consisting of eight sections (corresponding to twelve volumes, 1988–2005), edited by Günter Holtus, Michael Metzeltin and Christian Schmitt. The first two volumes, 1/1 and 1/2, offer a comprehensive overview of the discipline, its methodology and its history. The following two volumes, 2/1 and 2/2, address questions relating to the emergence of the Romance languages and their medieval stages. Tomes 3–6 (in six volumes) treat each modern Romance language individually, 7 concentrates on language contact and typology, and the extensive encyclopedic work is completed by a final volume containing indices and bibliographical information. Although the majority of articles are in German, the LRL also contains numerous articles written in Romance languages. The work is well structured and very rich; despite its volume, however, the treatment of certain aspects of diachronic linguistics is inadequate.

A second substantial handbook of Romance linguistics – *Romanische Sprachgeschichte* (RSG), published in Berlin by Mouton de Gruyter, in three large volumes (2003–2008), edited by Gerhard Ernst, Martin Glessgen, Christian Schmitt and Wolfgang Schweickard – was conceived as a complementary perspective to that of the LRL, with the aim of emphasising diachronic aspects. Over half of the articles in the RSG are written in French, Italian or Spanish, while the remainder are in German. The first volume addresses general issues relating to linguistic history, providing an outline of the external history of the various Romance languages both within and outside Europe. The second volume focuses on various contexts of language use (such as religious, literary

or economic language) and on external factors that influence language change. The third volume contains an overview of the internal history of the Romance languages (though it concentrates primarily on dominant varieties), in addition to an index.

The *Cambridge history of the Romance languages* (CambrHist), published by Cambridge University Press in 2 volumes between 2011 and 2013 and edited by Martin Maiden, John C. Smith and Adam Ledgeway, offers a different view on certain topics relating to diachrony. The first volume treats the internal history of the Romance languages, the second, their external history. Each chapter assumes a pan-Romance perspective (in contrast to the RSG, which describes each language separately), emphasising particularly salient features (whereas the LRL and RSG aim to provide a more general presentation). This encyclopedia is thus a useful complementary work, which further includes elements of a generativist approach to syntax.

This was followed by the *The Oxford Guide to the Romance languages* (OxfGuide) in 2016, published by Oxford University Press in a single volume consisting of 1200 pages, also edited by Adam Ledgeway and Martin Maiden. Its 64 chapters, each written by a different author, provide a brief presentation of the individual Romance languages and dialects (art. 8–24) and an analysis of very specific grammatical topics (art. 4–6, 27–31, 34, 42–64), as well as of a certain number of phonological/phonetic and grapho-phonetic phenomena (art. 25–26, 38–41). The great quality of this work resides in its consideration of dialectal evidence. External and even internal history remain peripheral, as the book was conceived as a complement to the CambrHist. Neither lexis nor variational and sociolinguistics receive much attention (*cf.* 2.4.2 below for the CambrHandb).

The *Oxford Encyclopedia* (OxfEnc), published under the direction of Michele Loporcaro and – during its first years – Francesco Gardani, will eventually cover a larger field. Its orientation is similar to that of the OxfGuide, with slightly more attention being paid to lexis and touching on a wider variety of topics. A print version would correspond to five volumes. Currently (June 2024), 142 of the 170 articles planned are accessible online.

Finally, an overarching series of more than 60 *Manuals of Romance Linguistics* (MRL), describing every Romance language individually as well as specific aspects of linguistic methodology, is in progress, under the direction of Günter Holtus and Fernando Sánchez-Miret (de Gruyter). The unstated aim of this collection is to produce an enlarged version of the LRL, judging by the content of the 36 volumes published between 2014 and 2024.

In addition to these works focusing specifically on Romance linguistics, the collection of *Handbücher der Sprach- und Kommunikationswissenschaft* (HSK) includes separate tomes on sociolinguistics, writing, lexicography, word-formation and typology (LangTyp), most of which are contained within two volumes. Numerous Romance scholars have contributed to the work, which contains many articles written in English.

The objective of the present *Companion* is not to replace these high-calibre works, but rather to provide an overview of the principal topics of Romance linguistics, thereby

rendering the specific and learned articles of the LRL, RSG, CambrHist, OxfGuide, CambrHandb (cf. 2.4.2), OxfEnc and MRL more accessible and facilitating their comprehension, since a structured overview of the discipline and a thorough understanding of the basics are necessary to reap the full benefits of the material contained within these specialised reference works.

In order to facilitate access to the major encyclopedias, references to the articles and chapters throughout this manual are provided in shortened form. The author's name and the title of the article (the latter in *italics*) are followed by the source and, where applicable, the number of the article ('art.') or chapter, as exemplified below:

- Schweickard, *Panorama des revues romanes*, LRL 1/2, art. 90c
- Koch/Oesterreicher, *Comparaison historique de l'architecture des langues romanes*, RSG 3, art. 220
- Andreose/Renzi, *Geography and distribution of the Romance languages in Europe*, CambrHist 2, 8 [= volume 2, chapter 8]
- Cruschina/Ledgeway, *The structure of the clause*, OxfGuide 31 [= chapter 31]
- Gardani, *Contact and borrowing*, CambrHandb 28
- Rainer, *The language of economy and business in the Romance languages*, OxfEnc

The individual articles are omitted from the final bibliography, since the works to which they belong are to be considered as general and indispensable references.

The LRL represents a special case: its 580 articles in 12 volumes, amounting to almost 10,000 pages, offer such an intricate and detailed view of the field that systematic cross-references would only serve as a hindrance to the reader. The most efficient and appropriate way to use this manual is to acquire a thorough knowledge of the structure of each of its volumes (cf. the complete table of contents in vol. 8, p. 13–32), and to consult the volumes which treat overarching topics in conjunction with those dealing with the individual languages, according to the nature of the research question. The complete bibliography, organised by language and topic and containing references to other works (vol. 8, p. 101–356), is particularly useful, as is the detailed subject index for the LRL itself (vol. 8, p. 383–635). The entries of the latter appear in a variety of languages, German being the most frequent. The cross-references to the LRL provided in this *Companion* are therefore not systematic, in contrast to those made to the other reference works.

2.4.2 Shorter manuals and textbooks

The present *Companion* takes its place among more than 90 Romance linguistics manuals and textbooks that have appeared since that of Lorenz Diefenbach in 1831 and that assume a more directly didactic perspective (see the author's comparative study: Glessgen, *Les manuels de linguistique romane*, 2000). Its aim is to provide a broad coverage of the various topics which make up Romance linguistics by means of a comparative, diachronic, variational and philological approach.

A number of introductory works have been written in English; these, however, are mostly outdated and do not cover all the salient aspects of Romance linguistics (the full citations are provided in the bibliography, 13.2):

- William D. Elcock, *The Romance Languages*, 1960, 21975: less an account of the content than a description of the state of the discipline and the way it was taught in the 1960s;
- Mario Pei, *The Story of Latin and the Romance Languages*, 1976: essentially a very well written and informed external history of the Romance languages which remains useful, as the subject is often neglected and as the style of the book is quasi-journalistic;
- Martin Harris / Nigel Vincent (eds), *The Romance Languages*, 1988: the first attempt at a modern approach to the Romance languages from a grammatical point of view; has since been replaced by *CambrHandb* and *OxfGuide*;
- Rebecca Posner, *The Romance Languages*, 1996: a synthetic view of internal history; remains of interest due to its holistic approach, but requires cautious consideration of both empirical data and general conclusions;
- Adam Ledgeway / Martin Maiden (eds), *The Cambridge Handbook of Romance Linguistics*, 2022 [= *CambrHandb*]: the most recent and most helpful overview, consisting of 940 pages written by 50 different authors, and representing a cross between a textbook and an encyclopedia highlighting the contribution of Romance linguistics to general linguistics; a useful complement to the *OxfGuide*. As in the case of the latter, the emphasis is on grammar and dialects, although the *CambrHandb* devotes more space to phonetics. No general description of the individual languages is provided. Lexis, diachrony and philology are not treated in depth, and four of its chapters are devoted to variation (relating to language, society and the individual).

For a more comprehensive viewpoint, the reader is obliged to consult works in German, Italian, French, and Spanish. The following two works are half a century old but they are still well worth reading (Tagliavini), studying (Bec), or consulting (Lausberg):

- Carlo Tagliavini, *Le Origini delle lingue neolatine*, 1982 (7th ed.)
- Pierre Bec, *Manuel pratique de philologie romane*, 1970–1971
- Heinrich Lausberg, *Romanische Sprachwissenschaft*, 1967–1972

Carlo Tagliavini's manual, which forms the basis for the conception of the present *Companion*, remains an indispensable complement. In addition, the two-volume work by Pierre Bec provides linguistic analyses of various Romance texts and is one of the most useful complementary approaches. Finally, for detailed information on historical phonetics and inflectional morphology in the Romance languages and dialects, Lausberg's opus – in small format but nonetheless very dense – remains a reliable source that has not yet been surpassed.

Three more recent and well-conceived introductions with a grammatical focus and based on a typological approach merit recommendation, the first longer, the other two (considerably) shorter:

- Lorenzo Renzi / Giampaolo Salvi, *Nuova introduzione alla filologia romanza*, 1994²
- Lorenzo Renzi / Alvise Andreose, *Manuale di linguistica e filologia romanza*, 2015⁴
- Georg Bossong, *Die romanischen Sprachen*, 2008

Other rather short volumes belonging to the Italian tradition have a philological focus:

Alberto Varvaro, *Linguistica romanza*, 2001 (fr. *Linguistique romane*, 2010)

Lino Leonardi / Laura Minervini / Eugenio Burgio, *Filologia romanza. Critica del testo, linguistica, analisi letteraria*, 2022 [the section on linguistics is by Minervini]

Marcello Barbato, *Le lingue romanze. Profilo storico-comparativo*, 2017

Pietro Beltrami, *La filologia romanza. Profilo linguistico e letterario*, 2017

The following two handbooks provide complementary insights into the history of research in the field:

José Enrique Gargallo Gil / Maria-Reina Bastardas (eds), *Manual de lingüística románica*, 2007

Hans-Martin Gauger / Wulf Oesterreicher / Rudolf Windisch, *Einführung in die romanische Sprachwissenschaft*, 1981 (very interesting, but written in German, like Bossong's)

Furthermore, there are numerous manuals and introductions dedicated to an individual Romance language or to a specific subject, such as the excellent *Profilo linguistico dei dialetti italiani* by Michele Loporcaro (2013) or our own 'Guide du FEW' (FEWGuide, 2019), which provides a detailed introduction to research in the field of diachronic Gallo-Romance lexicology as well as to the available resources.

The traditional importance of publications in German in the domain of Romance linguistics has already been mentioned in this brief overview, both as regards research on individual languages and the comparison of several languages, a perspective characteristically adopted by non-Romance-speaking authors. English was only recently introduced, but it is becoming increasingly dominant in particular areas, particularly those of grammatical research and synchronic linguistics. The use of French, Italian and Spanish, on the other hand, has remained relatively stable, and these languages are considered to be universally understood within the field of Romance linguistics. The remaining Romance languages are generally only used in writing by specialists in the relevant disciplines. The ability to understand literature written in the various Romance languages, as well as in English and German, thus remains a prerequisite for the in-depth study of Romance.

2.4.3 Ongoing bibliographies providing access to specialised monographs and articles

Handbooks and introductory manuals are only one type of resource available for the discipline of Romance linguistics. Thematic monographs and individual articles published in journals or collective works provide the most stimulating reading. A selection of these works will be cited throughout the present manual to illustrate milestones in research on the various topics covered. The contribution made by Internet resources to research is also on the increase, although the sheer volume of available data severely hinders any attempt at systematic use.

The goal of the *Companion*, however, is not to provide a bibliography of the type offered by the large manuals, by bibliographical publications such as the *Bibliographie sélective de linguistique française et romane* by Willy Bal (1997), or, more particularly, by the following ongoing bibliographies, nowadays all available electronically:

- the *Romanische Bibliographie* (= *Romance Bibliography*, RB, 1878–) is the most important bibliographical tool for Romance studies. It first appeared as a supplement to the *Zeitschrift für romanische Philologie* (1878–1960/64) before becoming independent. Published biennially for many years, it currently appears annually. Each issue consists of three volumes, reduced to two since 1998, containing an index, a list of abbreviations and sections for linguistics and literary studies (with the exception of French literary studies, which are treated in Klapp's *Bibliographie der französischen Literaturwissenschaft*). Organised by themes, it also includes reviews of monographic works published in scientific journals.
- The *Bibliography of linguistic literature* (= *Bibliographie linguistischer Literatur*, BLL, 1975–), published in Frankfurt am Main by Klostermann (online version: BLLDB), is an indispensable complement to the RB. It does not include monographs, but focuses exclusively on articles published in journals and collective volumes. It is thus a more comprehensive resource for the latter than the RB.
- *The Modern Language Association of America (MLA) International Bibliography*, New York, MLA, (1921–) is useful for publications within Anglo-American circles in particular, as well as for a number of specific areas (e.g. research on Romansh).
- *Linguistics and Language Behavior Abstracts* (LLBA, 1973–) is an extensive and up-to-date online resource, which concentrates on publications in general linguistics.

There are other ongoing bibliographies, mostly general, such as the *Bibliographie linguistique* (useful for work on minority languages) or the *Year's Work of Modern Language Studies* (which offers critical insights from current research). In comparison to the four bibliographies presented above, however, these works do not provide much new information about the field of Romance studies. Instead, more specialised bibliographies should be consulted, such as, for French, the *Bulletin annuel de linguistique française* or *Francis*.

An interesting partial though complementary means of access to Romance studies is provided by the free website *Online Contents Linguistik*, which offers a computerised list of a large number of tables of contents from linguistic journals. It encompasses 400 journals published from 1993 onwards and allows recent publications to be found easily.

An additional tool, which is most useful for locating monographs, is the online catalogue of all major libraries of the Western world, accessible through the website of Karlsruhe university library (cf. *Karlsruher Virtueller Katalog* (KVK): <<https://kvk.bibliothek.kit.edu>>; cf. also 12.1.4 for journals in Romance studies).

Part 2: **Romance languages and varieties today**

3 Presentation of the Romance languages

3.1 Terminology: languages, dialects and varieties in the *România*

The first question that must be asked, before any attempt is made to describe the various Romance languages, is: how should the term ‘language’ be defined in such a context? French is unquestionably a language. When speaking of the French language, one cannot, however, simply disregard the multitude of French ‘dialects’ (also known as the ‘*oïl* dialects’) spoken throughout northern France and Belgium until the First World War. It is no less necessary to specify that what is termed ‘French’ refers not only to a specific ‘standardised language’, but more precisely to a number of ‘varieties’ – standard and non-standard – which are in use in various European and extra-European countries and which differ slightly from one another in terms of their linguistic features. Things become even more complicated if one includes additional neighbouring languages and varieties such as Francoprovençal and Occitan in the discussion.

These facts call for more detailed terminological explanations: the term ‘language’ refers to a variety that ideally has a well-established identity, recognised by its speakers and based on internal linguistic features that distinguish it from all other languages of the world. In reality, however, it is sometimes applied to varieties that do not entirely fulfil these conditions. A ‘language’ must also necessarily be defined in relation to a ‘dialect’: a dialect is essentially a subvariety of a language, spoken within a geographically circumscribed area (‘regional dialect’⁶) or by specific social groups (‘sociolect’). In cases such as French, Italian or Spanish, the language coexists as a standardised variety with numerous regional, dialectal and sociolectal varieties.

To be more precise, the standard French variety and the *oïl* dialects together make up the language (or the ‘historical language’, in Coseriu’s terminology) properly called *French*, which is distinct from the Occitan dialects spoken within the same area as standard French. The standard Italian variety and the dialects of central Italy correspond to the language known as *Italian*, while the dialects of the southern or northern regions of Italy are as different from this language as Francoprovençal, Occitan or Gascon are from French. Thus, knowledge of French or Italian is no guarantee of being able to understand these varieties.

There are also languages which exist solely in the form of a dialect group, such as Francoprovençal. Francoprovençal has not undergone a process of standardisation, nor does it display a great deal of sociological diversification. Such cases are rather an exception in the *România* with its strong written tradition, but they are very common in other language families of the world.

⁶ The French term *patois* is synonymous with that of the dialect as a spatially defined entity. Nevertheless, it is often considered pejorative and/or archaic.

Finally, ‘variety’ is simply a neutral term, used to refer to standard and non-standard languages, regional dialects (again, defined in geographical, not socio-cultural terms), sociolects (e.g. when speaking of ‘popular’ or ‘standard’ varieties) and specialised registers of a given language (such as ‘medical jargon’) alike.

In summary, a language always comprises internal varieties. If it has not been standardised, it consists essentially of dialectal varieties. If it has undergone a process of elaboration (German *Ausbau*, cf. below), it exists as a standard variety and inevitably includes sociolects and registers, and usually also regional dialects.

These definitions can be further refined using the tools of linguistics. The identification of a variety as an entity in its own right and the evaluation of its status as a ‘language’ or ‘dialect’ are founded on two types of criteria: (1) language-internal criteria, based above all on phonetic, grammatical or lexical characteristics, and (2) language-external criteria, which take into account the degree of elaboration undergone by the varieties in question.

1. Language-internal criteria (or structural criteria) concern the degree of linguistic difference between a given variety and other existing varieties. This is a relative parameter, which establishes the distance between a specific variety and all other varieties in existence today, and it takes into consideration the notion of the intelligibility of this variety to speakers of any other variety. As an example, persons who have not learned French will not be capable of spontaneously understanding this language when hearing it spoken, even if they are a native speaker of another Romance language such as Italian or Portuguese. The French language thus displays significant linguistic distance (German *Abstand*) from other languages, even genealogically closely related ones. Hence, French can be considered a ‘language by distance’ (German *Abstandsprache*, after the terminology introduced by German linguist Heinz Kloss in 1952).

2. The language-external criterion of linguistic elaboration comprises several factors related to the role that a language plays in society. These include

- a) its use in writing, a rich textual tradition and, in some cases, elaborated scientific terminology;
- b) an awareness among speakers, who view their speech variety as a language in its own right and use it in a wide variety of contexts (e.g. in familiar and professional situations, in the media, etc.);
- c) an officially recognised status and possibly political importance.

A ‘language by elaboration’ (German *Ausbausprache*, Kloss), such as French, Italian or Portuguese, exhibits all of these features. Various other criteria have been suggested as a basis for language classification; however, the criteria of distance and elaboration are

often present in one form or another. They correspond, for instance, to the terms ‘autonomy’ and ‘standardisation’ used by Stewart (*A sociolinguistic typology*, 1968: 534–537).

The use of these two criteria helps clarify many linguistic situations. Some speech varieties do indeed differ considerably from even the nearest neighbouring forms on an internal level, while at the same time looking back on a tradition of elaboration. For other varieties, however, one of the two criteria, if not both, is fulfilled to a lesser degree or completely absent. Intermediate cases are common and varied within the *România*, as will be shown in the following chapters.

As an example, one can consider the dialects of Italy, which display marked internal differences, and most of which were put into writing to some extent from the Middle Ages onwards. Speakers of the Sicilian dialect will always have experienced difficulties in understanding the Venetian dialect (and prior to the formation of the Italian nation and the diffusion of standard Italian they had no common language). However, they most likely understand the neighbouring Calabrian dialects. Similarly, speakers of Calabrese understand Neapolitan dialect; moving up the Italian peninsula, the example can be repeated for the remainder of the dialects and their direct neighbours. From this perspective, Venetian and Sicilian clearly comply with the criterion of linguistic distance while remaining connected within the geolinguistic continuum via other intermediary dialect varieties. The existence of such continua with no clear dividing lines makes it difficult to identify distinct ‘languages’, despite the fact that the dialects concerned display a high degree of internal variation.

Based on these two criteria, then, varieties in the *România* can be divided into the following categories:

- a) languages by distance and by elaboration, such as French, Italian or Romanian;
- b) languages by distance without systematic elaboration, such as the example of Francoprovençal, discussed above;
- c) numerous regional dialects that have not undergone systematic elaboration and that, despite having distinct internal characteristics in comparison with other varieties in their proximity (such as a standard language or other dialects), do not qualify as languages by distance, owing in particular to their position within a dialect continuum, as in the case of Venetian described above;
- d) lastly, languages or varieties by ‘pure’ elaboration, with limited internal autonomy (as in the case of the regional dialects in no. 3), but which have undergone a higher degree of elaboration. Examples include the Corsican dialect group, politically severed from neighbouring Tuscan dialects and under the influence of French for two centuries, and Spanish in the Americas as opposed to Peninsular Spanish.

A final important parameter needs to be considered: standardised distance languages, which can also serve as national languages, act as ‘umbrella languages’ (*Dachsprachen*, literally ‘roof languages’ or ‘roofing languages’ in Kloss’ terminology) for all other languages existing within the same national space. This overarching function of a language

is independent of the parameter of language-internal proximity. A *Dachsprache* thus covers not only regional dialects (standard Italian in relation to the dialects of Italy, for instance), but also neighbouring languages (such as Sardinian) and even varieties from other language families (like Greek or Albanian in southern Italy). The territory covered by an umbrella language is sometimes referred to as its ‘communicative space’.

In the following sections, the various Romance languages and dialects will be presented in more detail by association of ideas, and thus in no particular order. Particular emphasis will be placed on varieties that can be identified as languages by distance. The issues surrounding their classification will then be addressed (*cf.* 3.5).

3.2 Romance languages in Europe and across the world

Romance varieties in present-day Europe cover a geographical area that is still tied to the former extent of the Roman Empire: the *Romània* spans a spatial continuum ranging from the Atlantic to the Mediterranean Sea, including the Iberian Peninsula, France, as well as parts of Belgium, the Alpine zones and the Italian Peninsula. Latin was spoken for centuries within this same area in Antiquity (*cf.* 10.2.1 and 10.3.2). Only Romania is separated from the continuous geographical area throughout which a large number of Romance varieties coexist alongside numerous other languages today.

As a suitable starting point for a general overview of the Romance languages, let us consider a map of the so-called *Romània continua* (the zones in which Latin evolved into a Romance language *in situ*), provided by Walther von Wartburg in 1950. An updated, more precise version of this map indicating the extent of Latinisation under the Roman Empire and displaying the linguistic situation of Romance dialects as spoken around 1900 is shown in fig. 4.

Note that dialects – not standard languages, which emerged later as a result of linguistic elaboration – are the only direct successors of Latin. The area attributed to French (8), for example, does not coincide with the communicative space of standard Modern French in 1900 (= France, and parts of Belgium and Switzerland), but instead with the area in which French dialects (= *langue d’oïl*) were spoken. The latter includes the region of Wallonia in southern Belgium, but excludes the regions of Brittany (Breton-speaking, D), Alsace and north-eastern Lorraine (German-speaking, B) in France. Romance-speaking Switzerland is also excluded, as it was originally a Francoprovençal-speaking area (9).

Romance languages

- 1 Romanian: Daco-Romanian (1a); Aromanian (1b); Megleno-Romanian (1c); Istro-Romanian (1d)
- 2 Dalmatian
- 3 Italo-romance: regional dialects spoken in central Italy, Corsica and northern Sardinia (3a), as well as in northern (3b) and southern (3c) Italy
- 4 Sardinian: Logudorese (4a) and Campidanese (4b) dialects
- 5 Romansh
- 6 Ladin
- 7 Friulian
- 8 French
- 9 Francoprovençal
- 10 Occitan
- 11 Gascon
- 12 Catalan (12a, including Balearic varieties), Valencian (12b); enclave of Alghero (Sardinia) (12c)

- 13 Spanish
- 14 Galician
- 15 Portuguese

Non-Romance languages in Romance-speaking countries

- A Varieties of German and of Hungarian (Transylvania, Banat)
- B German dialectal varieties (north-eastern France, South Tyrol)
- C Greek (southern Italy)
- D Breton (Brittany)
- E Basque (Atlantic Pyrenees, northern Spain)
- F Slovenian (Friuli, Italy)
- G Albanian (southern Italy)

Source: Wartburg, 1950, with major modifications

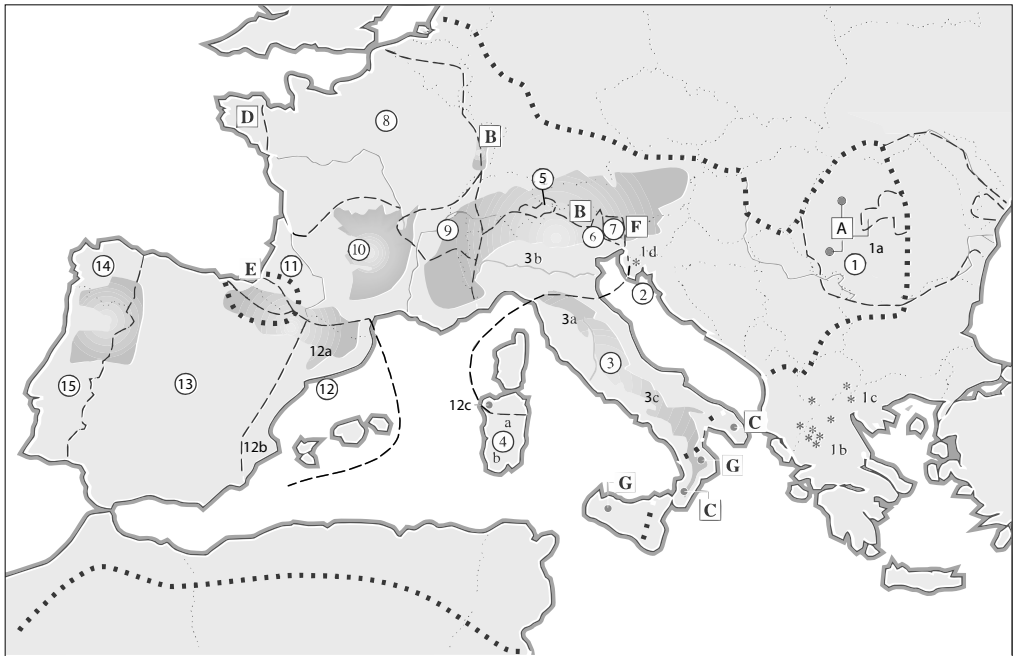


Fig. 4: Map of the *Romània continua*

-: National boundaries
- : Linguistic boundaries around 1900
- : Rivers (Loire, Rhône, Po, Tiber, Danube)
-: Maximum extent of Latinisation / Romanisation during the Classical period (around AD 200)
- =====: Mountains (Alps, Apennines, Pyrenees, Massif central, Vosges)

Today, the sociological status of all these languages is correlated with modern nations and their respective communicative spaces. Within this framework, national languages assume the role of dominant umbrella languages. They encompass (from west to east): Portuguese (15), Spanish (13), French (8), Italian (3), Romansh (or Rhaeto-Romance, in the canton of Grisons in Switzerland (5)) and Romanian (1).

In addition to the above, there are a few Romance languages that have their own autonomous areas. These include Galician (14), spoken in the independent region of Galicia in north-western Spain; Catalan and Valencian (12), linguistically very close to one another; but also Dolomitic Ladin (6) in the valleys of the Dolomites; Friulian (7) in the Friuli area of north-eastern Italy; Sardinian (4); and Francoprovençal (9) in the Aosta Valley in north-western Italy (whereas it has no particular status in French-speaking Switzerland or the French Alps, where it survives in several places).

Further languages, which do not have official status, include Occitan (10) in southern France and Gascon (11) between the Atlantic Ocean, the Pyrenees and the river Garonne.

The map also shows non-Romance languages which have their own territory within the *România continua*, or at least the most salient ones from a historical point of view:

- Franconian and Alemannic dialects in border areas (B) [thus, strictly speaking, outside the *România continua*];
- Breton, an Insular Celtic language, originating from the British Isles and introduced during the 5th and 6th centuries (D) [*id.*];
- Basque, which predates the Roman Empire (E) and is arguably a ‘pre-Indo-European’ language;
- Hungarian and German varieties in Transylvania and Banat, introduced under the Austro-Hungarian Empire (A);
- Greek (C), Albanian (G).

The majority of these languages have since lost considerable ground: in Romania, dictatorial politics were hostile toward minority languages. After 1989 the emigration of German speakers increased, whereas in Italy and France, the trend is largely due to natural reasons, though politics have also played a role. Basque has maintained its vitality in Spain, alongside Catalan and Galician, and the Alemannic dialects in Alsace are still spoken, strengthened by neighbouring varieties in the regions of Baden and Basel.

Wartburg’s map further indicates the territories where the Roman Empire more or less rigorously introduced Latin, but where it was subsequently lost. These include southern England, a German- and Slavonic-speaking fringe extending from the Netherlands to the Balkans, and northern Africa. In some of these places a Romance variety developed after the fall of the Roman Empire, only to disappear between the 9th and 11th centuries; they are therefore collectively known as the ‘*România submersa*’. In northern Tunisia and certain parts of the Balkans, as well as in areas bordering on the Germanic territo-

ries, traces of them can still be seen in the lexis and especially in proper names (e.g. in the Moselle Valley, the northern part of the Saarland or some places in the Black Forest).

The situation is slightly different for Dalmatian (2). This Romance language, once spoken on the Adriatic coast, is known to us through a small number of written documents dated between the 16th and 19th centuries and some oral transcriptions from the end of the 19th century. Influenced by Venetian dialects since the Middle Ages, it has nevertheless remained recognisably independent in terms of internal structure and it shows incidental interference with southern Slavonic languages. In some (limited) respects, it can be considered to represent a transitional state between Italo-Romance and Daco-Romance. It is generally assumed that this language became extinct with the death of its last speaker on the isle of Krk (formerly Veglia) in 1898. The area in which it was spoken thus belongs to that part of the *România continua* that has become the *România submersa* (cf. the bibliographical references in 3.2).

Today, however, the greatest number of speakers of Romance languages do not live in Europe – where there are roughly 175 million – but in the Americas, where Spanish and Portuguese have spread considerably, their speakers numbering 300 million and 180 million, respectively. The presence of Romance languages on extra-European soil is the result of colonisation from the 16th to the 19th centuries. Apart from creole languages, which are categorised separately and will be discussed later (cf. 3.6.3), they are the same languages as those spoken in Europe, despite the fact that they have followed a different evolutionary path from their ancestors in some respects.

This *România nova* differs from the *România continua* in that it does not share the geolinguistic continuity that has characterised the latter since Late Antiquity; nevertheless, it can look back on five centuries of internal and external linguistic history. Spanish displays the greatest amount of variation. It is spoken in nineteen countries in the Americas, listed below in the order of their geographical position from north to south and west to east (for an overview of the extra-European *România*, cf. the maps provided by Bosson, *Die Romanischen Sprachen*, 2008: 308–321, as well as those in RSG 1, art. 76–100, which offer a more detailed view):

- the Antilles islands, including Cuba, the Dominican Republic and the North American state of Puerto Rico;
- Mexico and several Central American countries, including Guatemala, Honduras, Costa Rica, El Salvador, Nicaragua and Panama;
- Colombia, Venezuela, Ecuador, Peru, Chile, Argentina, Uruguay, Paraguay and Bolivia.

Portuguese is the official, or semi-official, language of Brazil, the Madeira Islands, the Azores and Macau, as well as of other places where it coexists with a creole language that has a Portuguese lexical base. These include Guinea-Bissau, the Cape Verde Islands and São Tomé and Príncipe. French is the official language and mother tongue

in Quebec and in several smaller territories: these are the four overseas departments (*départements d'outre-mer* – DOM) Guadeloupe, French Guiana, La Réunion and Martinique and the four overseas territories (*territoires d'outre-mer* – TOM), as well as a few territories in the Pacific and Indian Ocean (in addition to Saint Pierre and Miquelon just off the North American coast).

Finally, there are countries where Romance languages are present, not as dominant varieties, but as minority or second languages. Examples are Spanish in the United States (its speakers being almost equal in number to those living on the Spanish peninsula), Portuguese in Angola and Mozambique (roughly 20 million speakers), and French as a *lingua franca* in the sixteen sub-Saharan African countries, the three Maghreb countries Morocco, Algeria and Tunisia as well as Madagascar. Furthermore, the high rate of mobility in Western Europe in recent decades has led to the establishment of large Romance-speaking communities outside their home countries, such as Portuguese in France, Italians in Switzerland or Germany, Romanians in Great Britain, etc. (cf. 10.6.1 for more detail).

→ RSG 1, sect. VI (*L'histoire linguistique de la Romania submersa*, art. 61–65), VII (*L'histoire linguistique externe de la Romania continua*, art. 66–75), VIII (*L'histoire linguistique externe de la Romania nova*, art. 76–96), IX (*L'histoire des langues de la Romania creolica*, art. 97–100)

3.3 Geolinguistic boundaries: the example of Gallo- and Italo-Romance dialects

When attempting a geographical description of the Romance languages, it is relatively easy to draw boundaries between standardised languages (e.g. between Italian and French or Spanish and Portuguese). Wartburg's map, however, is not based on standard languages but on dialectal varieties, the demarcation of which is a more sensitive issue. This is particularly clear in the case of Gallo-Romance and even more so in the case of Italo-Romance, as shown above by the example of Venetian and Sicilian (cf. 3.1).

Let us begin with Gallo-Romance. The Gallo-Romance group consists of the regional dialects of French (= *langue d'oïl*), Francoprovençal, Occitan (= *langue d'oc*) and Gascon, and it is characterised by a number of phonetic, grammatical and lexical features that set it apart from the majority of other Romance varieties. Within Gallo-Romance, spoken varieties of Francoprovençal occupy an intermediate position; in other words, they are linguistically closer to the *oc* and, more particularly, to the *oïl* dialects than French is to Occitan. Occitan varieties are internally divided into Northern and Southern Occitan. Furthermore, they are to be distinguished from Gascon, which is spoken in the south-western corner of France, west of the river Garonne.

The differentiation of these four Gallo-Romance languages manifests itself in the form of structural differences found in all internal domains of language. Dialectal varieties situated on either side of a dividing line traced on a map display a large number of

immediate differences, while at the same time sharing numerous features (*cf.* also 4.2.1 no. 3, fig. 13, map of Gallo-Romance).

An illustrative example (which will be easier to understand after having read section 6.5 on historical phonetics and 5.3 on grammaticalisation) concerns the Latin word *CASA* /*kasa*/ (n.f.) “house”, transformed through phonetic change over centuries to produce the following results:

O.Fr.	<i>chiés</i> /tʃjɛs/ > Fr. <i>chez</i> /ʃe/	(prep.)	“at, beside”
North. Occ.	<i>chas</i> /ʃas/	(prep.)	“at, beside”
Frpr.	<i>tsi</i> /tsi/	(prep.)	“at, beside”
South. Occ.	<i>casa</i> / <i>kazə</i> /	(n.f.)	“house”
Gasc.	<i>cazo</i> / <i>kazə</i> /, <i>case</i> / <i>kazə</i> /	(n.f.)	“house”

The phonetic features of these forms exhibit a pattern of differences that typically serve to distinguish the languages in question:

- Latin /k/ preceding /a/ evolves into /tʃ/ in Old French, then into /ʃ/ in Modern French and some northern Occitan varieties, whereas it changes to /ts/ in Francoprovençal, but remains unchanged in southern varieties of Occitan and in Gascon;
- stressed /a/ after /k/ in open syllables (i.e. syllables ending in a vowel sound) becomes /jɛ/ in Old French (which becomes /ɛ/ in Modern French), /i/ in Francoprovençal, while it does not change in Occitan;
- additionally, the whole of Gallo-Romance, with the exception of southern Occitan and Gascon, displays a change of word class: what was originally a noun becomes a preposition;
- the previous observation also explains the differences in the word ending: as a general rule, final /a/ is preserved in Gallo-Romance in a more or less weakened form; in the present case, however, this is only apparent in Southern Occitan and Gascon, as in the other languages the formerly lexical word has undergone reduction to become a grammatical word.

In this example, the difference between southern and northern Occitan varieties is more obvious than that between Northern Occitan and French. Linguistic differentiation within a single dialect continuum often involves relative rather than absolute differences. It is also interesting to note that, as far as the word *CASA* is concerned, Old French appears to be more similar to Francoprovençal than Modern French. Divergences between varieties, then, have increased over time (*cf.* also FEW 2/1, 449b *sqq.* s.v. *CASA* – the entry provides information on all historical and dialectal variation exhibited by this word within Gallo-Romance).

A further example (without detailed remarks) concerns the phonetic evolution of Latin *FICATUM* /*fikatu*/ “liver” (accompanied by metonymic change from a participle meaning

“stuffed with figs” to a homonymic noun meaning “liver of geese fattened with figs”, a culinary concept borrowed from Greek), which leads to the following results (*cf.* FEW 3, 490b *sqq.* s.v. FICATUM):

foie /'fwa/ in French

fedzo /'fɛdzo/ in Francoprovençal

fège /'fɛdʒe/ in Occitan

hidge /'hidʒe/ in Gascon

The lines drawn on maps to delimit such linguistic differences geographically are called ‘isoglosses’. Particularly salient boundaries where several isoglosses converge are known as ‘isogloss bundles’. All the linguistic boundaries within the *Romània* indicated on Wartburg’s map are represented in the form of isogloss bundles. The formation of Romance-speaking areas during the second half of the first millennium will be addressed in more detail in the corresponding section on external history. Nevertheless, the importance of mountain chains in the formation of boundaries between Romance languages can already be noted here (*cf.* 10.3.2 for more detailed argument).

Within Italo-Romance, dialectal diversification is even more marked than in Gallo-Romance. The dialects of northern and southern Italy encompass varieties that are quite distant from the standard language (‘Italian’), such as Genoese, Ticinese (in Switzerland) or Venetian, as well as Neapolitan, Apulian or Sicilian. The standard variety, Italian, is historically based on the literary Florentine models of the authors known as the ‘Three Crowns’ (Dante, Boccaccio and Petrarch) and is thus directly related to the Tuscan dialects, which also include Corsican and are close to the central (‘median’) dialects such as Umbrian (*cf.* *infra* and 3.4.5 no. 4).

Furthermore, a number of other Romance languages are spoken in geographical proximity to Italo-Romance dialects. They share certain features with the dialects, but at the same time display their own distinguishing characteristics that separate them from this multifaceted mosaic. The languages in question are Sardinian, Romansh, Ladin and Friulian, as well as – covering smaller areas – Catalan (spoken in Alghero, north-eastern Sardinia), Occitan (Piedmont and Guardia Piemontese, Calabria), Francoprovençal (province of Foggia) and French (Val d’Aosta).

Two main geolinguistic ‘frontiers’ can be observed which split the Italo-Romance dialects proper into three large dialect groups (*cf.* fig. 5 below). The frontier between the northern and central dialects is founded on a large number of diverging phonetic and lexical features, and coincides with the northern boundary of the Apennine mountain range. This dividing line is also known as the ‘La Spezia-Rimini line’. The boundary between central and central-southern dialects – the ‘Roma-Ancona line’ – is less marked, though recognisable. These divisions are neatly illustrated by the following map, which displays isogloss bundles for these two frontiers, established by Gerhard Rohlfs as early as 1937:

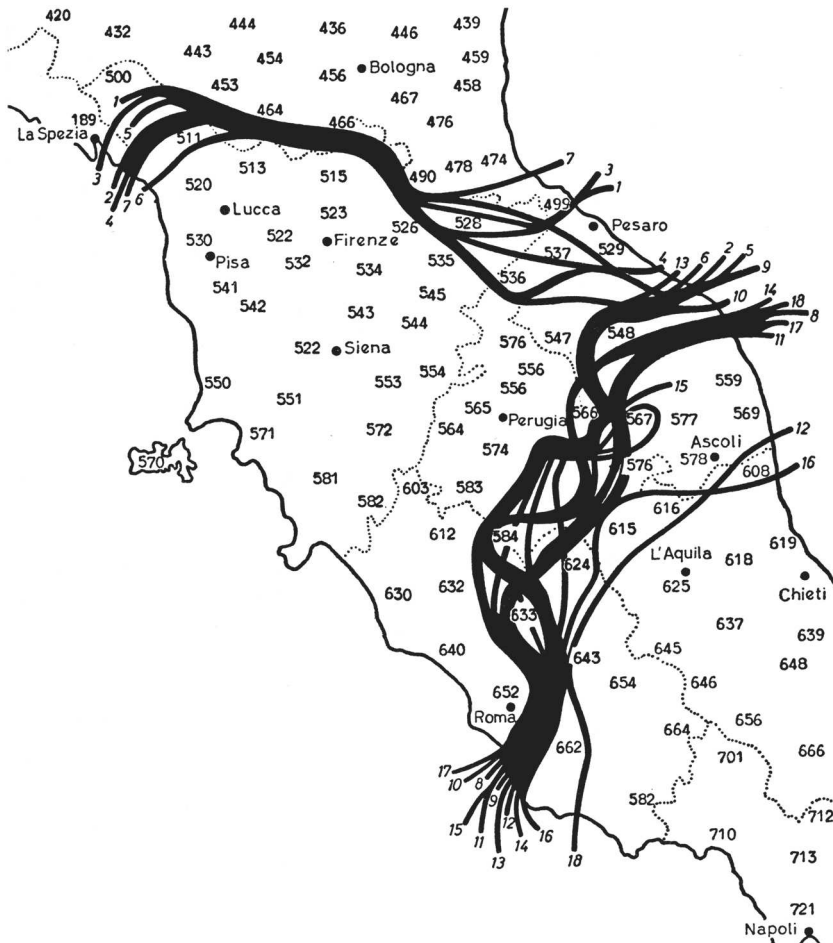


Fig. 5: Examples of isoglosses marking the linguistic dividing lines *La Spezia-Rimini* and *Roma-Ancona*

Source: after Rohlfs, *La struttura linguistica dell'Italia*, 1937, map 2; cf. Tagliavini, 1982, § 68 [the numbers 189 and 420–721 on the map correspond to the geographical inquiry points of the AIS, cf. 1.5.2 and fig. 6].

The following classic example makes use of the eighteen criteria cited by Rohlfs, in order to illustrate the methodology employed in the determination of linguistic boundaries such as these. The isoglosses forming the *La Spezia-Rimini* line are based on salient phonetic differences. Variant forms found south of the line are conservative (cf. 6.5):

Consonantal phenomena:

- 1: isogloss separating northern variants, in which the plosive /k/ in intervocalic position is voiced, from southern variants, in which the Latin plosives persist: e.g. Lat. *ORTICA* “nettle” > (north) *ortiga* vs. (south) *ortica*
- 3: intervocalic /p/ has evolved into a fricative north of the line, whereas it is maintained in the south: e.g. Lat. *CAPILLI* “hair” > (north) *cavei* vs. (south) *capelli*
- 4: double consonants have been simplified in the north, whereas they are maintained in the centre and south: e.g. Lat. *SPATHULA* “small sword” > **espadla* > (north) *spala* /*spala*/ “shoulder” vs. (south) *spalla* /*spal:a*/

Vocalic phenomena:

- 2: final, unstressed /e/ has been lost in certain areas of the north (within certain phonetic environments), but is maintained in the south: e.g. Lat. *SALE* “salt” > (north) *sal* vs. (south) *sale*
- 5: pretonic vowels have been dropped in the north, but are maintained in the south: e.g. Lat. *SELLARIU* “saddle” > (north) *slér*, *slar* vs. (south) *sellaio*
- 6: nasal consonants have been dropped and preceding vowels nasalised in the north, while nasals are maintained without vowel nasalisation in the south: e.g. Lat. *PANE* “bread” > (north) *pã* vs. (south) *pane*

Lexical phenomena:

- 7: lexical differences can also be observed between northern and southern varieties: e.g. Lat. (*HINC*) *HODIE* “today” > (north) *incø*, *incù* vs. (south) *oggi*

The isoglosses forming the Roma-Ancona line rely mainly on lexical phenomena, in addition to some morphological and phonetic characteristics:

- 8: (north) *fabbro* vs. (south) *ferraru* “blacksmith”
- 9: (north) *fratello* vs. (south) *frate* “brother”
- 10: (north) *donna* vs. (south) *femmina* “woman”
- 11: (north) *mio figlio* vs. (south) *figliomo* “my son” (possessive determiner appears as an enclitic)
- 12: (north) *ha le spalle larghe* vs. (south) *tene le spalle larghe* “he has broad shoulders” (*TENÈRE* as opposed to *HABÈRE* used as a verb to denote possession)
- 13: (north) *coscia* vs. (south) *cossa* “thigh”
- 14: (north) *la cimica* vs. (south) *lu cimice* “bedbug” (the gender is maintained in the south but changes in the north – a frequent phenomenon among words of the 3rd declension)
- 15: (north) *faggio* vs. (south) *fagu* “beech tree”
- 16: (north) *montone* vs. (south) *mondone* “battering ram” (assimilation of *nt* > *nd*)
- 17: (north) *denti* vs. (south) *dienti* “teeth” (metaphony or conditioned diphthongisation, cf. 6.5.2 no. 3)
- 18: (north) *aceto* vs. (south) *acito* “vinegar” (metaphony)

This illustration provides insights into the internal criteria that are used to identify boundaries between linguistic varieties. Distinctions made using these methods are

scientifically reliable, though they always remain relative. Two neighbouring dialects situated on either side of an isogloss bundle are still very similar and are almost always mutually intelligible (cf. 4.2.2). The same principle led Wartburg to group the Italo-Romance dialects north of the La Spezia-Rimini line with Gallo-Romance dialects (cf. 3.5.3). This is justified, particularly when their primitive medieval form is taken into consideration. Northern Italy thus forms a zone of linguistic transition between the *langues d'oc* and *d'oïl* and the central and southern Italian varieties.

→ Barbato, *Major isoglosses in Romània*, OxfEnc

3.4 External characteristics of the Romance languages

The geographical and geolinguistic defining criteria presented thus far (3.1–3.3) give a fair idea of the difficulties involved in describing a language adequately. These difficulties stem from the fact that a definition of the phenomenon of *language* and the description of a given *individual language* require different approaches, as individual languages are shaped by their external environment. An overview of the essential external characteristics of each of the Romance languages will therefore be provided, encompassing the following points:

1. Current speaker numbers in different territories
2. Current contact languages (within countries in which the languages in question have their own territory)
3. Elaboration, in the case of written languages
4. The importance of regional dialects (for languages whose native speakers are essentially speakers of dialectal varieties, such as Francoprovençal, this information has been incorporated into point 1.)

For this overview, we have chosen the fourteen modern Romance languages that most clearly fulfil the criterion of linguistic distance. These include: French, Occitan, Gascon, Francoprovençal, Italian, Sardinian, Romansh, Ladin and Friulian (the latter two being treated within the same section), Spanish, Catalan/Valencian, Galician, Portuguese and Romanian⁷. Each section contains references to the corresponding articles in the encyclopedias of the Romance languages (cf. 2.4.1)⁸, which essentially describe the same

⁷ No systematic reference will be made to the chapters on variation, language-internal characteristics and external history, where the information presented in the following will be placed within a broader historical context. Likewise, no explicit reference will be made here to the relevant sections of Georg Bossong's handbook (*Die romanischen Sprachen*, 2008) that have served as a guide for the present chapter.

⁸ For Bec, the relevant page numbers are given, for the LRL and the RSG the volume and article number, and for the OxfGuide, the article number.

fourteen languages⁹. Bec's *Manuel pratique* (1970/71) is particularly useful as it is the only work that includes a body of annotated texts written in the individual languages, thereby providing the reader with a concrete idea of their physiognomy. Like the MRL, the LRL devotes an entire volume to each of the national languages, while their history is to be found in the extensive and crucial volume 2/2. For the RSG, only the sections dealing with external history and language history will be referred to (volumes 1 and 3), omitting those which treat individual diasystematic topics. CambrHist and CambrHandb do not describe each language individually and will therefore not be mentioned here. The same applies to the OxfEnc as the current version does not offer a balanced coverage of the languages.

→ Andreose/Renzi, *Geography and distribution of the Romance languages in Europe*, CambrHist 2, 8
Jones/Pountain, *Romance outside the România*, CambrHist 2, 10

3.4.1 French and the *langue d'oïl* dialects

1. The French language has approximately 84 million native speakers who live in France, Belgium, Switzerland, Luxembourg, Monaco and Quebec (cf. 3.2 for the French Overseas Departments and Territories). It is more difficult to determine the number of second language speakers of French, particularly on the islands of Haiti, Mauritius and the Seychelles – where French coexists with a creole language with a French lexical base – and even more so in Africa. The number of speakers suggested varies between 25 and 130 million (in some cases considering all inhabitants of a country in which French is actually only spoken by a minority as 'French-speaking'); 65 million seems a reasonable number (cf. Bossong, *Die romanischen Sprachen*, 2008: 158–160).

2. In France, the national language coexists with numerous minority languages spoken within a specific territory, only some of which belong to the Romance family. They include Gascon, Occitan and Francoprovençal (cf. 3.4.2–4 below), Catalan (in the eastern Pyrenees), Corsican (cf. 3.4.5 no. 3) as well as some pockets of *oïl* dialects that are spoken within specific areas.

The precise delimitation of the area covered by a minority language is a matter of considerable difficulty, since the age of its speakers and the contexts in which it is used

⁹ Bec does not treat Galician; in the LRL, Francoprovençal and Gascon are only treated in one or two articles and they are absent from the RSG; Gascon is also absent from the OxfGuide. Romansh, Friulian and Ladin are usually treated together. On the other hand, the various encyclopedias, with the exception of the MRL, deal with Dalmatian, which is omitted here as it became extinct around 1900 (cf. 3.2 and the following encyclopedic references: Bec 2, p. 393–416 [by Ž. Muljačić]; LRL 3, art. 207 [M. Doria, *Dalmatico*: a) *Storia linguistica interna*; b) *Storia linguistica esterna*]; 2/2, art. 122; RSG 3, art. 226 [A. Bernoth, *Interne Sprachgeschichte des Dalmatischen*]; OxfGuide 9 [Maiden, *Dalmatian*]; OxfEnc [Maiden, *Dalmatian (Vegliote)*]).

are determining parameters, and since the situation of a minority language typically fluctuates (the issue is brought to light in an exemplary manner in the detailed cartographic representations presented by Bernissan 2012 for the case of Gascon, cf. 3.4.3 no. 1 below). In contrast, Italian, Portuguese or Spanish are not linked to specific geographical areas within France, even though they have a solid presence within the country due to immigration.

Among the non-Romance languages, the Alemannic dialects of Alsace, Franconian dialects in German-speaking Lorraine, Flemish (in Pas-de-Calais), Breton and Basque are all associated with a specific geographical area. In addition, some immigrant languages such as Arabic, Turkish or East Asian languages account for a significant number of speakers (at least 3 million of the 68 million inhabitants of France are Arabic-speaking, though no official census has ever been conducted). At the same time, French serves as a common reference point for speakers of each of the minority languages within the country.

Switzerland is unique in having four languages, three of them Romance, which share the status of official national languages. These are German (in German-speaking Switzerland, where Swiss dialects dominate in everyday spoken language), French (with about 2 million speakers in the '*Romandie*', where Francoprovençal used to be the spoken language, cf. 3.4.4), Italian (in the canton of Ticino and parts of Grisons), and Romansh or Rhaeto-Romance (in the canton of Grisons, cf. 3.4.7).

Belgium recognises two national languages, French (ca. 4.8 million speakers) and Flemish (ca. 6.7 million speakers), and a small area is German-speaking. Though the linguistic make-up of the country appears simple, it is the cause of significant language conflict. In Luxembourg, French coexists with the Franconian dialect Luxembourgish ('*Lëtzebuergsch*', the dominant mother tongue language) as well as with German. Finally, in Canada, the Province of Quebec is bilingual (with a French-speaking population of about 8.5 million), whereas the remaining provinces are mainly populated by monolingual English-speakers.

3. French as a written language was forged through a process of neutralisation from various spoken dialects, as well as on the model of Latin, from the 12th century onwards in particular. Medieval French was used throughout an area that included the British Isles (home to the Anglo-Norman dialect), northern Italy (where Franco-Italian was a literary minority language), as well as in the Latin kingdom of Jerusalem, where written French was used as a *lingua franca*. Its form varied from region to region, though it maintained a certain degree of general unity, which was not the result of a central model, but of the conditions of its emergence, which were the same everywhere. These included the desire for distance from the spoken dialects, a preference for supra-regional forms and the imitation of written Latin. Regionally marked writing in Old French undoubtedly differed significantly from the *oïl* dialects spoken at the same time, including those in the Parisian region, which played a key role in language neutralisation and standardisation processes, especially from the 13th and 14th centuries onwards.

4. The *oïl* dialects survived in the sphere of everyday use throughout their area of origin up to the beginning of the 20th century (cf. Carles/Glessgen, *Les écrits des Poilus*, 2020; cf. 4.2.1, fig. 14, map of *oïl*-speaking zones). From the 14th century onwards, however, the central zone of France underwent a process of de-dialectalisation, owing to the growing influence of the Parisian centre. Yet in the other regions, local dialects subsisted until as recently as the 1960s, finally succumbing to the pressure of modern-day communication (literacy, mobility, and, more recently, television).

Since the year 2020 or thereabouts, dialect speakers with mother tongue competence have been restricted to the oldest sector of the population, and to a few rare localities. This is particularly the case for Walloon in Belgium or, with even fewer speakers, Lorrain in the Val d'Orbey and the Val de Villé in the Vosges mountains (both under German administration during the period of intense de-dialectalisation between 1871 and 1918). There are also a reduced number of speakers of the Franc-Comtois dialect, some of whom are in the Swiss canton of Jura, of Gallo in eastern Brittany, of Norman in the Channel Islands, Guernsey and Jersey, and of Picard. Estimates of speaker numbers are generally exaggerated. Our own estimate is based on numerous personal enquiries; we consider that their number does not exceed 10,000 individuals in France (around 0.3% of the population in the *langue d'oïl* regions, which were still entirely dialect-speaking a century ago).

→ Bec 2, p. 1–132

LRL 5/1; 2/2, art. 138–145

RSG 1, art. 70; 76–81; 98–99; 3, art. 235–239

OxfGuide 18

MRL 8 (*Manuel de linguistique française*, eds Polzin-Haumann/Schweickard)

MRL 22 (*Manuel des francophonies*, ed. Reutner)

3.4.2 Occitan or the *langue d'oc* dialects

1. The varieties of Occitan are quite diverse from a linguistic point of view. Differences between southern forms (Languedocien and Provençal) and northern forms (the Limousin-Auvergnat group and Dauphinois) are particularly marked, as discussed above in relation to divergent patterns displayed by the phonetic evolution of Latin CASA. The Occitan-speaking regions do not have a politically autonomous status and there are no native speakers of a standardised variety. In other words, all mother tongue speakers are dialect-speaking.

The number of current native speakers of Occitan is not easy to evaluate, since the situation is complicated by a strong political and ideological undercurrent. Numbers suggested recently vary between half a million and two million. A recent and very detailed enquiry conducted in the Bigorre region of Gascony estimates the number of native speakers capable of spontaneous conversation in the whole of southern France to be approximately 100,000 (Bernissan, *Combien de locuteurs compte l'occitan en 2010?*,

2012). We predict that this number will have declined to ca. 30,000 speakers by 2030. This indicates a sharp and regrettable decline in comparison with the 9 or 10 million speakers presumed to have persisted at the beginning of the 20th century, even if one assumes that approximately 500,000 individuals are still able to understand Occitan or Gascon, particularly in the mountainous regions of the Massif central, the Alps and the Pyrenees.

In addition to these numbers, there are several hundred speakers of Provençal in various Alpine valleys of the Piedmont region. The same figure applies to the number of speakers of the Provençal variety displaced to Guardia Piemontese in Calabria by a medieval religious migration (the Vaudois).

2. Contact languages of Occitan are, in France and Italy, French or Italian, respectively. Speakers of Occitan varieties are all bilingual.

3. Occitan enjoyed a rich literary culture during the Middle Ages (an example of its manifestation being the poetry of the troubadours). The written *langue d'oc* is traditionally referred to as 'Old Provençal', even if this term is unfortunate owing to its potential for confusion with the dialects of the Provence region. The forms of Medieval Occitan that had undergone a process of elaboration were regionalised as a result of the abandonment of local spoken characteristics, although strong differences between the various areas remained.

The annexation of Occitan-speaking territories by the French crown began with Languedoc after the Albigensian Crusade (1208–1229) and ended for the most part at the end of the 15th century. The political control from the north drastically reduced the written use of Occitan after 1550 and prevented it from becoming standardised during the modern period (cf. 3.5.1 and 3.5.3 no. 5).

There have been two main attempts at reviving written Occitan in modern times. The first was championed by the 'Félibrige' movement towards the end of the 19th century; its protagonist was Frédéric Mistral (1830–1914), who received the Nobel prize for his verse novel *Mirèio*, written in Rhodanian Provençal (= the region of the Rhône delta). A second and more successful attempt was made by the *Institut d'Estudis Occitans* (IEO) in Toulouse, founded in 1945. The Félibrige developed an orthography inspired by French and based on pronunciation (under the influence of Joseph Roumanille, even though it is known as the '*graphie mistralienne*'), whereas the orthography developed for Occitan by the IEO is etymologically inspired and is based on the Languedocien dialect (the so-called '*graphie classique*'), used in most recent writing. Neither of the two movements, however, were really able to become established, nor did they have much impact on native speakers.

→ Bec 1, p. 395–462

LRL 5/2, art. 340–348; 2/2, art. 147–151

RSG 1, art. 71; 3, art. 240

OxfGuide 19; MRL 35 (*Manuel de linguistique occitane*, eds Esher/Sibille)

3.4.3 Gascon

1. Gascon essentially covers the region between the Garonne, the Atlantic and the Pyrenees. Its most striking phonetic characteristics go back to a regional differentiation that had already occurred in Latin, due, in turn, to language contact with Proto-Basque (cf. Chambon/Greub, *Note sur l'âge du (proto)gascon*, 2002). Internally, a distinction is made between the dialects of northern Bordelais and the southern dialects, the latter being further divided into the western dialects ('Gascon noir' in the Landes and western Béarn) and the eastern ones (in the rest of Béarn, in the Gers and in the Bigorre). In addition, further east, there is the Aranese variety of Gascon in the Catalanian Val d'Aran. Close contact with the neighbouring Occitan of Languedoc has resulted in a certain degree of convergence and in the development of a number of transitional varieties, but from a diachronic point of view the distinction between the two languages is clear.

Like Occitan, Gascon does not have a politically autonomous status in France, except in the Val d'Aran, where the regional language has acquired a high degree of independence under Spanish sovereignty, and about half of the 8,000 to 9,000 inhabitants of the area have at least some passive knowledge of the language.

Overall, the number of speakers of Gascon is relatively high compared to the other Gallo-Romance dialects, especially in the sparsely populated Landes region and in the southern Pyrenees (cf. 3.4.2 no. 1 above). Here too, however, the youngest speakers are currently already around 50 years old, while the most fluent speakers are closer to 70. In the Val d'Aran, speakers who actively use the language on an everyday basis are few, and they continue to gradually decline in number.

2. French is by far the most important contact language of Gascon, as it is for Occitan and Francoprovençal. Historically, there has also been sporadic interaction between Gascon and Aragonese, south of the Pyrenees. The Val d'Aran is trilingual, owing to the presence of Catalan and Spanish alongside Aranese.

3. The written elaboration of Gascon in the Middle Ages began around 1220, but was almost exclusively limited to the domain of (legal and administrative) documentary texts (cf. Glessgen, *Pour une histoire textuelle du gascon médiéval*, 2021; *L'étude linguistique du gascon médiéval*, 2022a). These were nevertheless abundant and they continued to be produced in the region of Béarn until the end of the 16th century. In more recent times, similarly to Occitan, a literary renaissance has emerged, to which a number of authors have contributed, and which is characterised by regional forms of writing as well as by the use of Occitan in music and the media.

→ Bec 1, p. 509–554
LRL 2/2, art. 152

3.4.4 Francoprovençal

1. Formerly, the area throughout which Francoprovençal was spoken included the largest part of Romance-speaking Switzerland (with the exception of the northern canton of Jura, where *oïl* dialects were spoken), and extended along the length of the Rhône valley, encompassing the large region around the former capital city of Lyon. It is also spoken in the Aosta Valley in Italy. An almost exclusively oral language, Francoprovençal thus consisted of a number of regional dialects sharing a series of specific internal characteristics. The vitality of the spoken language was still intact in the 19th century, after which it declined considerably – first in France, then in Switzerland, and finally in Italy. The autonomous status of the Aosta Valley slowed the process down somewhat; however, even in this region there are no longer many native speakers among the younger generations.

The last native speakers of Francoprovençal, elderly for the most part, are concentrated in the Aosta Valley and in certain parts of the Swiss canton of Valais (including, in particular, the village of Évòlène, which, at an altitude of over 1300 metres, is somewhat isolated from major traffic routes, and where the local variety is therefore still used by younger generations of speakers). Their number is very difficult to ascertain accurately; available estimates, which suggest several tens of thousands of speakers with solid linguistic competence, regrettably seem rather optimistic. A further variety of Francoprovençal is spoken by a very small number of persons in the two villages of Faeto and Celle San Vito in northern Apulia – the remnants of an expansion that took place during the medieval period.

Owing to the low numbers of young speakers, Francoprovençal is among the endangered Romance languages.

2. In areas where it is spoken, Francoprovençal is always a minority language alongside French or Italian. Its speakers are all at least bilingual. In the Aosta Valley, French also plays a role as a written language.

3. At no stage in its evolution was Francoprovençal subjected to a high degree of written elaboration. From the 13th to the 15th century, however, a comprehensive written tradition consisting of documentary texts developed in Grenoble, Lyon and the Forez, as well as on both the French and Swiss sides of the Jura. In the latter areas, a form of mixed language developed whose *scripta* was phonetically adapted to French but which contained vocabulary belonging to Francoprovençal. French only became established as the sole language from the 15th century onwards.

This explains why it was not until quite late on that Francoprovençal was identified as a language in its own right by the Italian comparative philologist Graziadio Isaia Ascoli (1829–1907; cf. 2.2.3). During the modern era, French was the written language in the French and Swiss zones mentioned, while this role was assumed by Piedmontese

in the Aosta Valley. In contrast, Francoprovençal appears sporadically in folkloric texts and dialectal literature.

→ Bec 2, p. 357–391
 LRL 5/1, art. 330; 2/2, art. 146
 OxfGuide 20

3.4.5 Italian

1. Italian is known among the population of all of Italy, including Sardinia and the Swiss canton of Ticino. It has about 60 million native speakers in Italy and up to 25 million more speakers throughout the world, as a result of large waves of emigration to other countries in Europe, North America, Argentina or Australia (*cf.* Haller, *La lingua degli emigranti e degli esilati: italiano*, RSG 1, art. 167). There are about 350,000 inhabitants in the canton of Ticino in Switzerland, alongside numerous Italian-speaking individuals living in other parts of the quadrilingual country (all together, these amount to more than half a million native speakers if one includes cantons other than Ticino; for details on the canton of Grisons *cf.* 3.4.7 no. 1 below).

2. As far as multilingualism is concerned, the situation in Italy is similar to that in France, except that the dialects of Italy still exhibit a strong presence (in contrast to Occitan or Francoprovençal in France) and the regional languages are generally supported by political autonomy. The national Italian language coexists not only with the numerous and highly diversified Italo-Romance dialects, but also with other Romance languages in several regions, which are presented here separately (*cf.* fig. 6): Sardinian, Dolomitic Ladin and Friulian (*cf.* below) as well as Francoprovençal, and, very sporadically, Occitan (*cf.* above).

Furthermore, five non-Romance languages are used within specific geographical areas: Germanic varieties (Tyrolean in South Tyrol, the Alemannic dialect spoken by the Walsers around the Monte Rosa massif, Bavarian in the Seven and Thirteen Communities [*Sette e Tredici Comuni*] on the south-eastern side of the Alps), Slovenian (in north-east Italy), Greek, Albanian and Serbo-Croat (introduced into southern Italy at different periods). Moreover, a rich array of recent immigrant languages is spoken in the country today.

3. As was the case for the *langue d'oïl* and *d'oc*, written medieval varieties of Italo-Romance dialects were differentiated according to macroregions, with the difference that major cities (particularly Genoa, Milan, Venice, Florence, Rome and Naples) played a much greater role in their formation. In contrast to French, however, the Italian standard language that was developed and institutionalised in modern times is not the organic result of a form that was subjected to several centuries of neutralisation and

diffused by the Royal Court. Instead, it is based on a specific regional dialect, Tuscan (and Florentine in particular), which underwent a somewhat artificial process of elaboration. Thus, in the 16th century, the archaic written variety of 14th-century Tuscan was used as a model for a supra-regional literary language that would eventually become established as the national language in the 19th century.

During the Middle Ages, written regional languages such as Neapolitan, Venetian, Sicilian or Tuscan coexisted in a relatively independent manner (although the influence of Tuscan began to make itself felt in several regions from the 15th century onwards). Moreover, in the modern period, literary texts of some importance continued to be produced in these regional languages up until the emergence of the present-day state; a dialectal literature worthy of attention still exists today. In Corsican, which was originally a Tuscan dialect but which has been in contact with French since the political integration of Corsica into France in 1768, textual production has attained considerable diversity in recent times.

4. The limited political centralisation of Italy during the modern period ensured that the spoken dialects remained vital throughout the 20th century. Still today, one in every two Italians regularly uses a dialect (especially in Calabria and Sicily or Veneto), even if the percentage of dialect speakers among younger generations has undergone a considerable decline. Today's media seem to have severed the younger generation from the dialects of their ancestors, so that within 50 years from now the dialectal situation may have changed quite drastically.

The map (fig. 6) represents a simplified version of the *Carta dei dialetti d'Italia*, which was presented by Giovan Battista Pellegrini in 1977. It provided the first ever cartographic overview of the Italian dialectological landscape in all its complexity. (It does however, contain an error: the territory labelled as 'Alto Adige' on this map actually corresponds to 'Trentino' and vice versa).

Several of the dialectal groups mentioned in the key (nos. 1–12) include distinct regional dialects. The map provides the points of inquiry of the AIS (*Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale*), the concrete distribution of which is neatly illustrated by fig. 5.

Various elements on the map can be refined with the aid of the bibliographical supplement of the *Lessico Etimologico Italiano* (LEI, LEISuppl), which offers what is currently the most detailed structural overview of the dialectal varieties of Italo-Romance, including the inquiry points of the AIS.

The recent dialectometric version of the AIS, realised by Hans Goebel, further specifies the distribution of dialect groups (Goebel, *La dialettometrizzazione integrale dell'AIS*, 2008, cf. the overview provided by the map on p. 101).



Fig. 6: Geolinguistic structure of Italy according to the AIS

Source: *Carta dei dialetti d'Italia* (Pellegrini, 1977), adapted from the reproduction in LRL, vol. 4.

The dialects of Italy are highly diverse (*cf.* fig. 6). As has already been mentioned, the various dialect groups can be divided into three (or even four) large geolinguistic entities: i) northern dialects, ii) Tuscan and central dialects, iii) central-southern and iv) extreme southern dialects (*cf.* 3.3). With the exception of Venetan (= dialects of the Veneto), the northern dialects share many features with Gallo-Romance dialects (*ibid.*); central dialects are spoken in Tuscany, but also in Corsica and southern Sardinia; this group is close to the ‘median’ dialects spoken in the northern part of the Marche and Lazio, as well as in Umbria. The southern dialects are further divided into central-southern (*alto-meridionali*) and extreme southern dialects.

This last division is indicated on Wartburg’s map (fig. 4) by the limit of Latinisation during Antiquity, which around AD 200 had not yet reached the extreme south of the Italian peninsula, which was Greek-speaking. Before the fall of the Roman Empire, this part of Italy was also Latinised; however, in the regions of Apulia, Calabria and even Sicily, a number of Greek-speaking villages persisted. Their vitality was reinforced during the Byzantine period between the 6th and 9th centuries (*cf.* Fanciullo, *Fra Oriente e Occidente*, 1996).

The internal linguistic diversity of the three (or four) main dialect groups is notable, and there is a particularly strong contrast between the northern (‘Gallo-Italian’) dialects and the others. This diversity is currently stable owing to the vitality that these traditional varieties still enjoy today, and they thus constitute an exceptional area of research – in phonetics, morphology, syntax, as well as lexis.

- Bec 1, p. 15–184
- LRL 4; 2/2, art. 126–136
- RSG 1, art. 68; 82a/b; 3, art. 230–233
- OxfGuide 13–16
- MRL 13 (*Manuale di linguistica italiana*, ed. Lubello)

3.4.6 Sardinian

1. The recent elaboration of Sardinian has resulted in a partial standardisation of the language. Nevertheless, native speakers today use a dialect, rather than a standardised variety, as their mother tongue. There are still a considerable number of speakers, surpassing one million individuals, although the younger generation follows the same tendencies as those observed in the Italian peninsula, its speakers shifting to the standard Italian language. Thus, despite Sardinia’s current autonomous status, the future of the Sardinian language is uncertain.

Sardinian dialects can be divided into two main groups: the Logudorese group spoken in the north-central region (4a in fig. 4), and the Campidanese group in the south (4b). Internal differences and variation within these groups are considerable; in the Logudorese group, Nuorese (in the east) stands out in particular.

2. Sardinian is in continual contact with Italian, which acts as an ‘umbrella language’ covering the traditional territory of Sardinia. It is not Sardinian but a dialect known as ‘Gallurese’ that is spoken in villages in the extreme north of the island, a variety close to southern Corsican (and thus to Tuscan). It is also close to Sassarese, itself a Tuscan variety containing some Genoese features, spoken on the north-western extremity of the island (the delimitation is indicated by a dotted line on Wartburg’s map). Alghero, on the northwest coast of Sardinia is a linguistic enclave, the result of Catalan domination during the Middle Ages (cf. 3.4.10 no. 1). Virtually all inhabitants of Sardinia speak Italian either as a first or second language.

3. Sardinian was first put into writing during the Middle Ages, though like Francoprovençal and Gascon, it remained restricted to legal and administrative texts. After the Byzantine period, during which it enjoyed a first semi-autonomous phase, medieval written Sardinian developed under the umbrella language of Catalan, following the Aragonese conquest of the island (1326). Catalan was subsequently replaced by Spanish in the 17th century, and it was not until the 18th century that the latter was replaced by Italian (1720). From the 15th century on, the Catalan period witnessed the appearance of a body of literature, which was for the most part religious in nature. Today, literature written in Catalan is abundant and of high quality.

The formation of a standard variety of Sardinian encountered many complications due to the great diversity of regional dialects, as was also the case for Occitan or Romansh. Today, Sardinian is written in two main dialectal forms, Campidanese and Logudorese, depending on the author’s origin. Although speakers are aware of the independence of Sardinian, its linguistic elaboration remains partial.

→ Bec 2, p. 263–304

LRL 4, art. 287–292; 2/2, art. 137

RSG 1, art. 69; 3, art. 234

OxfGuide 17

MRL 15 (*Manuale di linguistica sarda*, eds Blasco Ferrer/Koch/Marzo)

3.4.7 Romansh

1. Romansh is the only language that is native to Switzerland alone, which, moreover, is a plurilingual country (cf. 3.4.1 no. 2). There, the language is most commonly referred to as Rhaeto-Romance, a name which will be avoided here as it is often used imprecisely in linguistic discussions to designate a group of three distinct languages – Romansh, Ladin and Friulian (cf. 3.4.8 below).

Romansh is composed of five regional dialects that correspond to the five main valleys in the canton of Grisons: *Putér* in the Upper Engadin Valley, *Valláder* in the Lower Engadin Valley (including the Val Müstair), *Sursilvan* in the Vorderrhein region, *Sutsilvan* in the Hinterrhein valley and, finally, *Surmiran* in the Albula and Julia valleys.

There are practically no native speakers of the standard variety, *Rumantsch Grischun* (cf. no. 3 below). Today, about a third of its 60,000 native speakers live outside of the canton of Grisons and are to be found in Zurich in particular, where Romansh has a strong presence within academic settings. In Grisons, the highest concentration of speakers is found in the region of the Vorderrhein and Lower Engadin.

Even though the number of younger speakers is slightly declining, Romansh has received strong institutional support owing to its elevation to the status of national language (in 1938) and its role as an administrative language (at cantonal level since 1880, and at federal level since 1996). Its literary and cultural activity also ensures its continuing vitality.

2. Since the integration of the canton into the Swiss Federation in the 19th century, the inhabitants of Grisons have found themselves subject to increased linguistic dominance by Swiss German. Today, standard German and dialectal varieties of Swiss German are the most widespread languages in the canton, and practically all Romansh speakers master standard and/or Swiss German. In addition, Italian is also present as an official language together with Romansh in the canton of Grisons (in the southern valleys of Val Mesolcina, Val Bregaglia and Val Poschiavo), and there are a significant number of speakers of Italian origin in the Romansh-speaking zones.

3. From the 16th century onwards, a considerable tradition of written texts began to emerge, in Engadinian (*Valláder*) and Sursilvan in particular. In 1982, the Romance linguist Heinrich Schmid (1921–1999) developed a written modern language, *Rumantsch Grischun*, at the prompting of the *Lia Rumantscha*. This language, embodying a neutralised form created from different dialect varieties, has been taught in some schools in Grisons since 2007. On the other hand, it lacks centuries of habitual use, and the internal differences between the five regional dialects of Romansh are so great that native speakers do not always find it convenient to use this standard language. Engadinian and Sursilvan maintain a true dynamism as written regional languages alongside *Rumantsch Grischun*. In particular, they play a role in a diverse array of recent literary works.

→ Bec 2, p. 305–355 (includes Lad. and Friul.)

LRL 3, art. 226–233; 2/2, art. 125

RSG 1, art. 67 (includes Lad. and Friul.); 3, art. 229a/b

OxfGuide 12

3.4.8 Ladin and Friulian

Ladin encompasses the dialects spoken in five valleys of the Dolomites, particularly around the Sella massif (Val Gardena, Val Badia, Val di Fassa, Livinallongo, Cortina d'Ampezzo), while Friulian refers to a group of dialectal varieties spoken in the Friuli area.

These two languages, spoken in northe-astern Italy, have a similar external linguistic history, and, together with Romansh, they share a number of internal features (e.g. the retention of post-consonantal *l* and the Latin diphthong *-au*, plural forms ending in *-s*, the palatalisation of /k/ before /a/ into /tʃ/ and the diphthongisation of stressed Latin vowels /e/ and /o/ in open syllables, cf. 6.5).

Returning to the terminological question of 'Rhaeto-Romance', the Italian comparative philologist Graziadio I. Ascoli (mentioned above in 3.4.4 no. 3; cf. also 2.2.3) first recognised the characteristics shared by the three languages. His ideas were taken up by the Austrian Romance specialist Theodor Gartner (1843–1925), who assumed that the three groups of varieties formerly formed a natural linguistic unit, which he termed 'Rhaeto-Romance' (not in reference to the ancient population of the *Raeti*, but rather to the Latin name given to Grisons by the Romans, *Raetia*). Conversely, the Italian linguist Carlo Battisti (1882–1977) and the Swiss linguist Carlo Salvioni (1858–1920) defended the hypothesis that these languages are genetically related to the neighbouring Italo-Romance dialects. This, in turn, paved the way for the idea of the political and linguistic dominance of Italian over the three Alpine languages. The ensuing '*questione ladina*' included political discussions between Italy, Austria and Switzerland, and remains a matter of debate today.

The issue is thorny: typologically speaking, internal and external similarities between the three languages of this *Eastern Alpine Romània* justify their treatment as a group, lending them an intermediate position between northern Italian and Gallo-Romance dialects. From a genetic point of view, it seems reasonable to suppose that until the end of the 10th century, these three geolinguistic areas were part of a linguistic mosaic, which encompassed all of northern Italy with the exception of the Venice region (close to Byzantium and Dalmatia). From the 11th century onwards, the language groups of the *Eastern Alpine Romània* progressively detached themselves from northern Italy, resulting in the evolution of their current linguistic features.

1. The current number of speakers of Dolomitic Ladin is low (ca. 30,000), but in contrast to other minority Romance languages, it is slightly on the rise, owing above all to children learning it as a first language. Nevertheless, Ladin remains endangered.

Friulian is more widespread, even though native speakers of the new generations are falling in number. With more than half a million speakers today, Friulian is currently relatively stable. The two languages have official status in Italy, and are protected by various legal and administrative provisions.

2. Speakers of Ladin are for the most part trilingual (Ladin, a South Tyrolian variety of German, Italian), and all are at least bilingual. Speakers of Friulian are bilingual (Friulian, Italian).

3. Use of Friulian in writing is relatively limited, even though its textual tradition goes back to the 14th century for documentary texts. Ladin, which emerged as a written lan-

guage in the 16th century, has a more developed tradition of modern literature. The use of the language in education and administration in the Trentino-South Tyrol region has allowed the elaboration and diffusion of a written exemplary variety known as *Ladin Dolomitan* (or standard Ladin) modelled on the example of Rumantsch Grischun. The use of Friulian in writing is supported by its rather weak degree of internal variation and by the exemplary variety spoken in Udine, the historical capital of the Friuli region.

→ cf. Romansh

LRL 3, art. 218–225; 2/2, art. 124 (Lad.); 3, art. 210–217; 2/2, art. 123 (Friul.);

RSG 3, art. 228 (Lad.); 227 (Friul.)

OxfGuide 11 (Lad.); 10 (Friul.)

MRL 26 (*Manuale di linguistica ladina*, eds Videsott/Videsott/Casalicchio)

MRL 3 (*Manuale di linguistica friulana*, eds Heinemann/Melchior)

3.4.9 Spanish

1. Spanish is by far the most widespread Romance language today, with approximately 47 million speakers in the Iberian Peninsula, at least 430 million in Central and South America and about 40 million in the United States (cf. 3.2).

2. Apart from Spanish and Basque, there are several Romance languages that are associated with specific areas within Spain: Galician and Catalan, the Aranese variety of Gascon, Asturian and Aragonese. The autonomous status of these regions being firmly established, Catalan, Basque and Galician are co-official languages in their respective territories (cf. 3.4.10 and 3.4.11 below and 3.4.3 above).

Several examples of traditional, sometimes very pronounced bilingualism can be observed in Spanish-speaking countries of the *Romània nova*. These include Paraguay, where the indigenous language Guaraní not only enjoys the same official status as Spanish, but also has a considerable number of speakers, Bolivia (where Spanish coexists with Aymara), Peru (where it coexists with Quechua), as well as Guatemala and Ecuador. Despite the fact that the Spanish expansion in the Americas brought about a precipitous decline in pre-Columbian populations (cf. 10.5.1), the number of contact languages remains quite high in some cases, such as in Mexico (which has 54 officially registered indigenous languages). In the United States, on the other hand, Spanish remains a minority or dominated language, in contact with English. Puerto Rico is a particular case since it remains politically attached to the United States, though its population is primarily Spanish speaking.

3. Spanish was subject to linguistic neutralisation from the medieval period onwards, due to intense population mixing during the Arab-Berber conquest of the Iberian Peninsula and the ensuing *Reconquista* by Christian kingdoms and principalities. Effects of homogenisation on the spoken language considerably facilitated the elaboration of

the written language. In comparison to Gallo- and Italo-Romance languages, written Spanish has changed less since the Middle Ages; by the 13th century, it had attained a stage of elaboration comparable to that which French or Italian would only reach in the 15th or the 16th century. We will return to this question in the corresponding section on external history (*cf.* 10.4).

In the modern period, the diffusion of Spanish as a written language in the Americas and other Spanish-speaking territories was often accompanied by linguistic variation (mostly grapho-phonetic and lexical in nature) and numerous situations of linguistic interference involving contact languages.

Spanish literature has undergone a considerable development since the Middle Ages, catalysed by numerous cultural contacts and by the influence of the Iberian kingdoms and empires. It is characterised by an explosive increase in texts belonging to all genres, ranging from medieval scientific texts translated from Arabic in the ‘Spain of three religions’ of the Early and High Middle Ages and the mystical poetry of the 16th century to contemporary colonial accounts written by semi-literate authors and the rich Latin American literature of recent decades.

4. Even today, the geographical distribution and internal characteristics of Romance varieties on the Iberian Peninsula reflect the effects of the Muslim conquest of Spain in the 8th century. This conquest resulted in a weakening of the regional dialects that were in the process of formation in roughly four fifths of the peninsula, with the exception of a fringe in the north, reaching from Galicia to northern Catalonia. The Christian *Reconquista*, which had just as great an impact, brought about important migrations between the 9th and 13th century, by means of which the dialectal variety of Castilian, then spoken

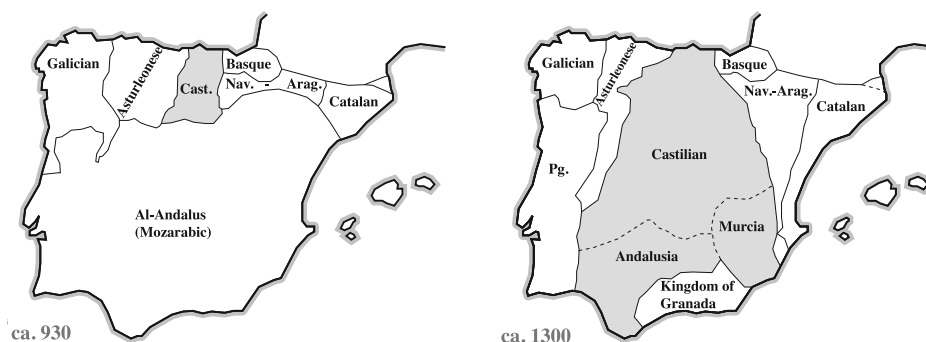


Fig. 7: Geolinguistic transformations in the Iberian Peninsula after the *Reconquista*

Source: translated from Dietrich/Geckeler 1990: 165; *cf.* the corresponding map by Bossong (2008: 322), as well as the detailed map by Martínez González/Torres Montes, RSG 1, art. 73, p. 859 (‘Expansión del castellano en la Península Ibérica’).

in a few Cantabrian valleys, between Santander, Burgos and La Rioja, was diffused over more than half of the peninsula (cf. 10.4.1 no. 1 for a more detailed description). These geolinguistic developments are represented on the two sketched maps above (fig. 7).

Like Castilian, Galician and Catalan also extended toward the south, with Galician reaching the Algarve (as a weakly diversified form of Gallego-Portuguese in the beginning), and Catalan reaching Alicante (= today's Valencian variety).

The terms *castellano* and *español* can be used interchangeably in most situations, although it should be understood that the first is more widely used in Latin America as a means of differentiation from the Spanish spoken in Spain. In this manual, 'Castilian' will only be used for the old regional dialects of Castile, while 'Spanish' will be used in all other cases.

As mentioned above, early Castilian was neutralised during this spectacular expansion, which would, in turn, bring it all the way to America. Outside their region of origin in northern Spain, Castilian varieties have thus never experienced the type of variation that can be observed in regional dialects of French, Occitan or Italian. The varieties resulting from the expansion of Spanish are therefore characterised by relatively low dialectal variation. They are not 'primary' dialects, which emerged as a result of the hereditary transmission of spoken Late Latin, but 'secondary' dialects, which emerged from a primary dialect through the neutralising effects of migration (cf. 4.2.3).

A particularly salient case among the secondary varieties of Spanish is Judeo-Spanish, which took shape during the Spanish Inquisition, after the expulsion of the Jewish population by 'Catholic' kings in 1492, and became the traditional language of the Sepharadi Jews. This variety, severed from its language of origin, was first written in Hebrew script and is characterised by a sizeable number of Hebrew loanwords. Its diffusion at the beginning of the 20th century in the Maghreb and the Ottoman Balkans in particular was severely limited as a consequence of persecutions and the Holocaust. Today, the majority of its tens of thousands of current speakers live in Israel.

Considering the enormous diffusion of Spanish, there are surprisingly few creole languages with a Spanish lexical base. The most widely spoken varieties include Chabacano, a group of creoles in the Philippines (which were under Spanish rule from 1565 to 1898) with more than half a million speakers. Papiamentu, the national language of the Caribbean islands of Aruba, Bonaire and Curaçao, is derived from a Portuguese-based creole, which was later Hispanicised but was also influenced by Dutch, and is spoken by at least 250,000 individuals. Finally, Colombian Palenquero, formerly a language of escaped slaves, is gravely endangered.

Let us return to the primary dialects, which are situated in the northern part of the Iberian Peninsula where Castilian dialects proper survived until the beginning of the 20th century. In addition, other Romance languages of northern Spain – Astur-Leonese to the west of Castilian, as well as Navarrese and Aragonese in the east (cf. fig. 7) – also display dialectal variation of the primary type. As in the case of the *oïl* dialects, despite attempts at revitalisation, the number of native speakers of these varieties is relatively

low (consider Aragonese, with less than 50,000 speakers). Asturian alone is healthier, as it is understood by approximately 500,000 persons and supported by official institutions (e.g. the *Academia de la Llingua asturiana*). In contrast, the vitality of the secondary dialects spoken in the south of the Peninsula is very strong, no doubt due to their close proximity to the standard language (which is based on Castilian), where more or less marked forms of Andalusian are spoken by a large portion of the population.

→ Bec 1, p. 186–310

LRL 6/1; 2/2, art. 156–159; 163–164 (includes Astur-Leonese and Navarro-Aragonese)

RSG 1, art. 73; 83–93; 100; 3, art. 245–249

OxfGuide 22

OxfEnc, *Judeo-Spanish (Judezmo, Ladino)*

MRL 14 (*Manual de lingüística española*, ed. Ridruejo)

MRL 20 (*Manual del español en América*, ed. Eckkrammer)

3.4.10 Catalan and Valencian

1. Catalan consists of two groups of varieties: Eastern Catalan, which includes Barcelona as a focal point, as well as the varieties spoken in the Balearic Islands, and Western Catalan, which extends southwards and thus comprises Valencian. These different varieties are used in all contexts of daily life by about seven to eight million speakers of all generations, the majority of whom are native speakers. The number of speakers of Catalan is growing and in recent decades it has gained in stability with regard to habitual use. From a legislative point of view, central Catalan and Valencian are recognised as official languages alongside Spanish.

Catalan also has a rather limited number of habitual speakers in the French department of Pyrénées-Orientales (the historical region of Roussillon with 10,000 to 15,000 speakers) and in Alghero, an enclave in Sardinia (with 5,000 to 10,000 speakers).

2. Catalan and Valencian coexist everywhere with Spanish (or with Sardinian and Italian in Alghero). Nearly all its speakers are bilingual or have a good knowledge of Spanish. On the other hand, slightly more than half of the inhabitants of Catalan- and Valencian-speaking areas are monolingual speakers of Spanish.

3. In the Middle Ages, the Catalan language developed a regionally diversified written culture, similar to that of Occitan, albeit a good century later. Notable authors include Ramón Llull of Majorca (who lived from ca. 1232 to ca. 1315) and Ausiàs March of Valencia (ca. 1397 to 1459). Due to the period of Aragonese domination, Catalan also played a role in Sardinia and southern Italy. As with Occitan, the written use of Catalan was greatly reduced in the 16th century after the personal union of the Castilian and Aragonese crowns (in 1479). Nevertheless, attempts to revive Catalan during the 19th

and at the beginning of the 20th century (*renaixença*) led to more favourable results and succeeded in saving this language, which at one time had seemed doomed to oblivion.

This success, however, failed to prevent a situation of identity conflict between Barcelona and Valencia, which manifests itself in the two standardised varieties of the language: Catalan proper, elaborated by the *Institut d'Estudis Catalans*, based on the orthography of Pompeu Fabra (1868–1948), and later Valencian, regulated by the *Acadèmia Valenciana de la Llengua* since its founding in 1998.

4. As previously mentioned, dialectal variation displays a first split into ‘eastern’ varieties (= central Catalan, the varieties of the Balearic Islands, as well as of those Roussillon and Alghero) and ‘western’ varieties, which extend from Lleida (Lérida) to Valencia in the south. The Valencian variety emerged as a result of expansion during the *Reconquista*, and thus displays the characteristics of a secondary dialect (*cf.* fig. 7 above). The northern varieties of Catalan (both eastern and western) exhibit a more marked variation characteristic of primary dialects.

The different varieties of Catalan and Valencian remain mutually intelligible without any difficulty, even though Spanish speakers in particular easily notice the differences between eastern and western forms, the latter being closer to Spanish.

→ Bec 1, p. 463–508

LRL 5/2, art. 349–358; 2/2, 154–155

RSG 1, art. 72; 3, art. 241–244

OxfGuide 21

MRL 25 (*Manual of Catalan linguistics*, eds Argenter/Lüdtke)

3.4.11 Galician

1. and 2. Galician is spoken exclusively in Galicia, in the north-western corner of the Iberian Peninsula. It has had the status of an official language alongside Spanish since 1981; its approximately 2.7 million speakers are nearly all bilingual.

3. At a relatively early stage of the *Reconquista*, Galician was exported towards the south, reaching the Algarve in 1248. Consequently, the written medieval language of Galician was not fundamentally distinct from Portuguese. This language group, frequently referred to as ‘Gallego-Portuguese’ or ‘Galician-Portuguese’, is particularly well known for its troubadoursque poetry. The spoken varieties of the two languages began to diverge after their political separation in the 12th century, and the gap widened as Galicia fell into the hands of the Spanish crown after the treaty of Alcáçovas (1476). Had it not been for these political circumstances, it is likely that the relationship between Portuguese and Galician today would be much closer, comparable to that between Catalan and Valencian.

After the subjection of its territory to Spanish rule, Galician was no longer used in writing, similarly to the way in which Catalan was replaced by Spanish or Occitan and Gascon by French. As with Catalan, attempts to revitalise Galician in speech and writing since the end of the 19th century (*rexurdimento*) were successful, despite a lower number of speakers. A final common point between Galician and Catalan is their involvement in significant identity conflicts, which are apparent in the form of lively debates on orthography, for example, pitting the Spanish and Portuguese models against each other.

4. Regional dialectal varieties of Galician evolved from spoken Late Latin and are thus primary dialects. They display more or less marked divergences from the standardised language, giving rise to a complex situation as regards oral communication.

→ LRL 6/2, art. 410–417; 2/2, art. 160

RSG 1, art. 74; 3, art. 250a/b

OxfGuide 23 (includes Portuguese)

MRL 19 (*Manual of Galician linguistics*, eds Sousa/González Seoane)

3.4.12 Portuguese

1. Portuguese is currently spoken by about 10 million speakers in Portugal, more than 215 million in Brazil and more than 10 million in Africa (Angola, Mozambique). In addition, there are several million speakers of Portuguese as a second language in Africa, as well as Portuguese immigrants in Europe and northern Africa. The number of speakers in Asia, once sizeable, is very low nowadays, with the exception of East Timor, where, arguably, as many as 100,000 speakers are still to be found. This makes Portuguese the second most widely spoken Romance language after Spanish.

2. Portuguese is the only Romance language whose country of origin (Portugal) is monolingual, at least in terms of regional languages. This is once again a distant effect of the medieval *Reconquista*, as the expansion of Gallego-Portuguese created a relatively extensive area of linguistic homogeneity.

In Brazil, the strong regression of indigenous languages since the 18th century has caused a regrettable reduction in speaker numbers. Nonetheless, over 150 different languages have been identified within the country, most of which are critically endangered.

As previously mentioned (cf. 3.2), in Guinea-Bissau, the Cape Verde islands and São Tomé-and-Príncipe, Portuguese functions as the official language alongside various Portuguese-based creoles that have the status of recognised regional languages.

3. When Portuguese was first put into writing in the Middle Ages, it was essentially indistinguishable from Galician; nevertheless, the majority of texts were written in

the area of present-day Portugal. During the modern period, Portuguese underwent phases similar to those experienced by Spanish, with regard to both its extra-European expansion and its textual elaboration (consider, for instance, the poet Luís de Camões, 1524–1580). Today, there are two standardised forms of Portuguese, that of Portugal and that of Brazil, which have fairly major phonetic and lexical differences, but relatively homogeneous grammar as well as a strong tendency towards orthographic unity.

4. The fact that Portuguese emerged as a variety of Galician following the expansion of the latter explains why its dialectal varieties are exclusively of the secondary type. Their diversity is thus weak and they are all close to the standard language. As is the case in southern Spain, local varieties coexist without great difficulty alongside the written language.

In Brazil, regional variation is also present, but it does not follow the same pattern as in Europe, due to strong migratory movements throughout the centuries.

→ Bec 1, p. 311–393

LRL 6/2, art. 418–457; 2/2, art. 161–162

RSG 1, art. 74; 94–96; 100; 3, art. 251–255

OxfGuide 23 (includes Galician)

MRL 16 (*Manual de linguística portuguesa*, eds Martins/Carrilho)

MRL 21 (*Manual of Brazilian Portuguese linguistics*, eds Kabatek/Wall)

3.4.13 Romanian

1. Today, Romanian is the national language of Romania and Moldova with ca. 21 and 3 million speakers, respectively, in addition to about a million speakers in neighbouring countries (such as Ukraine, Serbia and Hungary) and several million emigrants around the globe.

Wartburg's map identifies three further varieties of Romanian in the Balkans, mostly in Greece, at the beginning of the 20th century: Aromanian and Megleno-Romanian, as well as Istro-Romanian in Istria. These varieties have lost a great number of speakers and are critically endangered, though there are arguably still as many as 150,000 speakers of Aromanian, which is considered by recent research to be a language in its own right.

2. Romania was a multilingual and multicultural region during the medieval period, a status which, regrettably, could not be maintained due to the tragic eradication of Hungarian, German and Romani speakers under the dictatorship of Nicolae Ceaușescu in the years between 1965–1989 and the massive emigration in the years following his fall. Thus today, the vast majority of Romanians are monolingual. In Moldova, the use of Russian, which was formerly an official language, has declined since the country

acquired its independence in 1991, even though monolingual speakers are still numerous there.

3. In medieval Romania, Old Church Slavonic was used as a written, scholarly language. No texts written completely in either Romanian or Latin exist for that period, but a large number of fragmentary Romanian elements (lexemes, toponyms and anthroponyms) are to be found in Slavonic texts (*cf.* Bolocan, *Dicţionarul elementelor româneşti din documentele slavo-române (1374–1600)*, 1981). The first complete works written in Romanian did not appear until the 16th century. The language was first written in Cyrillic script; standardisation in Latin characters did not begin until as late as 1860, intensifying at the end of the 19th century, influenced by French and Italian in particular. Moldova abandoned Cyrillic script writing even later, when it became independent in 1991.

4. The remaining Romanian-speaking groups south of the Danube and in Istria point to the importance of migratory movements within this population, which was characterised by its peripatetic culture during the Middle Ages. Thus, like the majority of the Spanish-speaking population of the Iberian Peninsula, the territory of Romania (in which Daco-Romanian is spoken) seems to be the result of a migratory movement. The subject remains under debate, though it seems certain that the zone situated south of the Danube was more intensely Latinised during the period of the Roman Empire than today's Romania and that Slavonic invasions during the 6th century subsequently reduced the Romanian presence north of the Danube. This medieval Romanian-speaking population seems to have been, at least partially, the result of migrations towards the north beginning in the 11th century (*cf.* 10.3.2; 10.4.1 no. 1).

Consequently, geographical variation in Romania and Moldova is relatively limited and can be compared to that of the secondary dialects spoken in the Iberian Peninsula. The traditional varieties spoken in Moldova also differ only slightly from those of the former region of Moldavia in Romania, with which they form a geolinguistic continuum.

→ Bec 2, p. 133–262

LRL 3, art. 165–206; 2/2, art. 121

RSG 1, art. 66; 3, art. 221–225

OxfGuide 8

MRL 9 (*Manuel de linguistique roumaine*, eds Dahmen/Munteanu)

3.5 Aspects of intra-Romance typology

3.5.1 'Languages by distance' and 'languages by development' in the *România*

Now that we have presented the Romance languages, we have an empirical basis upon which to address the distinction between 'language' and '(regional) dialect'. Consider-

ing the Romance languages individually highlights the extent to which these abstract criteria are relative, as well as the fact that in real situations, varieties are situated on a continuum, and there are many possible stages between distance and development.

→ Pountain, *Standardization*, OxfGuide 37

1 Fully-elaborated languages by distance

A combination of the two criteria of distance and development are essentially present in the cases of the five national languages of Portuguese, Spanish, French, Italian and Romanian, which were subject to elaboration from an early stage. French and Italian display both notable internal distinctiveness and a long tradition as written languages. Romanian is particularly remarkable owing to its significant internal distance from all other Romance languages, although it displays a lower degree of elaboration due to the fact that the existence of written evidence is a relatively recent phenomenon, as is its standardisation. Spanish and Portuguese both have a tradition of elaboration that goes back to the Middle Ages, but they are very close to one another in terms of internal features – a factor which, in turn, reduces their degree of autonomy.

The group of Catalan varieties also belongs to this category, since the two standard forms, Valencian and Catalan, are used extensively by a large number of speakers today (*cf.* 3.4.10).

2 Partially-elaborated languages by distance

This situation applies to Romansh and Galician, both of which are national, co-official languages. Romansh displays distinguishing internal characteristics and its standard variety has been elaborated systematically, fairly recently; nevertheless, the elaboration remains partial and the use of the language is restricted, even within its natural territory. Moreover, the standard variety of Romansh is at a disadvantage owing to its considerable distance from the regional dialectal varieties, which are firmly rooted both in writing and in speech.

Galician only became distinct from its daughter language, Portuguese, at the end of the medieval period, under the influence of Spanish. Its standard form is recent and, similarly to Romansh, notably divergent from dialectal varieties. Given its structural proximity to Portuguese, Bosson (Die romanischen Sprachen, 2008) even classifies it as an ‘elaborated dialect’ (see no. 4 below) of Portuguese among the ‘languages by development only’ (see no. 3 below).

This second category of varieties, characterised by their partial or weak degree of elaboration, also includes the remaining six languages presented above: Occitan, Gascon, Sardinian, Francoprovençal, Friulian and Dolomitic Ladin. If historical evidence is taken into consideration, Dalmatian also belongs to this group. Several remarks can be made with regard to these languages: medieval Occitan was a written language

in its own right, earlier than Catalan, Italian or Spanish; however, the tradition of Occitan writing languished considerably during the 16th century and attempts at restoring it during the 19th and 20th centuries were not particularly successful. In addition, the spoken use of Occitan declined over the course of the 20th century. Gascon, though genetically a distinct language, shares a large number of internal characteristics with Occitan and has a similar external history. Though it has a larger speaker base than Occitan today, it does not display a higher degree of elaboration, and awareness among speakers that their language is distinct from Occitan is practically nonexistent.

From an internal point of view, the type of autonomy (in Stewart's terminology, cf. ch. 3.1) displayed by Sardinian is similar to that of the national languages; however, external factors are lacking, mainly because its written tradition and its use in the public sphere remain limited. The situation with regard to Friulian and Ladin is similar: although they have marked internal characteristics and are supported by a well-developed awareness among their speakers, they are still in the course of elaboration in terms of their use, both in writing and in everyday life, as well as in terms of standardisation.

Finally, the elaboration of Francoprovençal in the 13th and 14th centuries remained partial as far as textual varieties were concerned, and was interrupted by French in the 15th century. Since then, the role of Francoprovençal has essentially been limited to that of a spoken language. It thus represents the most clear-cut case of a language by distance with no real elaboration.

3 Languages or varieties by development only

From an internal perspective, these varieties are very close to another well-established distance language; they have, however, undergone at least partial elaboration, mainly for political reasons. This applies, for instance, to Moldovan, the official language of Moldova, which is historically identical to the Moldavian dialect of Romanian spoken in Romania, but has developed some differences from the latter due to the political annexation of Moldova to the USSR and its later independence. Differences are arguably more marked between varieties of French spoken and written in Switzerland or Belgium and those of France. Romance languages in the Americas are even more illustrative examples: Portuguese in Brazil, Spanish in the nineteen Latin-American countries mentioned previously, and French in Quebec. Although these varieties have distanced themselves from their ancestors in Europe, this has not resulted in problems of mutual comprehension, nor have they evolved into separate languages.

This category also includes regional dialects of the primary type that have known some degree of elaboration, which, in turn, distinguishes them from dialects that are only spoken. Asturian and Aragonese, spoken in the northern Pyrenean fringe of Spain, belong to this group: these two varieties have undergone some elaboration in writing and their respective territories have enjoyed relative political autonomy. Their internal distance with regard to Spanish, however, is limited in comparison with the situation

described for Sardinian or Occitan, as is their degree of elaboration. The same goes for the elaborated dialect of Corsican, mentioned above in relation to Italian/Tuscan (cf. 3.4.5 nos. 3 and 4), for Aranese in Spanish territory (compared with Gascon in French territory, cf. 3.4.3), and for varieties such as Judeo-Spanish (cf. 3.4.9).

4 Regional dialects and ‘elaborated dialects’

The classification of languages according to the criteria of distance and development leaves a number of Romance varieties unaccounted for, despite the fact that these show some degree of autonomy. This concerns the primary Romance regional dialects in particular, which have developed highly diversified phonetic, lexical and (to a lesser degree) grammatical features. Since they rely more or less directly on a hereditary evolutionary process from spoken Late Latin, their individual history goes back at least 1,500 years. As mentioned above, a significant number of urban regional dialects have even undergone a degree of elaboration, catalysed by the production of literary and documentary texts, and can therefore be considered to be ‘elaborated dialects’ (German *Kulturdialekte*). However, from the 9th to the 19th century there were almost as many regional dialects as there were inhabited centres in the present-day areas of Italy, France and northern Spain. Due to their great number and their internal proximity to immediately neighbouring varieties, they cannot be classified on the basis of the criterion of linguistic distance.

5 Languages and dialects in the *România*: reprise

This brief overview shows that a relatively independent language is not easy to differentiate from a dialect or variety with a lesser degree of autonomy. There are many transitional situations along this continuum of autonomy, ranging from fully autonomous languages to partially autonomous or elaborated languages to regional dialects displaying little or no elaboration, though varieties that are less marked are certainly no less interesting. In general, a study of all Romance varieties contributes to a better understanding of language change and language functioning.

The number of varieties considered to be actual ‘languages’ by authors of Romance manuals has increased continually since the start of the 19th century, beginning with five literary languages in 1818, four of which ‘are cultivated (...) today’ (that is, in Schlegel’s time: ‘l’italien, l’espagnol, le portuguais et le François’, alongside ‘le provençal’ as well as ‘une variété infinie de dialectes et de patois’), which became eight in Gustav Gröber’s *Grundriss* (1886–1906). Walther von Wartburg (1945–1971 cites nine: It., Sp., Pg., Fr., Occ., Rom.n, Sard., Romansh and Cat.), eleven are noted by Carlo Tagliavini (1952–1982: including Dalmatian and Francoprovençal in addition to the other varieties previously mentioned), while there are fourteen according to Pierre Bec (1970–1971). The LRL lists sixteen (including the aforementioned eleven languages, along with Friulian, Ladin, Galician, as well as the elaborated dialects of Aragonese/Navarrese, Asturian/Leonese

and Corsican, but excluding Francoprovençal), and, finally, approximately twenty-four are described by the *Manuals of Romance Linguistics* (cf. Müller, *Die 'Sprachen' der romanischen Sprachwissenschaft*, 1994).

The *Romània* with all its varieties, the historical development of which is comparatively well known and excellently documented, illustrates in an exemplary manner the general difficulties involved in determining the status of a speech variety, even with the aid of the very detailed distinguishing criteria presented above. For lesser-known languages and dialects – the majority of the world's language families belong to this category – difficulties are far greater. This observation alone implies that major databases that register a maximum number of languages (such as the WALS) should be used with caution.

The languages of the *Romània* also illustrate the current global tendency towards a loss of linguistic diversity (cf. 1.2.5). We have mentioned the great number of varieties that are endangered today: speakers of *oil* dialects or of Francoprovençal are very few in number, and even Occitan, Gascon and varieties spoken in northern Spain have seen their use reduced; Istro-Romanian is almost extinct, and Megleno-Romanian is critically endangered. Other varieties, such as Friulian and Sardinian, the dialects of Italy, Aromanian or even Romansh, are still very much in evidence today, though their speaker numbers are declining as well. Conversely, if a language has been standardised, as is the case of Catalan, its survival is more or less guaranteed, though this necessarily implies a decrease in its natural dialectal diversity. This loss of linguistic diversity that results from the spread of a standard variety can also be observed in part for Galician, and will soon apply to Dolomitic Ladin. It is almost complete in the case of French, the 'success' of which has essentially eliminated nearly all the old primary *oil* dialects.

In all Romance-speaking countries, the use of a dominant national language is accompanied by that of other territorial languages, which are often numerous and of a particular sociolectal and functional importance. A standard language constitutes an 'umbrella language' within the communicative space of its political state (cf. 3.1 and 3.2), but it is always only one of the multiple varieties of everyday language use. Even nowadays, the world of Romance languages remains large and diversified, containing unexplored treasures both for linguistic analysis and for reflection on questions of identity.

Since this *Companion* is addressed first and foremost to students and scholars of the most widely taught Romance languages, i.e. French, Spanish and Italian, our approach will focus on these languages and their history. Nevertheless, it will also incorporate – as far as this is possible – the other Romance languages and regional dialects presented above (cf. 3.4). Depending on each specific case, a greater or lesser number of internal or external parameters will be taken into consideration. An in-depth treatment of even these languages and varieties, however, is beyond the scope of this manual. Our objec-

tive is not to study specific segments of a language in detail, but to provide a general interpretative framework that allows a global perspective.

3.5.2 Illustrating variation among Romance languages

Part 2 of this manual attempts to describe the most salient linguistic characteristics of the Romance languages. As a first insight, a brief illustration of the similarities and differences within this language family is provided below. This time, the individual languages will be presented from east to west.

To this aim, we will consider the sentence ‘*panem nostrum cotidianum da nobis hodie*’ (“Give us this day our daily bread”) from the *Pater Noster* (Matth. 6, (5) 9–13), translated into one variety of each of the languages presented above (after Heger, *Die Bibel in der Romania*, 1967). Note that this is only a selection of specific variants among many others that may exist for the same sentence in the various languages. Even in the case of the Latin version, a choice had to be made, and the *Vetus latina* version was opted for, rather than the *Vulgate* version (the latter replaces *quotidianus* with the theological concept of *supersubstantialis*, which is present in the Friulian version). The difference between *quotidien* “daily” (Fr.) and *de cade jou* “(of) every day” (Occ.) reflects not only a linguistic, but also a theological difference. Nevertheless, this short passage is of illustrative value:

(Lat.) Latin	<i>panem nostrum cotidianum da nobis hodie</i>
(Pg.) Portuguese	<i>o pão nosso de cada dia nos dá hoje</i>
(Gal.) Galician	<i>o pan noso de cada dia dánolo hoxe</i>
(Sp.) Spanish	<i>danos hoy nuestro pan de cada día</i>
(Cat.) Catalan	<i>el nostre pa de cada dia doneu-nos, Senyor, el dia d’avui</i>
(Gasc.) Gascon	<i>baillat-nous ouey nousté pa quotidian</i>
(Occ.) Occitan (Provençal)	<i>dounà-nou vuèi nouèste pan de cade jou</i>
(Frpr.) Francoprovençal (Fribourg)	<i>le pan dè ti lè dzoa, baidè no le, ouè</i>
(Fr.) French	<i>donne-nous aujourd’hui notre pain quotidien</i>
(Rom.) Romansh (Sursilvan)	<i>nies paun de mintga di dai a nus hoe</i>
(Lad.) Ladin (Dolomitan)	<i>dànes encuei nost pan de vigni di</i>
(Friul.) Friulian	<i>dànus uè lu nèstri pan (soresostanzeôs)</i>
(Sard.) Sardinian (Logudorese)	<i>su pane nostru de ogni die danoslu hoe</i>
(It.) Italian	<i>dacci oggi il nostro pane quotidiano</i>
(Rom.n) Romanian	<i>pâinea noastră cea de toate zilele dă-ne-o nouă astăzi</i>

When considering these variants of the sentence, the great similarity between Galician and Portuguese is quite clear. The proximity of Catalan, Occitan, Gascon, Francoprovençal and French to each other is also apparent. The proximity between Italian and Sardinian is less clear, though perceptible. The differences between the three Alpine languages of Romansh, Ladin and Friulian are also recognisable, as is the distance

between Romanian, the only language that is separated from the geographical continuum of the *România* in Europe, and the rest of the family.

The following section constitutes a brief commentary on these examples, focusing essentially on their grammatical features and applying terminology that will be presented and discussed in part 3. A separate commentary is provided for each of the main constituents, i.e. the verb phrase ('give us'), the noun phrase, which functions as a direct object ('our daily bread'), and the adverbial phrase ('this day').

- verb phrase: the Latin verb DARE ("to give") has survived in most languages (It., Sard., Sp. *da-*; Lad., Friul. *dà*; Pg., Gal. *dá*; Rom.n *dă*; Rom. *dai*; Lat. *da*). In the Fr.-Occ.-Cat. group, it is replaced by a descendant of the verb DONARE, originally meaning "to give a present" (Fr. *donne*, Occ. *dounà*, Cat. *doneu*): the word has undergone a process of metonymic semantic change. In Gascon and Francoprovençal, a third term (also common in varieties of Occitan) is introduced: BAJULARE ("to carry"), which has also evolved by metonymy;
- the personal pronoun (functioning as an indirect object complement in the verb phrase) is nearly always placed after the verb (Sard., Sp., Cat. *-nos*; Gal. *-no*; Fr. *nous*; Occ. *nou*; Rom.n *-ne* and *nouă*; a complementary marker is included in Rom.: *a nus*. In Lat., it is represented by the inflectional form *nobis*; in It., one can recognise the old demonstrative pronoun *HIC(CE) > *-ci*, the grammatical function of which has changed); in Pg., the pronoun is placed before the verb (*nos da*), as a result of the non-initial position of the verb;
- the direct object is repeated in the form of verbal complements in Sard., Frpr. and Rom.n (*-lu*, *le*, *-o*);
- noun phrase: the Lat. noun PANE (most nouns in Romance languages have evolved from accusative forms in Latin; in this case from PANEM, with loss of word-final -M from the 1st century on) is present in all languages, though it has undergone different phonetic changes. The final vowel has been dropped in Gal., Sp., Occ., Friul., *pan*; Rom. *paun*, as has the nasal consonant in Cat., Gasc. *pa* and Pg. *pão* and – as far as pronunciation is concerned – in Fr. *pain*. The stressed vowel is nasalised in Pg. and Fr; it is diphthongised in Rom. *paun* as well as in O.Fr. (a > æ > ej ~ ε). Moreover, the noun's gender has changed in Rom.n (from M. to F.), a phenomenon not uncommon among the other nouns of the 3rd declension class;
- the possessive determiner is based on Latin NOSTRU (in Rom., a process of metaphony results in its diphthongisation: *nyes* > *nies* before word-final -u); in the central languages (It., Fr., Occ., Cat., Sp.), the determiner precedes the noun: $n_{\text{DET}} p_{\text{N}}$, whereas it is postposed in peripheral languages (Pg., Rom.n, as well as Sard.: $p_{\text{N}} n_{\text{DET}}$);
- when the possessive determiner is placed before the noun, the accompanying article is omitted, except in It. *il nostro pane*, Friul. *lu nèstri pa* and Cat. *el nostre pa*. When the possessive determiner follows the noun, the article remains intact (Pg.,

Gal. *o* < ILLU and It., Friul., Cat., Sard. *su* < IPSU, Rom.n. [*pâine*]-*a* < ILLA: postposed to the noun). Lat. does not use articles, these being an innovation in Romance languages;

- the qualifying adjective can be identified as a learned borrowing from Latin (i.e. a form borrowed from Latin in its use as a written language in the *România*, as opposed to a form inherited from spoken Late Latin): It. *quotidiano*, Sp. *cotidiano*, Fr. and Gasc. *quotidien* < QUOTIDIANU (*cf.* the Latin version of the sentence). In the other languages, which introduce a nominal complement, a variety of adjectives meaning “all” are used: Rom.n. *toate* < *TOTTA(E); Sard. *ogni* < Tusc. *ogni* < OMNI; Por., Gal., Cat. *cada*; Occ. *cade* < Late Lat. CATA < Gr. *katá* (“by means of”), Romansh *mintga* < Germ. **manigibô* (“large quantity”, *cf.* 9.6.2 no. 2) – originally a noun, which became an adjective due to its frequent use in pre-nominal position. Most of the languages use Lat. DIEM (M.: Sard. *die*; Pg., Cat. *dia*; F.: Rom.n. *zi*-); Occ., on the other hand, uses the nominalised adjective form *jou* < DIURNU;
- the adjunct adverbial has evolved from HODIE (see the Latin version) in most languages (once again displaying varied phonetic forms: It. *oggi*; Por. *hoje*; Gal. *hoxe*; Rom. *oz*; Sard. *hoe*; Friul. *uè*; Occ. *vui*; Cat. *avui*; Sp. *hoy*). In French it is reinforced by *au jour d’(hui)*. Rom.n. *astă-zi* corresponds to ISTA DIE, Lad. *enc-uei* to HANC DIE;
- finally, constituent order corresponds to the Latin order of object-verb in Sard., Rom.n., Pg., Cat. and Rom.; it is inverted (verb-object) in more central languages (It., Occ., Fr., Sp.).

Even without considering detailed nuances, especially those of a phonetic nature, this simple sentence, with its basic vocabulary and close adherence to the biblical Latin model, illustrates the vast array of possible variations that manifest themselves in the Romance languages.

3.5.3 The classification of the Romance languages

The presentation of the Romance languages has already shown how difficult it is to make clear-cut distinctions or to establish classifications within the *România*. This is true for both external and internal questions relating to phonological, grammatical or lexical differentiation. Even though the five or six modern standard languages appear to diverge considerably from an internal perspective, the differences are obscured if all the Romance varieties in question are taken into account, and even more so when they are considered from a historical perspective. Differentiating between Spanish, Catalan and French, for instance, becomes much more complicated if the primary dialects of early Castile, Aragon, Gascony, Catalonia, Languedoc, the Massif central, the western Alps and the Franche-Comté are also included in the comparison. Seen from this point of view, transitional situations are much more frequent than clear boundaries. Only

when intermediate forms and older stages of a language are glossed over do divisions appear clear and simple.

Consider the example of a number of forms of medieval French and Occitan, once again taken from the *Pater noster* cited above (after Heinimann, *Oratio dominica romanice*, 1988). This comparison illustrates the extent to which differences that are apparent today were less pronounced in earlier stages:

O.Fr. (British Isles, ca. 1150)	<i>nostre peïn chaske journal dune nus hoi</i>
O.Fr. (Paris, ca. 1250)	<i>nostre pain de chascun jor nos dones hui</i>
O.Occ. (Waldensian, ca. 1300)	<i>dona nos encoy lo nostre pan cotidian</i>
O.Occ. (1 st half 14 th c., ms. 15 th c.)	<i>dona huey lo pan de tos dies</i>
M.Fr. (Lorraine, ca. 1365)	<i>donne nous du jour d'ui nostre pain cottidien et de chesques jour</i>
Mod. French (1524)	<i>donne nous aujourd'huy nostre pain cotidian</i> (~ <i>quotidien</i> , 1535)

The French versions of the sentence, taken from the translations of the Bible by Lefèvre d'Étaples (1524) and Calvin (1535, *cf.* 10.5.3 no. 2), already resemble their present-day form, which, moreover, is reflected in the Middle French version from Lorraine dated some 160 years earlier. The Old French versions produced in England and Paris display notable differences in constituent order, as well as in the forms of the complement (*hui*) in comparison to the later examples given above. It is difficult to determine whether they are closer to the forms from the Lorraine Bible or to those from the Provençal and Waldensian Bibles.

The differences among Romance varieties that are visible at a specific point in history are thus relativised when the increasing divergence of the languages over time is observed (*cf.* 4.2.2 and 10.4.1 no. 4).

Until now, we have focused in particular on comparing close varieties, considering the possibility and usefulness of distinguishing them with regard to both their internal and their external status. Another fundamental question concerns the constitution of subgroups for the various representatives of a language family, in this case on the basis of internal criteria alone. In contrast to what one might expect, such a classification runs into even more insurmountable difficulties than the categorisation based on the criteria of autonomy and elaboration.

This can already be seen in the simple binary categorisation suggested by Wartburg in 1950. Wartburg distinguished the Eastern *Romània* (central and southern Italo-Romance, Dalmatian, Romanian) from the Western *Romània* (all other Romance languages, with Sardinian occupying an intermediate position between the two categories). In his model, the dialects of northern Italy belong to the Western *Romània*, in contrast to standard Italian and the dialects situated south of the La Spezia-Rimini line (*cf.* 3.3), which forms the limit between the Western and the Eastern *Romània*.

This split suggested by Wartburg is principally based on a phonetic criterion involving two phenomena – the outcome of the Latin plosives /p t k/, single and double, in intervocalic position – as well as on the morphological criterion of plural formation ending in -i (It., Rom.n *amicī*) as opposed to -s (“sigmatic plural”, as in Fr. *amis*, Sp. *amigos*, etc.).

These differences among the Romance languages can be easily observed. The following example illustrates the voicing of Latin intervocalic plosives, which involves both the series of voiceless consonants /p t k/ and voiced consonants /b d g/. Latin FORMICA contains an intervocalic /k/, which is conserved in Rom.n *furnicā*, Dalm. *formaika* and It. *formica*; in contrast, the phoneme is voiced in Sard., Occ., Cat., Pg. *formiga*, Sp. *hormiga* /ɣ/, and it has disappeared completely in Romansh *furmia* and Fr. *fourmi*. A complete list of these phonemes is provided below (cf. 6.5.2 no 5.1). The forms in our example are taken from the corresponding article of the *Romanisches Etymologisches Wörterbuch* (REW) by Wilhelm Meyer-Lübke (cf. fig. 8 below), which contains hundreds of other examples of lexemes illustrating the same split. This reference work is now dated (³1935), but it provides a succinct and thorough overview of the main lexical (and phonetic) similarities and differences among Romance languages and dialects, structured etymologically.

3445. formīca „Ameise“.

Rum. *furnicā*, vegl. *formaika*, it. *formica*, log. *formiga*, engad. *furmia*, frz. *fourmi*, prov., kat. *formiga*, sp. *hormiga*, pg. *formiga*; afrz. *formiz*, prov. *formitz*, dauph. *frēmize*, wald. *frümizi*, wallon. *fromiš*, eigentlich FORMICAE M.-L., Rom. Gram. 2, 17. — Ablt.: rum. *furnicar*, it. *formicaio*, engad. *furmier*, frz. *fourmilière*, prov. *formiguier*, sp. *hormiguero*, pg. *formigueiro* „Ameisenhaufen“; ...

Fig. 8: Example of an article in Meyer-Lübke’s REW (art. 3445)

Remarks:

The entry includes forms of the Vegliot (Vgl.) dialect of Dalmatian (spoken on the Isle of Krk), the Logudorese (Logud.) variety of Sardinian, and the variety of Romansh spoken in Engadin (Engad.), Grisons. Among variants of the alternative pattern ending in /-ts/ (explained by Meyer-Lübke as derived from the plural form FORMICAE, but interpreted differently by Wartburg in his FEW as coming from *FORMICEM), Meyer-Lübke cites variants of modern Occitan (‘prov.’, including Waldese from the Italian Piedmont), Francoprovençal Dauphinois and Walloon (the *oil* dialect spoken in Belgium). Moreover, the same variation of the intervocalic consonants affects derived forms (Germ. *Abl.* = ‘Ableitung’) such as It. *formicaio*, Fr. *fourmilière*, Sp. *hormiguero* “anthill”.

Parallel to the processes of voicing and the complete loss of these consonants, Latin double plosives in intervocalic position are degeminated (simplified): It. *cuppa* (< Lat. CUPPA) contrasts with Sp., Pg., Cat. *copa*, Fr. *coupe*. The process also reaches Romanian (*gută* < Lat. GUTTA).

Differences between plural forms (cf. 7.3.2) are also easily recognisable; the geolinguistic distribution of this phenomenon is shown on the map provided by Gerhard Rohlfs, which displays divergences among the Romance languages in visual form.

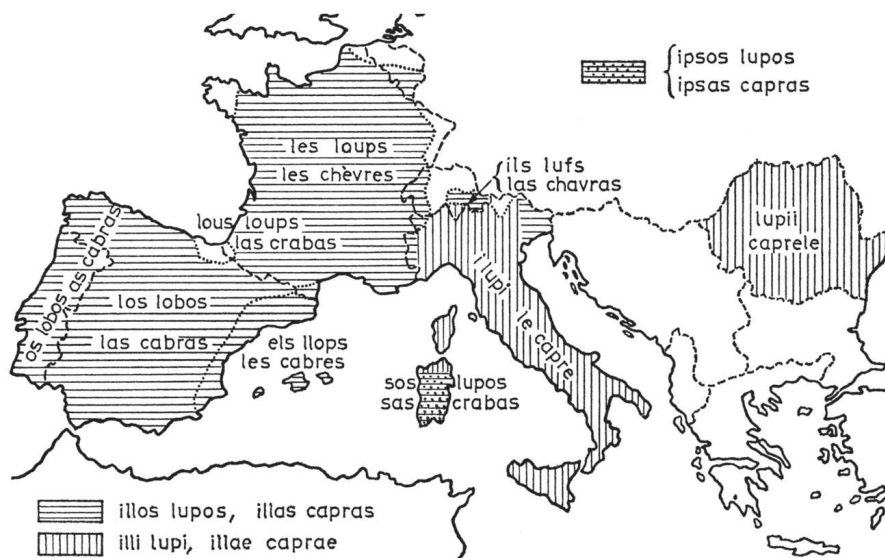


Fig. 9: Plural formation in the *România*

Source: Rohlfs, *Romanische Sprachgeschichte*, 1971, map 10, p. 245; for other maps, cf. *id. Panorama delle lingue neolatine*, 1986.

The linguistic form of the definite article represented on Rohlfs's map also serves to emphasise the particular position of Sardinian: *sos* has evolved from Lat. *IPSOS*, in contrast to forms deriving from *ILLOS* in almost all other Romance languages (cf. 3.5.2 above).

The three criteria established by Wartburg (voicing and degemination of intervocalic plosives, plural forms ending in *-s* vs. *-i*) are not completely isolated phenomena. There are a number of isoglosses based on grammatical and lexical features that group together southern Italian and Romanian (cf. 8.4.5 for the subordinating conjunction south. It. *ca*, Rom.n *că*, < *QUIA*). In general, however, the criteria are inconclusive: as we have already seen, degemination also occurs in Romanian and voicing does not affect some valleys where Béarnese (Gascon) or Navarrese (on the other side of the Pyrenees) are spoken. Furthermore, grammatical and lexical isoglosses easily emerge between two given Romance languages, as the example of the *Pater noster* in the previous section has shown.

Indeed, the two linguistic atlases by Rohlfs (1971, 1986) illustrate the highly variable nature of the solidarities that emerge among different areas within the geolinguistic

continuum of the *România* when phenomena are grouped according to shared characteristics. As an example, consider variant names for 'MARE' as a specific concept from rural life (fig. 10).

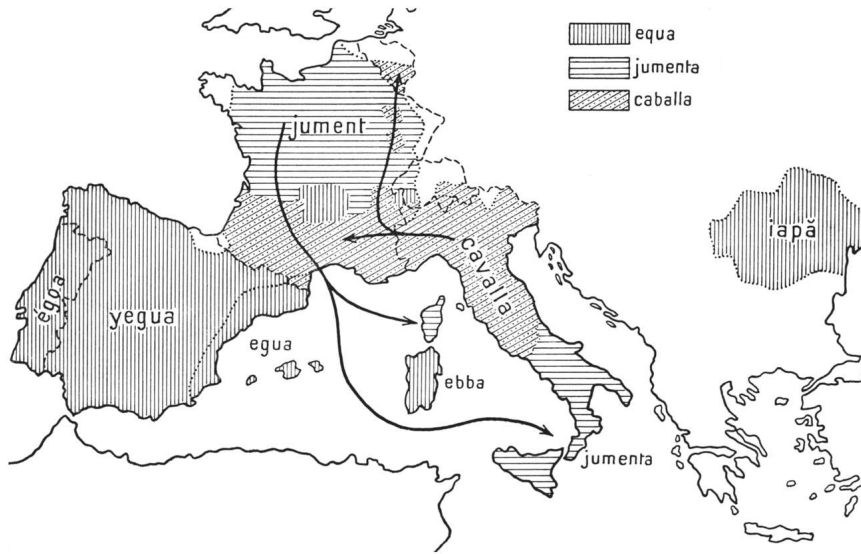


Fig. 10: Terms for 'MARE' in the *România*

Source: Rohlfs 1971, map 36, p. 271.

Remarks:

1. the habitual Latin form *EQUA* /'ekwa/ has survived in the two 'lateral areas', in Romanian (*iapă*, which has undergone the phonetic change /kw/ > /p/) and in the Iberian Peninsula (*(y)egua*, with voicing of the intervocalic plosive) as well as in Sardinian, which is conservative (displaying the regular change /kw/ > /bb/);
2. forms deriving from *EQUA* were also maintained in medieval spoken *oïl* and *oc* varieties (O.Fr. *ive*, O.Occ. *eg(o)a*). This is not shown on the map, but can be deduced from the fact that forms of *EQUA* are conserved in southern Occitan varieties and Francoprovençal;
3. variants of the Spoken Latin innovation *CABALLU* (→ *cavalla* f.) were first concentrated in a central area (northern and central Italian, Friulian, Ladin, Romansh) before spreading to southern Occitan, Gascon and even to eastern *oïl* dialects (extending as far as Walloon);
4. in French, the older type *EQUA* (like *CABALLU* in the majority of the eastern zones) was in competition with and was eventually eliminated by a word denoting "beast of burden" deriving from Latin *JUMENTU*. This innovation is also found in southern

- Italo-Romance, where it was reinforced by the French model (introduced by the Normans during the 12th century as a term belonging to courtly culture);
5. the French term was finally introduced into Corsican after the island was annexed by France in 1768. The two waves of innovation represented by *CABALLU* and *JUMENTU* thus partially overlap; they did not, however entirely eliminate previous forms (for more details, cf. Rohlfs, pp. 84–87).

Rohlfs' map illustrates the extent to which modern linguistic groupings are an effect of geolinguistic developments that have taken place over one and a half millennia, highlighting the fact that the constellations that can be observed today have no absolute value. Other maps show different results again and this diversity is confirmed by the distribution of grammatical phenomena (cf. 8.5). When a large number of internal parameters are considered, the division of the *Romània* into an innovative Western *Romània* (thought by Wartburg to be influenced by Celtic) and a conservative Eastern *Romània* loses all its classificatory and explanatory value.

The only strong linguistic solidarities that emerge from the consideration of a large number of phenomena are those between neighbouring languages, for example between Portuguese and Spanish, Catalan and Occitan or northern Italian, Romansh, Ladin and Friulian. These groupings usually function in two directions: Occitan, for instance, also displays unifying characteristics with French, Catalan with Spanish, and northe-astern Italian with Dalmatian. These are thus relative characteristics, and are not a suitable base upon which to found an accurate general classification of the Romance languages.

For strictly practical purposes, the traditional designations 'Ibero-Romance', 'Gallo-Romance' and 'Italo-Romance' can be used. These 'categories' involve both geographical aspects ('the Iberian Peninsula') and political aspects ('present-day France'). From a linguistic point of view, they are inaccurate, owing in particular to the intermediate position of Catalan between the 'Ibero-Romance' languages (Sp., Gal., Pg.) and the 'Gallo-Romance' languages (Fr., Frpr., Occ.) and to the difficulties involved in categorising Sardinian as well as Romansh, Ladin and Friulian ('the *Romània* of the eastern Alps') in relation to the 'Italo-Romance' dialects. 'Daco-Romance' (or 'South-East Romance') of today would only encompass varieties of Romanian. As for Dalmatian, it would further be necessary to take into consideration the ways in which it was influenced through contact with Venetian dialects. Caution should therefore be used when applying such designations, which can, nonetheless, be helpful for the practical treatment of linguistic data.

The above analysis has shown that it is impossible to achieve a single coherent classification of the Romance languages based on internal criteria. The reason for this is linked to their emergence: these languages all developed from spoken Late Latin, from which they all branched off around approximately the same time. Therefore, a general

hierarchical categorisation consisting of groupings by branches cannot be established. The only identifiable unifying characteristics are the effect of multiple geographical, socio-political and cultural factors.

The difficulties involved in classifying the Romance languages – a well-known language family – indicate *a fortiori* how arduous a task the classification of African, Asian or indigenous American languages would prove, considering that they are not well described and lack historical documentation (*cf.* also 1.1, fig. 1, for the Indo-European languages). Here, once again, the *Romània* offers us a lesson in humility in the face of the multifaceted complexity of language.

- LRL 7, art. 495–511, *Analyses contrastives, classification et typologie des langues romanes*
 Hoinkes, *Prinzipien der genealogischen Klassifikation der romanischen Sprachen*, RSG 1, art. 11
 Jacob, *Prinzipien der Typologie und der sprachinternen Klassifikation der romanischen Sprachen*, RSG 1, art. 12

3.6 Language contact in the *Romània*

3.6.1 Diglossia, bilingualism and language contact

Nowadays, the presence of two or more languages is the norm in almost all countries of the *Romània*, whereas monolingualism is the exception. The same can be said of almost all contexts in which language is present, both current and historical. Diglossia and polyglossia (the coexistence of two or more varieties in a single communicative context) are thus frequently accompanied by bilingualism or plurilingualism (the individual's competence in two or more languages). In societies marked by diglossia, it is mainly the speakers of minority languages who are bilingual. Their bilingualism is thus of a functional nature and their choice of one or the other language depends on speech partners and situations (*cf.* Fishman, *Bilingualism with and without diglossia; diglossia with and without bilingualism*, 1967).

The concepts of diglossia and bilingualism can be applied to different languages as well as to internal varieties of a particular language. The mechanisms at work are identical, from a cognitive, sociological and communicative point of view. As an example, a doctor from Prato will easily speak standard Italian within his professional sphere, but will switch to his native Tuscan dialect within his family circle and to English at international conferences; thus, for functional purposes, he is capable of using two distinct varieties of one language, as well as a foreign language. We will return to the question of varieties in chapter 4.

In the presentation of the Romance-speaking areas, we have already seen that language plays an important role in identity conflicts. In a diglossic society, the choice of one language over the other is never neutral and always corresponds to the assumption of a sociological and ideological position. Moreover, a poor grasp of language use and idiom-

atic expressions necessarily results in value judgements on the part of other speakers. A positive attitude with regard to one's own linguistic identity is of primary importance, both for individuals and societies. Linguistic awareness can help overcome reflexes resulting from prejudice that lead us to reject 'others'; this is one of the most important tasks of linguistics, which is all too often misused in these situations.

From a language-internal perspective, bilingualism leads to linguistic 'interference' (or, in broader terms, 'transfer'), by means of which the two (or more) linguistic systems used by a single group of individuals influence one another. The nature of the interference depends on the social prestige of the two languages in question, which is unequal in the majority of cases: the dominant language, possessing a higher level of prestige (the 'high variety' (HV) according to the terminology introduced by Ferguson, 1959) tends to exert a greater influence on the less prestigious language (the 'low variety' (LV), cf. 4.3.4). Thus, the former often functions as a 'donor language', while the second acts as a 'recipient language'.

Language interference first occurs within the brain, and then at the level of utterances. The most widely spread transfer phenomena are borrowings, which are mostly lexical, and less often of a grammatical or phonetic nature. Borrowing always depends on pre-existing interaction between two languages and implies a contact zone where they coexist, which facilitates the transfer of elements from one to the other.

Situations of diglossia often lead to so-called 'code-switching' in spoken language (as in the sentence *Sometimes I'll start a sentence in English y termino en español*, an example taken from the diglossic situation in Puerto Rico; cf. Poplack, 1980). Sometimes, specific words end up being borrowed. A successful borrowing is thus merely the salient result of multiple situations of interference, which, however, for the most part remain without any long-term linguistic consequences.

A contact zone that is large enough to induce borrowing depends on either widespread diglossia or cultural contact within restricted groups of a population. In France, for instance, diglossia exists between French on the one hand, and languages such as Alsatian, Occitan or Arabic on the other. In the same way, there is interference between Spanish and Catalan or Galician. In these examples, language transfer is reciprocal: the French spoken in Alsace and the Spanish spoken in Catalonia contain a significant number of Alsatianisms and Catalanisms, respectively; Alsatian and Catalan, in turn, borrow elements from their respective umbrella languages to varying degrees.

Cultural contact is apparent, for example, in recent borrowings from English into specialised terminologies (e.g. in the fields of music, fashion or the sciences). Borrowings find their way into the Romance languages in the form of Anglicisms, introduced by English-speaking individuals, which then become widespread among communities of speakers who are often not English-speaking (cf. 9.6.2 no. 5). The same phenomenon can be observed with regard to the introduction of Arabic medical-biological or astronomical terminology into the Western world during the medieval period.

- As an example, there seem to have been three parallel borrowings of Arabic *az-zaʿ-
farān* “saffron” into the Romance languages: (1) Sp. *azafrán*, Pg. *açafrão*, (2) Cat.
safrà, medieval Lat. *safranum* > Fr. *safran* and (3) It. *zafferano*. The Spanish-Port-
uguese type additionally includes the agglutinated Arabic article, a phenomenon
characteristic of code-mixing in speech, which respects morphemic boundaries.
- A further example is Arabic *as-samt* (*ar-raʿs*) “zénith” > Medieval Lat. *zemt*, *zenit*
> Sp. *cenit* > It. *zenit*, Cat. *zenit*, Pg. *zénite*, Fr. *zénith*. This borrowing was spread
through writing alone as it arose from a misreading error: the Romance forms
seem to be based on a false interpretation of the three minims (down strokes) of
the grapheme *-m-* as *-ni-*. This error, very frequent in medieval manuscripts, must
have been present in a copy of a Latin text translated from Arabic.

In situations of cultural contact, transfer is restricted to small groups, and borrowing is often unidirectional (e.g. from medieval Arabic to Spanish or from modern English to French but not vice versa).

Nevertheless, situations of diglossia often do not have a lasting linguistic impact: the presence of numerous immigrants in France has hardly led to the appearance of any Italianisms, Hispanisms, Lusisms (i.e. borrowings from Portuguese) or Turkisms in everyday French. Transfer phenomena are stronger in the opposite direction, though they remain confined to communities of immigrants or even to specific individuals, and such borrowings are seldom long-lived.

→ Kabatek, *Diglossia*, OxfGuide 36

3.6.2 The importance of language contact in the *România continua*

Language contact that has occurred both in the European and the extra-European *România* has given rise to a large number of temporary and permanent borrowings, mostly lexical in nature and affecting only a few clusters of concepts with similar meanings. Nevertheless, some historical events led to language mixing on a large scale that had a lasting effect.

This is the case for medieval English, which was transformed considerably due to intense contact with French and written Latin during the Anglo-Norman period of English history (cf. 10.4.1 no. 1). Consequently, the number of words of French or Latin origin in English vocabulary is very high, reaching an average of 50% depending on the textual genre and language register. Grammatical and phonetic transfer phenomena are less obvious, though they merit more thorough investigation (note, for example, English constituent order, which follows the French model S-V-O, as opposed to the order T(opic)-V-X (X being a free constituent) that characterises the Germanic languages). Hence, the study of English by Romance specialists, particularly with regard to its vocabulary, is as justified as their study of Romance-based creoles, for instance (cf. 3.6.3 below).

A comparable situation characterises Romanian, which was strongly influenced by its intense contact with south Slavonic languages from the 6th to the 17th centuries. Romanian contains at least as many lexical borrowings from these languages as it does words of Latin and English origin, and numerous phenomena of grammatical transfer serve to set this language apart from the classic Romance type. Its current physiognomy as a Romance language is partially the effect of re-Latinisation and Gallicisation processes during its elaboration in the 19th century, which reduced the number of elements of Slavonic origin.

Other mixed languages in Europe, born of contact between a Romance and a non-Romance variety, are less widespread:

- as a consequence of British rule in Gibraltar, the dialect of Andalusian origin spoken in this enclave has developed into a mixed Anglo-Spanish language, known as *yanito* (< *llano*, probably originally meaning “plain-spoken” or popular language; cf. Kramer, *English and Spanish in Gibraltar*, 1986);
- a large number of French loanwords were incorporated into the Franconian dialect of Luxembourg, today’s *Letzeburger Sproch* (cf. *Dictionnaire étymologique des éléments français du luxembourgeois*, DictEtFrLux);
- Maltese is a dialect of Arabic which has been subjected to a high degree of modification due to a very large number of Italian and English borrowings.

Of the many cultural contact languages in the *România*, written Latin has been by far the most influential during all historical periods. It has had a decisive influence on all languages derived from spoken Late Latin, from the Middle Ages, when they first began to be written and elaborated, up until the 21st century. In texts written in present-day French, Italian or Portuguese, more than a quarter of the vocabulary has been borrowed from written Latin and does not reflect hereditary phonetic change (cf. 6.6). As far as the grammar of Romance languages is concerned, it displays greater autonomy with respect to its Latin model, the latter having nonetheless left numerous traces. Continual reference to Latin has resulted in a general increase in convergent evolutionary patterns within the *România* and a parallel reduction of divergent tendencies. This is particularly true for the written Romance languages, which resemble each other much more closely than do the spoken varieties of Romance, including those that have been standardised.

It is not necessary to present the numerous situations of language contact in the *România* here, since the different types of language transfer – which we will discuss in the chapter on lexis (cf. 9.6.3) – are recurrent, though the details always vary. It should be borne in mind that languages are by no means impermeable entities; rather, they evolve in complementarity and in contact with one another, mutually influencing one another. The degree of such influence is relatively variable and depends on sociological constellations. Languages, as such, are sufficiently similar to allow for every form of transfer, even grammatical ones, without difficulties. Based on this observation, one

could say that all languages of the world form a vast continuum, covering history and prehistory, and that by way of their strong similarities, they are susceptible to interference at any time.

→ Gardani, *Contact and borrowing*, CambrHandb 28

RSG 2, sect. XII, *Contacts linguistiques et migration*, art. 134–169

3.6.3 Romance-based creoles

Many situations of language transfer are limited to lexical borrowing; in some cases, however, grammar and phonology are involved. Creole languages present the most marked case of language interference. As noted previously, creoles emerged in the regrettable historical context of European colonisation at the dawn of the modern period (16th to 18th centuries, cf. 2.3.2 no. 2, as well as 10.5.1). They arose from attempts at communication between the colonisers and their subjects (natives or, often, slaves brought from other, especially African, countries): the latter spoke one or more languages that were very different from those of the former. The multiple examples all indicate a starting point of almost total incomprehension among speakers, which made the classic process of learning the dominant language difficult. In this way, ‘minimal systems of communication’ established themselves, consisting of a lexical base of frequently used words taken from the vocabulary of the dominant language and structured by means of new grammatical principles.

The establishment of these systems nevertheless relies on habitual processes of second-language learning. A speaker of a given language (L1) gradually learns elements of a second language (L2), his/her individual competence slowly moving toward that of a speaker of L2. At every stage of the learning process, this newly-acquired competence is interpreted by the learner as a coherent linguistic system – an ‘interlanguage’ (using the term introduced by Selinker in 1969). This explains phenomena such as false analogies, which are very frequent during the process of language learning (cf. also Corder/Roulet, *The notions of simplification, interlanguages and pidgins and their relation to second language pedagogy*, 1977). The formation process can be illustrated by the following figure:

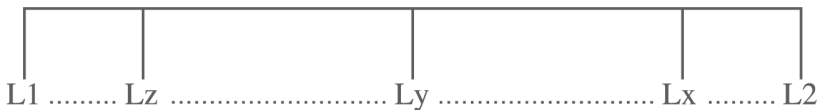


Fig. 11: The formation of an interlanguage

Once the process of learning has reached a suitably advanced stage, the individual language system takes on the shape of Lx and is thus close to L2, while at the same time maintaining some interference with L1. A less advanced stage (Ly) corresponds

rather to a mixed language that combines elements of L1 and L2. Conversely, when the process stalls due to a lack of information on L2, the extrapolated system (Lz), curiously, does not consistently reflect either L1 or L2, but instead represents an ‘interlanguage’, a ‘minimal system of communication’.

Under the normal circumstances of language learning, such an interlanguage disappears at a very early stage; however, in the case of *pidgins* and *creoles*, this system persists, becoming a language in its own right. The first speakers of minimal communication systems such as these do not develop them into regular languages – at this stage, one speaks of ‘pidgins’ –, in contrast to their children, who begin with the pidgin as their mother tongue, transforming it into a fully-fledged creole language with the aid of core linguistic input.

There are creoles with Luso-Spanish, English, French and even Arabic lexical bases (the latter, almost unknown, are spoken in sub-Saharan Africa). These languages developed in geographically disparate locations, often islands, in the Americas, Asia or Africa.

The majority of Romance-based creoles have already been mentioned:

- Papiamentu, with a Hispano-Portuguese (and Dutch) base, spoken on the Antillean ‘ABC islands’ (Aruba, Bonaire and Curaçao);
- other Portuguese-based creoles spoken in Guinea Bissau, the Cape Verde Islands and in São Tomé-and-Príncipe (cf. 3.2);
- among the creoles with a Spanish lexical base, the group of Chabacano creoles spoken on the Philippines, and Palenquero in Columbia (cf. 3.4.9, no. 4);
- French-based creoles, more numerous than those mentioned above, spoken in Guiana, Martinique, Guadeloupe, Haiti, La Réunion, Mauritius, the Seychelles and New Caledonia (cf. 3.4.1 and the example sentence in 8.4.3 no. 6).

Though the external history of the *Romània creolica* is parallel to that of the *Romània nova* in that it involves colonisation, typologically speaking (i.e. with regard to their internal structure), creoles with a Romance lexical base bear no resemblance to the Romance languages themselves. Instead, they are closer to the other creoles – those with an English or Arabic lexical base. Thus, the *Romània creolica* is beyond the basic scope of this manual, despite its interest for general linguistics (cf. 2.3.2 no. 2; 6.2; 8.4.3 no. 6 and 10.4.1 no. 2 for the *lingua franca* with an Italian lexical base).

→ Bollée/Maurer, *Creoles*, OxfGuide 24

Bachmann, *Creoles*, CambrHist 2, 11

OxfEnc (*Creole Languages*, Muysken; *Spanish-based Creole Languages*, Schwegler/Jacobs/Quint; *French-based Creole Languages*, Zribi-Hertz; *Portuguese-lexified Creoles*, Clements)

RSG 2, sect. IX, *L'Histoire des langues de la Romania creolica*, art. 97–100

4 Variational linguistics and the Romance languages

4.1 The impact of variation

4.1.1 Linguistic variation – an empirical constant

The presentation of the languages and dialects of the *Romània continua* has shown that a language is never homogeneous. Each of the Romance languages encompasses a number of varieties that diverge on a geographical level, including old, hereditary dialects and, in some cases, varieties that have been exported to other parts of the world, such as Spanish or Portuguese in America. In addition to these geolinguistic varieties, most Romance languages also have a modern standard variety.

However significant all these differences may appear, they are in fact the mere tip of the iceberg of the effects induced by linguistic variation, which is omnipresent in speech. In the act of utterance, all elements of language are potentially subject to variation: phonetics and phonology, morphology and syntax, as well as lexis, are thus the internal *variables* involved. *Variants* appear sporadically and tied to specific situations at first; they may subsequently become habitual for an individual (idiosyncratic variation, distinctive for an individual person), for a larger group of people or in a specific communicative context, potentially leading to the emergence of linguistic *varieties*. It is in this manner that, over time, variation in synchrony leads to regional, sociological, contextual or medium (i.e. spoken/written) differences within a given language. In the most extreme cases, variation in space leads to the emergence of new languages (cf. 1.2.5). Thus, synchronic variation is the *sine qua non* condition for diachronic variation. As well as contributing to the internal diversity of a language, the omnipresence of variance in utterance is the driving force behind linguistic expressivity and language change.

In the following chapters, many examples of synchronic variants relating to the various domains of language will be presented. These include:

- phonetic and phonological variants, such as the allophones [R] and [r] in French and Italian (cf. 2.1.1 no. 1.2);
- syntactic variants, e.g. the omission of the negative particle *ne* in French, as in *je ne sais pas* vs. *je sais pas* (cf. 4.4.3 no. 1);
- variants in macro-syntax, involving, for example, phenomena of ellipsis in spoken language; e.g. Fr. (*je n'ai*) *jamais fait ça* or (*c'est*) *génial!* (cf. 4.4.3);
- lexical variants, which are very numerous e.g. Peninsular Spanish *coche*, *ordenador*, *patata* vs. American Spanish *carro* (or *auto*), *computadora*, *papa* (cf. 9.1.4).

Many of the morphological or syntactic variants occur predominantly in speech, as opposed to written language, as the latter is more strongly stabilised by standardisation. Specialised terminology and complex syntax, in contrast, are characteristic of writing,

which has allowed them to attain an advanced stage of evolution and to develop a higher degree of variation.

The majority of variants have no lasting effect on the linguistic system and its organisation. Nevertheless, if a number of them coincide with specific language-external parameters, relating to (i) a particular geographical area, (ii) a social group or (iii) situation, they may constitute true *varieties*. In the following, we will describe these three decisive parameters in more detail (cf. Glessgen/Schösler, *Repenser les axes*, 2019: 17–26).

The first, geographical or so-called ‘diatopic’ variation, produces the sort of varieties that are the most deeply rooted in the evolution of languages. It is this parameter that is ultimately responsible for the formation of new languages: all Romance languages developed from Latin as a result of geographical separation; similarly, geographical separation caused the divergence of Galician and Portuguese from a common medieval base (cf. 3.4.11/12). Diatopic variation manifests itself throughout the *Romània* in the following very diverse forms:

- hereditary regional dialects that have evolved directly from Latin. These are known as ‘primary’ dialects and may be unintelligible to a speaker of the respective standard language;
- more recent regional dialects that have developed from the expansion of varieties that already belong to Romance. They are known as ‘secondary’ dialects and include examples such as Andalusian, Canarian and Mexican Spanish (varieties that have developed from Castilian Spanish), or Canadian French (which has evolved from the variety spoken in France);
- ‘regiolects’ of a language, such as the French spoken in Paris, Marseilles, Strasbourg or Lausanne. This type of variation occurs only after a standard variety of a language is in place, and is dependent upon it.

The second type of variation, known as ‘diastratic’ variation, is based on the differences in linguistic prestige that arise from different social groups. Diastratic variation requires a diversified, stratified society. In tribal or clan-type societies, where the members of a tribe, usually consisting of between 50 and 100 individuals, form one society, it cannot develop to any great extent. This is probably also true in the case of fully rural societies with little mobility, which were dominant throughout the emerging Romance territories between the 6th and the 8th centuries (cf. 1.2.4 no. 2). In a modern society, however, utterances vary considerably according to the socio-cultural group to which the speaker belongs, and according to his or her level of education: a farmer, a labourer, a craftsman, a shopkeeper, a teacher, a priest or pastor, a doctor, a lawyer and an academic are quite easily identifiable by their manner of speaking or writing. These differences give rise to distinct varieties which are mutually intelligible – albeit with a tendency towards misunderstandings – and which play a significant role in social inclusion or exclusion.

Both from a theoretical and an empirical perspective, diastatic variation overlaps with the third type, which results from different communicative situations, and is termed ‘diaphasic’ variation. It includes stylistic registers and – according to Coseriu (*Sprachkompetenz*, 1992: 280) – also involves differences related to biological groups (defined by gender or age) and professional groups, despite a strong overlap with diastatic variation. It thus has an affinity with the notion of ‘discourse traditions’, which include scientific language or the language of the media (cf. 4.5.1 no. 1). Differences between diaphasic varieties can further be likened to those between spoken and written language (cf. 4.4).

Like diastatic variation, diaphasic variation characterises societies with a high degree of socio-cultural and professional differentiation. In the Western world, its development has intensified since the end of the 19th century, following increased specialisation of the professional world and the diversification of society into micro-groups.

The communicative importance of these different language varieties lies in the fact that their use within a society mirrors the awareness its members have of their own identity. Social roles and the relationships established among the members of a society become apparent through the varieties they choose, and through the particular form these varieties assume.

From a language-internal point of view, the emergence of all these specific language forms – or ‘lects’ – is the direct result of the variability inherent in language itself, and of its openness to the spontaneous alternatives that regulate its functioning. Although no single variety is impermeable and each one, in turn, contains a large number of variants, it is habitual use that reinforces the internal cohesion of linguistic varieties. Thus, if a new form is used a few thousand times in a given context, it eventually becomes characteristic of this particular context – be it diatopic, diastatic or diaphasic.

4.1.2 The limits of variation: intelligibility, grammaticality and acceptability

Though language variation may be particularly pronounced, it nevertheless has its limits. When one considers a specific language at a given point in history, it is possible to identify not only *variable*, but also *invariable* elements. If the latter were also subject to variation, they would no longer correspond to what speakers consider to be part of the language in question. In this sense, intelligibility constitutes the extreme limit of variation: if an utterance cannot be understood, it can not be considered to be a successful variant (cf. 1.2.4 no. 2). Though the intelligibility of utterances is strongly context-dependent, it remains an inter-subjective dimension.

A second parameter that restricts variation was identified by Chomsky: the presence or absence of grammaticality, i.e. the property of an utterance to be well formed in grammatical terms, that is, in adherence with the rules of syntax governing the language in question (= ‘grammaticalness’, Chomsky). A purely syntactic parameter, grammaticality does not take semantic factors into consideration. Thus, an utterance may be

ungrammatical but intelligible, or grammatical but unintelligible (the classic example suggested by Chomsky being the sentence *Colorless green ideas sleep furiously*). Grammaticality concerns all varieties of a language, though one rarely finds utterances that are grammatical in one variety and ungrammatical in another (but cf. e.g. South West England (Devon) *She be coming* or *She am here* vs. standard English *She is coming* and *She is here*). On the other hand, grammaticality is context-independent within each individual variety; in short, the same forms remain grammatical in all contexts.

A third parameter, more difficult to pinpoint, is acceptability. An utterance may be intelligible and grammatical but unacceptable in a given context. It may also be ungrammatical but acceptable (e.g. in the case of incomplete sentences such as *Would you please ...?*, uttered by a host and followed by a gesture inviting the guest to sit down). Acceptability strongly depends on situational context, the language varieties involved, the relationship between speakers, and their shared linguistic knowledge. Language use that is considered unacceptable leads to negative reactions from other speakers. The parameter of acceptability therefore has strong sociological and pragmatic implications (concerning the appropriateness of language use in a particular context; cf. 4.3.2). To sum up, linguistic variation needs to be both intelligible and acceptable in order to be successful, and ideally also grammatical.

4.1.3 The variational approach to language

Variational linguistics replaces the idea of a single indivisible language by that of a mosaic of partially identical, partially divergent forms. A language, then, is made up of the elements of all linguistic varieties used by its speakers. These varieties share the vast majority of grammatical rules as well as a comprehensive basic vocabulary, ensuring linguistic cohesion despite existing differences.

Following this line of thought, written and spoken French in France, the *oïl* dialects, as well as written and spoken French in Canada form part of one and the same variational space; the same applies to standard Spanish in Spain, its regional dialects spoken throughout the Iberian Peninsula and its written and spoken varieties in America. It is important to distinguish between the concept of ‘variational space’ and that of the ‘communicative spaces’ of a given language, the latter corresponding to a political nation. The communicative space normally includes several languages grouped beneath one dominant umbrella language, which represents the standard language of the respective nation: in France, the French standard language spans Occitan, Breton, Alsatian as well as Arabic dialects, etc.; in Italy, Italian as a roofing language covers a multitude of genetically distinct Italo-Romance dialects, as well as Romance and non-Romance languages; in Spain, the Spanish standard coexists with Catalan, Galician and Basque (cf. 3.1 and 3.2).

The variational space of a given language thus exists within different communicative spaces. The French variational space includes the communicative spaces of France,

Quebec or Morocco, but it does not completely fill any of them. At the same time, this international variational space constitutes a linguistic reality of its own: the multiple channels of communication in the modern world (media, travel, professional mobility) contribute towards bringing all varieties of one language into contact with each other, with the result that if one element changes, all others are affected. Varieties such as youth language and the language of the media in France, Canada or Switzerland exhibit as many shared evolutionary patterns as divergent ones. To express this idea of a complex linguistic entity, the varieties of a language viewed as a whole are known as its ‘architecture’ or its ‘diasystem’ in variational linguistics (*cf.* 4.3.1).

The variationist perspective allows us to better discern the concrete physiognomy of a language and the path of the changes it undergoes. It emphasises the dynamic aspect of a given language, which at any one time encompasses divergent norms. At the same time, however, it is faced with numerous epistemological obstacles and must constantly have recourse to concepts that are ‘fuzzy’ by their very nature, such as the ‘continuum’, ‘relative distinctions’ or the ‘empirical approach’, all of which partly rely on intuition.

The distinction of different varieties is more a result of a statistical study than a linguistic reality immediately recognised by speakers. This is clearly illustrated by the means used to identify the limits of regional dialects (*cf.* 4.3). Linguistic cohesion requires formal and semantic continua. This starting point led R.A. Hudson (1980) to the idea of an ‘item-based model’ (focusing on individual linguistic phenomena rather than broader variational groupings), which assumes that each individual linguistic phenomenon (i.e. each ‘item’) “is associated with a social description which says who uses it and when” (Hudson 1980, 51). The identification of varieties thus consists in the establishment of groups of phenomena that have been analysed according to these criteria.

As an example, an apico-alveolar (or ‘rolled’) [r] appears in the regionally marked variety of French spoken in Marseilles, but also in the French spoken in Tunisia, or in the French of dialect speakers in northern France (*cf.* 2.1.1 no. 1.2). In contrast, the presence of final nasal consonants in *main* [mɛ̃n] (vs. standard [mɛ̃]), only appears in southern varieties of French of France. Thus, for the description of the regiolect of Marseilles, both phenomena are pertinent: it is their co-presence that distinguishes the regiolect of Marseille from other varieties, and elsewhere, apico-alveolar [r] coexists with other phenomena.

The types of variation presented above will now be described in more detail, beginning with diatopic variation, which is the most fundamental, despite the relatively weak presence of regional dialects in the present-day Romance world. We will then discuss the other types (diastratic and diaphasic variation, spoken-written and immediacy-distance strategies, discourse traditions and textual genres), which call for in-depth theoretical reflection. Variation in geographical space enables a better understanding of the other forms of variation, which are more complex and more interdependent.

4.2 Geographical or diatopic variation

4.2.1 Primary dialects

The three types of geographical varieties – primary dialects, secondary dialects and regiolects – have quite different characteristics (*cf.* 4.1.1). Primary dialects are the oldest and, in the *România*, the only type to have known a hereditary continuity since spoken Late Latin. In theory, these dialects are seen as having emerged as a result of an uninterrupted chain of communication, from generation to generation, within the same inhabited locality. When they are observed from a broad perspective, this view corresponds to reality, despite population movements within the *România continua* and situations of linguistic contact, including the influence of the standard variety on dialects (*cf.* 10.4.1 no. 3).

These spoken regional dialects are the starting point for the elaboration that has resulted in different written and standardised varieties of the Romance languages; this, in turn, explains why they are termed ‘primary’. These language forms were neutralised and enriched on the basis of models provided by other written languages, above all Latin (but also Old Slavonic in Romania and Arabic in Spain). In this sense, dialects are the ‘natural’ varieties of a language, whereas standardised forms are ‘semi-artificial’ varieties (*cf.* 4.3).

From a geographical point of view, owing to the vagaries of history, primary dialects do not cover the whole of the space occupied by the Romance languages; they traditionally characterise a geolinguistic continuum that encompasses Italo- and Gallo-Romance as well as a fringe of the northern Iberian Peninsula of approximately 100 to 200 km in breadth. In contrast, Spanish varieties, as well as Valencian or Portuguese, are essentially the result of linguistic expansion and they are consequently known as secondary dialects. This also applies to Romanian, albeit as a result of a different historical context, and even more so to the whole of the *România nova* (*cf.* 4.2.3).

→ RSG 3, sect. XIV, *Histoire des dialectes dans la Romania*, art. 212–215

1 Status and characteristics of primary dialects

Differences between primary dialects and the standard variety within the same diasystem may be so great that the former are unintelligible to monolingual speakers of the latter, despite a large degree of structural similarity. These differences are often restricted to the domain of phonetics and to specific lexemes, as shown by the following excerpt from a folk tale in Lorrain dialect:

In hôme de Frimbô atô mérié avô eune fôme qu’ atô méchante, pire que lo diape;
 Un homme de Fraimbois était marié avec une femme qui était méchante, pire que le diable.
 “A man from Fraimbois was married to a woman who was malicious, worse than the devil”.

Elle fiô tojo le contrère de c' qu'i d'mandô ; ç'atô eune manre bête
 Elle faisait toujours le contraire de ce qu'il demandait ; c'était une mauvaïse (moindre) bête.
 "She always did the opposite of what he asked; she was (like a) beast of poor quality".
 [...]

Quand toute celé ettu velé, 'en r'vont é Frimbô
 Quand tout cela avait été avalé, ils s'en reviennent à Fraimbois.
 "When all that had been swallowed, they went back to Fraimbois".

Lé fôme atô in pô tournisse (ça bîn fé à moins, lé peute [< PUTIDA] bête!)
 La femme avait (était) un peu le tournis (ça [c'était] bien fait au moins, la vilaine bête).
 "The woman felt a little dizzy, that taught her a lesson, at least, the nasty beast".
 [...]

Edited by Jean Lanher, *La Femme qui se noie*, in *Les Contes de Fraimbois*, 1983. The tale was written around 1900 by Athanase Grandjacquot, a former teacher at the elementary school of Fraimbois. The transcription of the text is amateurish, influenced by French.

Primary dialects have played a determining role in the history of the *Romània*, and the contribution to methodology provided by their study is no less significant. This is due in particular to the high degree of linguistic variation they exhibit, as well as to the fact that their variation is related to the easily identifiable parameter of space. Results obtained from the analysis of dialect variation in space can thus serve as a model for the identification and description of diastatic and diaphasic varieties, especially with regard to boundaries between neighbouring varieties and the interactions that arise between them. Primary dialect variation illustrates both the way in which a multifaceted linguistic diasystem functions as a cohesive entity and the potential it holds for extreme diversification. Moreover, dialectology complements and enriches the sociolinguistic study of linguistic identity, insecurity and conflict.

The unique value of primary dialects as witness to of 1,500 years of 'natural' or 'spontaneous' linguistic differentiation explains the key position that linguists have long allotted them. It should also be mentioned that around 1870/80, when the field of dialectology emerged, primary dialects were still of strong sociological importance within their traditional territories. These varieties, which were part of everyday reality for the great majority of speakers in France, Italy, Switzerland and Belgium, found themselves in conflict with the national standard languages, which were in the process of expansion. The weak presence of these regional dialects in today's Romance-speaking world (cf. 3.4 above), in contrast, led to a decline in interest in dialectology in the 1980s. However, this trend was reversed during the second decade of the 21st century. The surviving dialects and the large collections of dialectal documentation from the past have once again become the object of study, in particular through the use of digital humanities (cf. Thibault *et al.*, *Nouveaux regards sur la variation dialectale*, 2019). The dialects of the *Romània* are thus in the process of regaining their rightful status as the outcome of and the witnesses to the diversification of Late Latin: indeed, primary dialects constitute the principal object of observation for comparative Romance linguistics.

The fact remains, however, that the empirical study of Romance dialects developed mostly between 1870 and 1960, manifesting itself in two forms. These were i) monographs detailing the internal characteristics of a variety in a given locality, and ii) linguistic atlases covering a rather broad area. In both cases, efforts were mainly of a descriptive nature, concentrating on phonetic variation, the compilation of traditional vocabulary and the establishment of morphological paradigms. The interpretative component was generally neglected by authors of such studies, perhaps with the exception of their analysis for etymological and onomasiological purposes.

Finally, syntax remained peripheral during the heyday of dialectology, even in terms of mere description. It is true that variation is weaker in syntax; indeed, its relative homogeneity ensures the cohesion of the diasystem. Recent work, however, has both revealed and begun to exploit the potential offered by the field of micro-syntax in particular, concentrating on the highly diversified dialects of Italy. Comparative research in this area over the last three decades has thus brought to light an extremely marked degree of microvariation (cf. the three-volume work by Manzini and Savoia, *I dialetti italiani e romanci* published in 2005, and, much earlier, Rohlfs, *Grammatica della lingua italiana e dei suoi dialetti*, vol. 3: *Sintassi e formazione delle parole*, 1969).

2 The language atlases of the *Romània*

The first major language atlas of a Romance language was the *Atlas linguistique de la France* (ALF, 1902–1912). It was conceived by the Swiss scholar Jules Gilliéron, whose previous work, the *Petit atlas phonétique du Valais roman* (1881), coloured by hand, was the first of its kind in the *Romània*. Gilliéron held the first chair of dialectology, created for him at the École pratique des hautes études in Paris in 1894 (cf. 2.2.3). Field enquiries for the ALF were carried out by a man who had received no formal linguistic training but who had an excellent ear. Edmond Edmont was a cheese merchant with an interest in the diversity of dialects spoken in his native region and who collected his observations on the dialect of Saint-Pol-sur-Ternoise (Pas-de-Calais) in a reliable dictionary. For the ALF, Edmont travelled throughout France over the course of four years, interviewing dialect speakers in 639 villages with the aid of a questionnaire.

The 1,920 entries in the questionnaire essentially cover concepts belonging to rural everyday life (parts of the human body, farm animals, agricultural tools, festivals, etc.). The answers further allow the establishment of phonological inventories and the identification of a limited number of grammatical features (the latter including verbal, nominal and pronominal paradigms, nominal determination and – albeit very superficially – the syntactic characteristics of simple sentences; cf. Chauveau's rich description of the development of the objectives of the ALF, RSG 1, art. 8).

Answers given by informants were written on individual maps, each of which reproduces a geolinguistic panorama for one concept. The maps are raw materials that require further elaboration before they can be fully exploited. In particular, their legibility is improved by the grouping of the phonetic variants according to etymological

types, as shown by the map of the representants of the verb *scier* “to saw” in Occitan and Gascon regions:

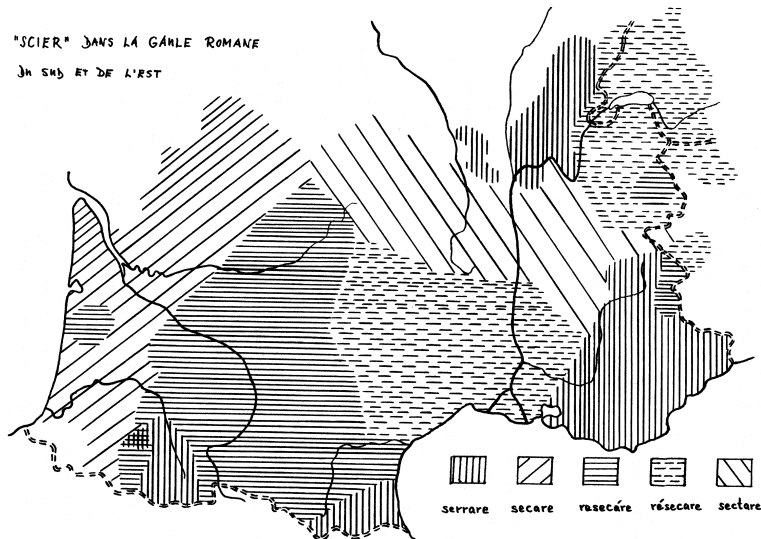


Fig. 12: Digitised map of the ALF for the verb “to saw”

Source: Lothar Wolf, *Aspekte der Dialektologie*, 1975: 58 (based on a study by Gilliéron/Mongin 1905).

The map shows how descendants of the Latin lexeme *SERRARE* (“to saw”) have declined, owing to their homophony with continuants of *SERARE* (“to close, lock”):

1. Initially, the descendants of *SERRARE* must have covered most of the territory, as evidenced by the relict-like preservation of the word in peripheral areas. Forms that have developed from *SERRARE* persist in four zones: in (Gascon-speaking) Béarn; in (Catalan-speaking) Roussillon; in the central and eastern parts of Provence and western part of the Dauphinois (= Occitan-speaking); in the (Francoprovençal-speaking) areas north of Lyon.
2. *SERRARE* is first replaced by *RESE'CARE*, which survives in eastern Gascon and western Languedocien and, in the form of the variant *'RESECARE*, in eastern Languedocien as well as western and southern Francoprovençal.
3. According to Gilliéron, *RESE'CARE* loses its prefix in favour of *SE'CARE* which prevails in all French dialects (not shown on this map) including western Limousin and western Gascon ('Gascon noir').
4. The area covered by *'RESECARE* appears discontinuous due to the introduction of the even more recent form *SECTARE* (“to cut”) in varieties of northern Occitan.

The ALF is currently available in three different digitised versions (cf. FEWGuide 179):

- University of Innsbruck: <<http://diglib.uibk.ac.at/urn:nbn:at:at-ubi:2-4568>>
- University of Grenoble: <<http://lig-tdcge.imag.fr/cartodialect4/>>
- University of Toulouse: <<http://symila.univ-tlse2.fr/alf>>

The survey methods of the ALF paved the way for the elaboration of language atlases of further Romance languages, which appeared throughout the 20th century. An alternative method was implemented by Georg Wenker for the German linguistic atlas (*Deutscher Sprachatlas*), begun slightly earlier than the ALF (although publication did not begin until 1926 and remained partial). Wenker had sent questionnaires to schoolmasters, church ministers and village mayors, who were asked to indicate the forms used in their local dialects for a range of concepts. This method also provided useable results, even though it relied on data from a large number of non-specialists, which greatly reduced the quality of phonetic transcriptions and increased the probability of error.

The methods of direct enquiry were subsequently refined, beginning with the *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* (AIS, *Sprach- und Sachatlas Italiens und der Südschweiz*, published in 8 volumes from 1928–1940), conceived by the two Swiss professors Karl Jaberg and Jakob Jud. The most important methodological differences with regard to the ALF are the following:

- they chose not one, but three scholars to conduct the enquiries: Paul Scheuermeier for northern Italy, Gerhard Rohlfs – a renowned dialectologist and lexicographer specialising in southern Italo-Romance dialects, whose field travels by donkey would later become legendary – and Max Leopold Wagner, author of the *Dizionario Etimologico Sardo*, for Sardinia;
- in contrast to the ALF, their enquiries included large cities;
- the maps of the AIS are organised according to thematic groupings (= in onomasiological order, i.e. by concepts);
- they are also accompanied by drawings indicating the shape of the designated objects, since the choice of designation for an object is often motivated by its shape;
- finally, whereas the points of inquiry in the ALF follow a purely geometrical order (according to their position within a circle), they are organised according to geolinguistic regions in the AIS (cf. fig. 12).

We have already presented the isoglosses separating regional dialects in the north (points 1–499), the centre (pp. 500–) and in central-southern Italy (pp. 745–; cf. 3.3 and fig. 4, the latter showing the points of the AIS for central Italy in detail). An additional dividing line appears in the extreme south (Salentino, southern Calabrian and Sicilian). Moreover, the AIS includes Sardinian, Ladin and Friulian, despite the fact that these languages are not part of the Italo-Romance continuum; Corsican, on the other hand, is absent, even though it belongs to the Tuscan dialects (cf. 3.4.5 no. 3).

At almost the same time as the AIS, the *Atlante linguistico italiano* (ALI) was begun in 1924 in Turin. Publication, however, began only recently (9 volumes have been published to date, the first appearing in 1995). The documentation it provides is more extensive than that of the AIS, which it complements, and it is methodologically more advanced.

The more recent series, the *Nouveaux atlas linguistiques et ethnographiques de la France*, has attained a new level of precision. It comprises about twenty works in multiple volumes that cover all regions of France. These new atlases not only expand the network of enquiry points covered by the ALF, but also make use of the advanced methodology created for the AIS; in addition, they rely on interviewers who are themselves dialect speakers. Recent atlases incorporate an increasing number of sociological factors, an example being the *Atlas lingüístico diatópico y diastrático del Uruguay* (ADDU, directed by Harald Thun, 2000–). The exemplary *Atlante linguistico del ladino dolomitico e dei dialetti limitrofi* is also published online, thus allowing electronic searches (ALD-1 and ALD-2, directed by Hans Goebel, 1998 and 2012); it includes detailed descriptions and interpretative dialectometrical maps that merit particular attention.

Using the linguistic atlases is no trivial matter and requires additional tools. For the AIS, Scheuermeier assembled a large number of photographs that complement the drawings on the maps and bear witness to the everyday culture described in the Atlas (*Bauernwerk in Italien, der italienischen und rätoromanischen Schweiz*, 1943–1956). The volume of commentaries (Jaberg and Jud, *Der Sprachatlas als Forschungsinstrument*, 1928) and the lexical index of the AIS (Jaberg 1960) are indispensable tools. Moreover, new insights into geolinguistic configurations can be gleaned from the index of the *Lessico etimologico italiano* (LEI, created by Max Pfister), which includes the material of the AIS as well as – for northern regions – a computerised study by Goebel (*Dialektometrische Studien*, 1984; an expanded version covering the entire scope of the AIS is in progress).

For the ALF only a rough index exists (Gilliéron and Edmont, *Table de l'Atlas Linguistique de la France*, 1912) in addition to a volume of supplements (1920), but the lexicological articles of the fundamental *Französisches Etymologisches Wörterbuch* (FEW, by Walther von Wartburg) take into account all the material present in the ALF. The ALF has also been digitised by Goebel, as have the only published volume of the *Atlas lingüístico de la Península ibérica* (ALPI) and the comprehensive Catalan atlas (ALCat).

The production of language atlases is one of the greatest achievements of Romance studies. Not only the primary dialects of the *Romània continua*, but also the varieties of the Iberian Peninsula and Romania, as well as numerous Latin American countries and Quebec, have been studied in the context of linguistic cartography projects. There are even plans for an *Atlante linguistico mediterraneo*, which aims to map linguistic data from coast to coast in the Mediterranean, as well as the pan-Romance project *Atlante linguistico romanzo* (ALiR, directed by Gaston Tuaillon and Michel Contini). In total, more than 400 linguistic atlases have been produced for the *Romània*.

→ Chauveau, *Histoire des langues romanes et géographie linguistique*, RSG 1, art. 8
Lausberg/Winkelmann, *Romanische Sprachatlanten*, LRL 1/1, art. 87

3 The contribution of dialectometry: the example of *Gallo-Romance*

The field of dialectometry represents an important innovation in the presentation of geolinguistic data. It involves the digitisation of the forms contained in a selection of maps, which are subsequently compared using computerised methods; the results of the comparison are projected onto a polygonal map. Each point of the atlas, then, corresponds to a single polygon. In this way, similarities and differences between linguistic forms are displayed founded on reliable statistical calculations, allowing the identification of zones of (relative) linguistic interruption or dialect ‘frontiers’.

As we have already seen, some isogloss bundles are easily identified without the aid of information technology; however, quantified data vastly increases the accuracy of both macro- and micro-divisions. Based on an analysis of 1,421 digitised maps of the ALF, Goebl was able to delimit the major linguistic zones of Gallo-Romance with remarkable precision. The polygons in the following map correspond to the inquiry points of the ALF, distinguishing between the main dialect groups of French (ABCD), Francoprovençal (E), Occitan (GHI) and Gascon (F):

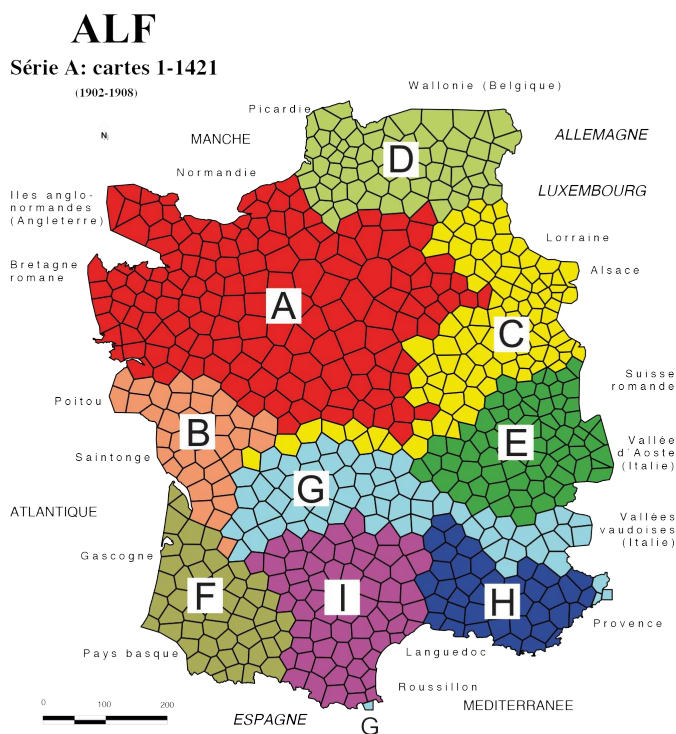


Fig. 13: Dialectometrical map of linguistic zones of Gallo-Romance (source: Goebl 2003, map 20)

Remarks:

1. like the ALF, the map excludes non-Romance-speaking Bretagne and Lorraine, as well as Alsace. In contrast, it includes – in contradiction with the title of the atlas ('... de la France') – Wallonia and Romance-speaking Switzerland;
2. an adequate reading of the map requires other aspects of dialectometrical methodology to be taken into consideration: these are i) hierarchical classification, ii) the distribution of similarities, and iii) the different ways in which the data are displayed. The following observations are based on these elements of classification:
 - maximum contrast appears between French (A-D) and Francoprovençal (E) on the one hand, and Occitan (G-I) and Gascon (F) on the other;
 - within the *oïl* domain, one can distinguish between a central-west group (A) a north-eastern group (CD) and the south-west, which was originally Occitan (B);
 - Gascon (F) is clearly differentiated from the Occitan dialects: Limousin, Alvernian and Dauphinois in the north (G), Languedocien (I) and Provençal (H) in the south.

A more detailed classification of linguistic data for the *oïl* area keeps the dialect identity of the southwest (Poitou-Saintonge) intact, but leads to an internal differentiation among the two remaining groups, as displayed in the following figure:

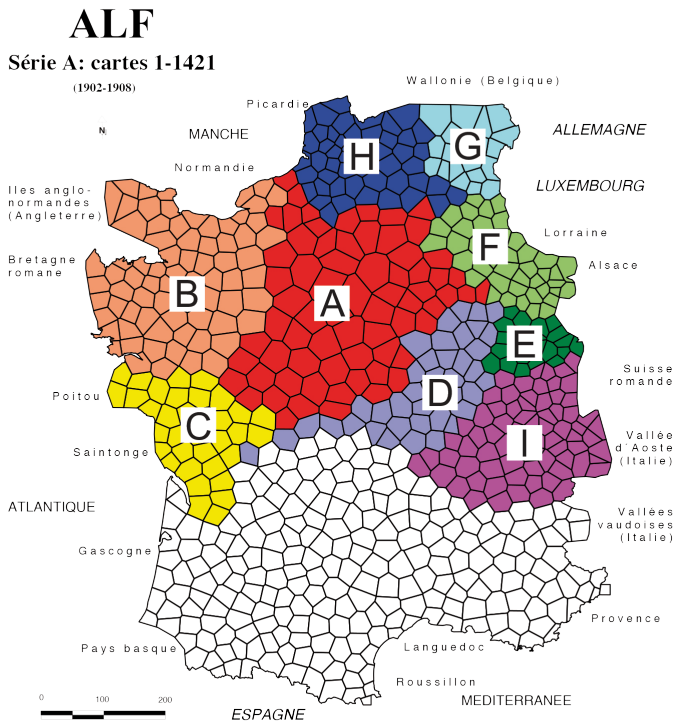


Fig. 14: Dialectometrical map of the *oïl* area (source: Goebl 2003, map 22)

The following areas can further be distinguished:

- the south-western and the Francoprovençal area (here C and I [= B and E on fig. 13]);
- within the central-western group [= A on fig. 13]: a western group (B) and a central group (A); the western group includes Normandy and Romance-speaking Brittany; the central group encompasses the zone around the Île-de-France, de-dialectalised since the 18th century, the *Centre* region *per se* and Champagne;
- within the northern group [= D on fig. 13]: Picard (H) and Walloon (G);
- within the eastern group [= C on fig. 13]: Lorrain (F), Franc-Comtois (E) and Bourguignon (D).

→ Goebel, *Romance linguistic geography and dialectometry*, OxfGuide 7

4 Dialectological descriptions and studies

Parallel to the language atlas projects, there are numerous monographs on local dialect varieties, particularly for Italy, France and the northern fringe of the Iberian Peninsula, many of which were written prior to the beginning of the 1960s. These studies often take the form of university theses of the type ‘the dialect of ...’. Like the atlases, these publications place great emphasis on phonetic and lexical phenomena, whereas morphology is less consistently treated and syntax, regrettably, is almost completely neglected. A bibliography of the older dialect dictionaries for Italo-Romance was published by Prati in 1931 (*I vocabolari delle parlate italiane*). For Gallo-Romance, Wartburg and Keller’s more recent annotated bibliography of the *Dictionnaires patois gallo-romans* (1969) illustrates the great variety and quality of dialectological studies. Modern dialectological monographs are more of a rarity and are more likely to describe Italo-Romance varieties (such as Loporcaro’s *Grammatica storica del dialetto di Altamura*, published in 1988, or Ledgeway’s *Grammatica diacronica del napoletano*, 2009; works based on a synchronic approach include Parry’s *Sociolinguistica e grammatica del dialetto di Cairo Montenotte*, 2005 and Tortora’s *A Comparative Grammar of Borgomanerese*, 2014).

Like the atlases, these dialectological studies were stimulated by the *Wörter und Sachen* (‘words and things’) movement during the first half of the 20th century, which focused on the question of the etymological motivation of dialect forms. There are hundreds of publications, studying subjects such as the designations for ‘plough’, ‘poppy’, ‘glow-worm’, etc. in Gallo-Romance or Italo-Romance dialects, or in Romance languages and dialects (cf. 9.8 and 2.2.4 no. 1). Dialectological studies on the naming of specific concepts are catalogued and described in an exemplary manner in Quadri’s *Aufgaben und Methoden der onomasiologischen Forschung* (1952).

Atlases and dialectological studies have produced a wealth of data that provide raw material for the study, among other things, of the semantic and derivational principles involved in the formation of the vocabulary associated with traditional life in Europe. Curiously, the great potential inherent in all these materials assembled by several generations of researchers remains largely under-exploited. The ALF and monographs on

Gallo-Romance dialects form the principal basis of the FEW (the main etymological dictionary of Gallo-Romance), and the AIS and other dialectological sources for Italy play a similar role for the LEI. Apart from this, however, dialectological work has not been extensively utilised, apart from by the authors of the works themselves and by lexicographers with an interest in etymology.

→ Barbato, *Non-quantitative approaches to dialect classification and relatedness*, OxfEnc

4.2.2 Language evolution and geographical space

The primary reason for mapping dialect data is to observe the evolution of language in relation to space. All dialects of the *Romània* form a geolinguistic continuum, which extends from Galicia in the southwest to Wallonia in the north and to Sicily in the southeast. Within this large group, internal linguistic diversification increases progressively with distance. Thus, two dialects that are geographically close are always more similar than two that are further apart. For example, there are numerous intermediate varieties separating a variety of Picard from a variety of Provençal, such as the dialects of the French *Centre* region, the transitional zone between French and Occitan (known as the *Croissant*) and northern Occitan. The degree of cohesion between geolinguistic areas correlates with geographical distance and proximity. This even applies to standard languages, elaborated from dialect varieties: Galician resembles Spanish more than it does Catalan, Spanish resembles Catalan more than it does French, etc.

Nevertheless, and in contrast to what can be observed for elaborated languages, divisions between dialects are never particularly marked. Determining dialect frontiers is an abstract operation, and differs considerably from the perception of dialect speakers themselves. Whereas linguists draw isogloss bundles in order to identify dialect groups (*cf.* 3.3), speakers nearly always see themselves in continuity with their neighbours. It should further be mentioned that the same methodology is used to identify other forms of linguistic variation and perhaps, more generally, the many forms of socio-cultural variation: in order to define the characteristics of an informal variety of a Romance language, or even to compare Old French and Modern French, linguists establish isoglosses that serve to structure the continua under study.

The central idea of a geolinguistic approach is that geolinguistic configuration reflects the historical evolution of a language; this illustrates a strong interdependence between the parameters of space and time. As we have seen from the example of the words for 'TO SAW' (fig. 11), the distribution of linguistic forms on the maps of an atlas makes it possible to distinguish innovative forms from archaic forms that bear witness to earlier language stages. Innovative forms are often situated in central zones or are part of continuous zones, whereas conservative forms display a disjunct distribution pattern and are often situated in peripheral or 'lateral' zones, to use the terminology introduced by Matteo Bartoli in his *Introduzione alla Neolinguistica* (published in 1925,

cf. the methodological overview by Weinhold, *Sprachgeographische Distribution und chronologische Schichtung*, 1985).

The first to study these phenomena systematically was, once again, Gilliéron, in his monograph *Généalogie des mots qui désignent l'abeille*, based on the ALF (1918), as illustrated by the following map:

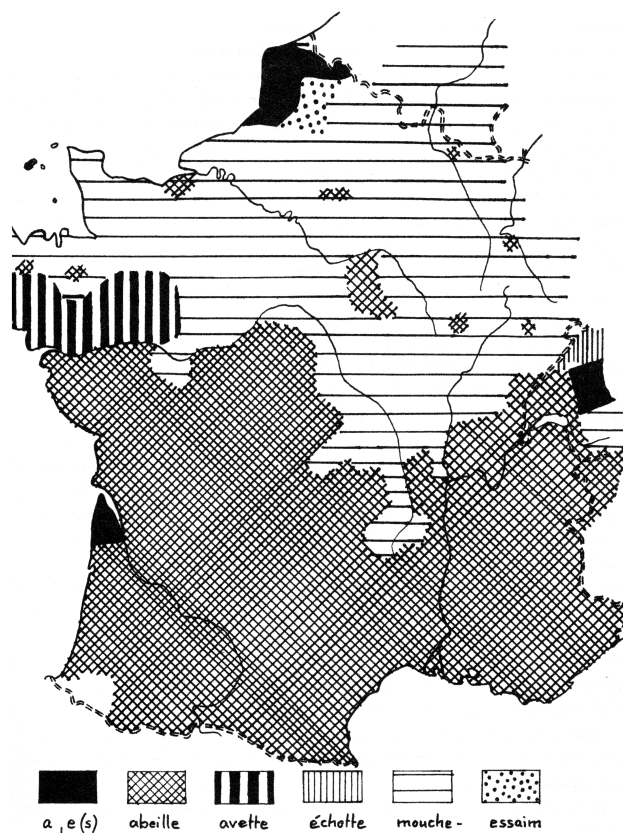


Fig. 15: Simplified map of the denominations for 'BEE'

Source: ALF, map 1, after Wolf, *Aspekte der Dialektologie*, 1975: 59.

Remarks: the map presents several concurrent forms of the Modern French word *abeille*; their geographical distribution allows the following chronological layers to be identified:

1. The oldest type is *a, e(s)*, corresponding to the Latin word of origin *APIS/APE*; this type is found in the northern tip of the Gascon-speaking area (Médoc: *aps*, ALF points 548, 549, 650), in the north of France (*eps*, *es*, *Ê*, ALF point 295, etc.), on the island of Guernsey (*eys*, point 399) and in Eastern Francoprovençal (*a*, point 60,

etc.). This distribution can only be interpreted as the result of a Late Latin or early Romance diffusion of Lat. *APE* / Old French **ev* across the whole of the *oïl* domain, before this phonetically weak form was replaced by longer forms.

2. The first dominant type on the map is Occitan and Francoprovençal *abelha* (Galliscised as *abeille*), deriving from the Latin diminutive *APICULA*, which is recorded as a concurrent form of *APIS*. The two Late Latin words must have coexisted in emerging Gallo-Romance, *APIS* being the older, and *APICULA* a more recent form that asserted itself in the whole of the south of the territory, extending north to the Loire River and east to the Jura Mountains. The distribution of these two words during the 8th to the 10th centuries split Gallo-Romance into two parts, illustrating the rather limited area covered by French around ca. 1000.
3. The second dominant type, *mouche à miel*, is an innovation in Old French. Phonetic weaknesses of the type affecting *APE* in the north (which, in French, evolves into *es*, *ef* or *ee*, cf. FEW 24, 11a s.v. *APIS*) must have brought about its replacement by this new form, which spread throughout the *oïl* domain, eventually, in the modern period, also driving the *abeille* type southward (movements are clearly visible along the valleys of the Loire).
4. Two other diminutives of the successors of *APE* can be found in zones bordering on the areas covered by *abeille*: *avette* in the west (< *ef* + *-ITTA*, probably involving interference with *abeille*) and *échotte* in the east (a variant of *aichette*, *aichatte* < *aissatte* < *ey*s + *-ITTU*).
5. On occasion, other lexical innovations appear spontaneously, including *essaim* (< *EXAMEN* “swarm”, formed by metonymy).
6. Finally, standard French *abeille* is a loanword from regional French, which presumably entered the varieties spoken in Poitou-Charentes as a borrowing from Occitan *abelha*. *Abeille* appears in French as early as the 14th century (cf. FEW 24, 10a s.v. *APICULA*), where it coexists with the dialect form *mouche à miel*, even replacing it sporadically during the 19th century (by a process known as ‘de-dialectalisation’). The areas in which this replacement occurred are also shown on the map (Normandy, the Centre, etc.).

This map shows how a given form can come to occupy a considerable area: first, the form emerges by innovation in one or sometimes several different locations; it then spreads around one or more centres, eventually covering a larger, continuous area. This model applies to every sort of linguistic change. Innovations always occur in a small subentity of the diasystem: a city, a few villages, a socio-cultural group (such as the royal court, journalistic circles or suburban youth) or a specific discourse tradition (such as lyric poems or scientific texts). These subentities may accumulate at different levels: an innovation can thus occur concomitantly (1) in poetry, (2) at the royal court and (3) in Paris or Versailles.

From there, the innovation spreads to other varieties of the diasystem, finally becoming generalised. Moreover, this process can be interrupted at any time, as can be

seen from different innovations that remain locally circumscribed on the map (*avette*, *échetote*, *essaim*). The disappearance of linguistic forms follows the same path, but in reverse order: a form first disappears from individual parts of the diasystem before being abandoned everywhere.

Dialect variation thus provides a clear explanation for innumerable phenomena of divergence within the diasystem (cf. 4.3.1). It also shows that a coherent interpretation of the evolution of a historical language (cf. 9.1.4) can only be achieved if the evolution of its different varieties is also taken into account, as far as this is possible. The emergence of words in standard Modern French, Italian or Spanish can only be explained through the combination of medieval sources and modern dialect forms, the latter reflecting past evolutionary stages. The same principle, which forms the basis of major etymological dictionaries like the FEW, the LEI and the Ibero-Romance dictionaries of Corominas, applies to every grammatical or phonological phenomenon. The cohesion of the diasystem extends to its historical stages – as in Coseriu's idea of the '*langue historique*' (cf. 4.3.1).

4.2.3 Secondary dialects

Today, secondary dialects are spoken throughout most of the *Romània*. These are not directly descended from spoken Late Latin, but from varieties that were already Romance and that were exported by their first speakers during migratory movements. Andalusian, for example, is a variety of the Spanish that originated in Castile, and it was introduced into Andalusia in the middle of the 13th century as a result of the *Reconquista*. By that stage, however, Spanish had already become an elaborated written language. The linguistic base of what was to become Andalusian was already homogeneous to a certain extent (cf. 10.4.1 no. 1), as a result of neutralisation induced by population mixing during migration and by the written use of the language. Consequently, it did not have the opportunity to diversify.

In contrast, the high diversity of primary dialects is the result of relatively stable population settlement, combined with a weak influence of written models (Latin or Romance). From an internal perspective, in contrast to primary dialects, secondary dialects are thus characterised by the effects of linguistic neutralisation and homogenisation.

Terminologically speaking, the distinction between 'primary' and 'secondary' dialects, established by Coseriu (1981), is based on the different relationships that these two types of dialects have with a standardised, neutralised language. Primary dialects precede the standard language, which is forged on their basis; they can also exist without a standard variety being derived from them (as is the case for Francoprovençal, the northern and southern Italo-Romance dialects and, to some extent, Occitan). Sec-

ondary dialects, on the other hand, depend on a language that is in the process of neutralisation, or even on the existence of a standard language.

The same observations apply to Portuguese as a secondary dialect of Galician, or to Valencian as a secondary dialect of north-western Catalan. In the case of Romanian, diversification of the spoken language was hindered by the limited geographical stability (*stabilitas loci*) of its speaker population, following movements of the population from south of the Danube and, finally, nomadic displacement. Linguistic mixing thus obstructed the formation of highly diversified dialect varieties of the primary type in Romania (Daco-Romanian varieties are, however, clearly distinct from Aromanian, Megleno-Romanian and Istro-Romanian).

Romance varieties outside Europe are, without exception, secondary dialects, since they emerged from exported European varieties. French in Quebec, for example, is based on dialects from the west of France in their 17th/18th-century form, which through mixing and interaction with written French formed the starting point for new developments. Today, popular spoken French in Quebec nevertheless displays notable regiolectal variation, along with clear variation according to socio-cultural factors. A secondary dialect can thus diversify over the course of centuries.

On a larger geographical scale, similar evolutionary patterns explain the rich lexical innovations in Spanish and Portuguese in the Americas (*cf.* 10.5.1 and 10.6.1). Peninsular varieties, especially from the south, were exported from the 16th century onwards in several waves, generally via the Canary Islands, and they have undergone an evolution of their own on the American continents. After the emergence of Latin American countries in the 18th century, secondary dialects finally developed their own new prestigious varieties, similarly to Quebec, through interaction with the standard language of their respective metropolitan centres. These standard languages are no longer considered secondary dialects, but so-called ‘regiolects’ – they will be discussed in the following chapter (*cf.* no. 4.2.4 below).

A secondary variety of major cultural importance is Judeo-Spanish, which developed in the Mediterranean basin and south-eastern Europe after the expulsion of the Jewish population from the Spanish kingdom in 1492 (*cf.* 10.5.1). Another noteworthy example is spoken French in the Maghreb and various sub-Saharan countries, which, despite being merely a second language for part of the population, displays the characteristics of a secondary dialect.

Today, linguistic research on secondary varieties is as advanced as work on primary dialects. There are numerous language atlases relating to the *România nova* (such as the excellent *Atlas lingüístico diatópico y diastrático del Uruguay* by Thun, mentioned previously) as well as a large number of monographs on its dialects. Above all, secondary dialects allow the study of phenomena of language mixing and neutralisation, or of variational phenomena relating to discourse traditions. For example, the comparison between a conversation held in a shop in Mexico and a similar conversation held in

Castile immediately brings to light the politeness characteristic of the novo-Hispanic culture, which contrasts with the brisk manner of the inhabitants of the central Peninsula. As far as research on diaphasic or pragmatic variation is concerned, secondary dialects offer a field of research that remains largely unexplored.

4.2.4 Regiolects or tertiary regional variation

The particular linguistic characteristics of regiolects are much less noticeable than those of primary or even secondary dialects. This is not surprising, since regiolectal varieties – ‘tertiary’ dialects according to Coseriu (*Los conceptos de ‘dialecto’, ‘nivel’ y ‘estilo de lengua’*, 1981: 14) – by definition presuppose the existence of a preceding elaborated or even standardised variety, from which they arise. Regiolects of French, Italian or Spanish, then, were only truly able to develop after the standardisation of the respective languages had begun in the 16th century. The evolution of regiolects continued in parallel with that of the corresponding standards, which intensified during the 20th century. This tertiary level of variation concerns phonetic nuances and prosody, as well as the use of regional words; grammatical variation, on the other hand, is rare.

Though internal differences are weak, they nevertheless allow a speaker of French from Strasbourg, for example, to recognise whether his or her interlocuteur is also a native of the region of Alsace, or whether he or she is from ‘inner’ France. A Milanese speaker can easily recognise a central or southern Italian speaker, and a Galician will know if he or she is speaking to a fellow native speaker or to an Andalusian. Regiolectal variation thus has a strong impact as an identity marker – a characteristic feature of all diasystematic variation.

Regiolectal lexical variation has only recently attracted the attention of researchers. While the situation is still far from satisfactory for the majority of the *Romània*, several high quality resources exist for the various French-speaking countries, which, all together, provide an analysis of many thousands of regional lexemes:

- the exemplary *Dictionnaire des régionalismes de France* (DRF), by Pierre Rézeau (2001), and the recent dictionary *Les Mots des Poilus* published in 2018 by the same author; cf. Glessgen / Thibault, *La lexicographie différentielle du français*, 2005, as well as Carles / Glessgen, *Les écrits des Poilus*, 2020
- the *Dictionnaire suisse romand* (DSR) by André Thibault (1997, 2004)
- the *Dictionnaire historique du français québécois* (DHFQ) by Claude Poirier (1998)
- the *Dictionnaire des belgicismes* by Michel Francard *et al.* (2010, 2021)

Regiolectal variation in French was already displayed by the written medieval language, which, far from being homogeneous, was divided into regional *scriptae* (i.e. written traditions), such as those of the British Isles, Picardy, Lorraine or the Île-de-France (cf. 10.4.3 no. 3). More accurately, the *scriptae* were a phenomenon of pluricen-

tric codification (*cf.* below), since the standard language was not yet fully formed but in the course of elaboration on the basis of these interdependent *scriptae*. The situation changed slightly toward the beginning of the 16th century, once the variety of French practised at the royal court and by the members of the higher ranks of society in the capital had attained the status of a standard and had become a model for the elite of the whole nation. This standard variety, which had been restricted to the domain of writing, was then able to slowly make its way into spoken language, which, in turn, led to the development of spoken regional or ‘regiolectal’ French in southern France or in Romance-speaking Switzerland. There is only limited information available on the old regional varieties of French, though we know that many of today’s forms go back to the 16th to 18th centuries (*cf.* Chambon, *Méthodes de recherche ...*, sect. 3, *Régionalité et variation lexicale*, 2017, 1, 355–401; 503–597).

Spoken regiolectal varieties are thus based on the spoken form of a neutralised variety. They are enriched by borrowings from further indigenous varieties (such as the *oïl* dialects, Occitan, Breton or Alsatian), by semantic and derivational innovations, and they also conserve forms that are no longer in use in the standard variety (such as *septante* instead of *soixante-dix* for “seventy” in Switzerland and Belgium).

The history of regiolectal variation in Italy is even more recent than in France, since written medieval Italo-Romance varieties displayed greater divergence than those in France and since the standard language, which was established in the 16th century, remained a purely written language for two or even three centuries. Hence, regiolectal spoken varieties of Italian were only able to emerge in the 20th century. Regrettably, the state of research on *italiano regionale* remains unsatisfactory.

The variational situation is particularly complex in Spain, where secondary dialects and (informal or standard) regiolectal varieties form a single vast continuum. This applies to both sociolinguistic aspects and internal linguistic characteristics. As an example, the urban dialect of Seville (a secondary dialect) and the regional variety of Spanish spoken in the same city are not clearly distinct from one another, and a plethora of intermediate forms exist that may manifest themselves at any time, depending on the circumstances and speakers involved. A distinction of this type is thus in no way comparable to the distance observable between the urban dialect of Venice (a primary dialect) and the Italian regiolect spoken there – these could easily be considered to be two distinct languages.

In the Romance-speaking countries outside Europe, regiolectal variation and the respective national standard varieties overlap: the standard French of Quebec can simply be considered a regiolectal variant of standard French in general, and the standard Spanish of Argentina a variant of standard Spanish. In both cases, regiolectal variation within the country may display further differentiation.

From this perspective, the standard French of France is merely one variety among others and is consequently a regiolectal variety of French. The existence of terms that

are used in France but not in Canada or Romance-speaking Switzerland would thus be characteristic of this 'French' regiolect. This is similar for Portuguese; it is even more marked for Spanish, which is spoken in many countries of the Americas, and by a much larger number of speakers than on the Iberian Peninsula. From this point of view, Spanish is similar to English with its different standard varieties in Great Britain, the USA, Australia, Africa and India, alongside numerous other varieties in countries where it functions as a second language.

French in France or Spanish in Spain undeniably maintain a certain prestige within this complex tapestry of interdependencies, but rather as *primus inter pares* (= first among equals). New varieties of standard (and non-standard) French, Spanish or Portuguese are diffused throughout whole countries thanks to the media, travel and migration. This phenomenon is called 'pluricentric codification', based on the idea that norms are forged simultaneously in different places (cf. 10.6.1).

In the world of today, regiolectal variation has gained considerable importance. At a time when primary dialects are losing ground, secondary dialects persist and regiolects are flourishing.

→ RSG 3, sect. XIV, *Histoire des variétés régionales (et urbaines) dans la Romania*, art. 212–219

4.3 Diastratic and diaphasic variation and their theorisation

4.3.1 Foundations of variational theory

Whereas the study of diatopic variation raises many methodological questions and calls for knowledge in different areas (historical phonetics and etymology in particular), it nevertheless does not require a highly elaborated theoretical background. The other types of varieties, in contrast, are neither easy to identify and describe, nor to understand from a theoretical point of view. The theorisation of variational linguistics began relatively recently, in the mid-20th century. In 1951/52, the Norwegian Romance scholar Leiv Flydal (1904–1983) laid the foundations of a line of thought that became established in modified and elaborated form owing to the work of the previously-mentioned structuralist Eugenio Coseriu (1966 etc.). Flydal observed that a given language in synchrony encompasses several subsystems, which he called 'language structures' ('*structures de langue*') (1951/52), and that, moreover, speakers are aware of earlier stages of the language they speak. Flydal stressed the possibility of speakers borrowing forms from other 'structures' of a language or from 'older structures' or (sub)systems ('*structures anciennes*'), especially for stylistic purposes (this generates 'extrastructuralisms' ['*extra-structuralismes*']). Using the same logic, Coseriu later spoke of 'functional languages' ('*langues fonctionnelles*', 1966).

As a designation for the sum of all these 'structures' and 'extra-structuralisms', Flydal introduced the term 'language architecture', and shortly thereafter, Uriel Wein-

reich introduced the alternative designation ‘diasystem’ (1954), which is more commonly used today. Coseriu added the concept of ‘historical language’, which encompasses both synchronic variation, in whichever form it may exist, and diachronic variation. The ‘historical language’ thus provides the general framework for all forms of studies on a given language.

Flydal had also previously identified the two concepts of variation according to geographical area and variation according to social hierarchy, introducing the terms ‘diatopic’ and ‘diastratic’ used here. Variation associated with language use was, in his opinion, an ‘operation of stylistic nature’. The term ‘diaphasic’ was only introduced later on by Coseriu (cf. 4.1.1), who used it to refer not only to stylistic choices that can be made by any speaker and to differences arising between different groups of professionals, but also to those arising from biological factors (such as sex and age), which impose greater constraints.

The Anglo-American linguistic tradition typically considers diatopic and diastratic varieties to be primary varieties, grouped together under the term ‘dialects’ (cf. 3.1; see also Halliday, 1978, who speaks of ‘varieties according to users’). These are distinguished from situational varieties or registers (Halliday: ‘varieties according to use’), whose use and linguistic form depend on the relationship between speakers (also known as ‘style’ / ‘tenor’ / ‘attitude of discourse’), the topic of conversation, the situation (or ‘field’) and the discourse ‘mode’ (i.e. spoken or written). Halliday and Hudson also emphasise the fact that a single variety may assume a different value for different speakers or groups: ‘one man’s dialect is another man’s register.’

→ Koch/Oesterreicher, *Comparaison historique de l’architecture des langues romanes*, RSG 3, art. 220

4.3.2 Interdependence among the varieties of a language

Both Flydal and Coseriu noted the intense interdependence among the three dimensions of variation. No diastratic and diaphasic varieties exist in ‘pure’ form, and important dictionaries, grammars or computerised databases which allow a variety to be identified or recognised only exist for the standard language. However, even these resources offer only a very approximate description, and they should be complemented by the observation of the language used in major newspapers, which accurately reflects current standard varieties.

In reality, linguistic varieties themselves display a great deal of overlap and are interwoven: a dialect variety or, even more so, a regiolectal variety, may include a range of different language registers or stylistic features. Similarly, a diastratic variety can manifest itself in a number of different regional forms and it can also involve stylistic variation. Finally, an utterance connected to a given situation may vary according to regional and sociolinguistic parameters (cf. Coseriu ²1992: 284). Hence, the three dimensions of variation are simultaneously present in every utterance (cf. 4.6, fig. 19). In prin-

ciple, moreover, speakers have a command of several varieties of their native language, which can be considered as a limited form of multilingualism (cf. Coseriu 1992: 290).

Coseriu also observed that a diatopically-marked element can fulfil the same function of sociological distinction as a diastratic element. 'Speaking dialect' has the same effect of social inclusion or exclusion as speaking a popular variety in a society that no longer speaks dialects. In the same way, a diastratically-marked element can evoke a diaphasic effect: a popular term, for instance, can emphasise a moment of intimacy in an informal familiar dialogue. The three types of variation are thus connected by a unidirectional 'chain of variation': 'a dialect [*dialecto*] can function as a register [*nivel*] and as a linguistic style [*estilo de lengua*], and a register likewise as a style but not the other way round' (cf. Coseriu, *Los conceptos de 'dialecto', 'nivel' y 'estilo de lengua'*, 1981). The following figure illustrates the chain of variation:

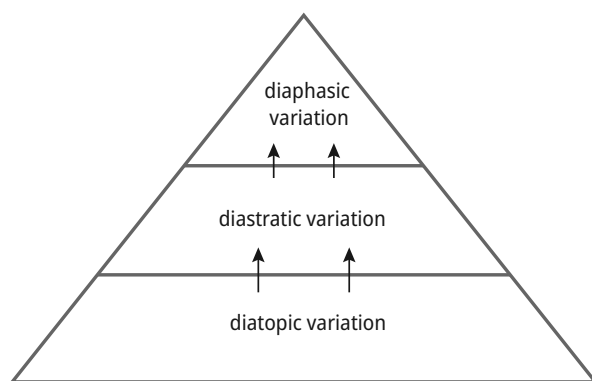


Fig. 16: Chain of variation

Source: Coseriu 1988: 133; cf. Coseriu 1966.

As in any situation of diglossia, the coexistence of different varieties results in numerous variational borrowings, above all of a lexical nature. Terms may find their way from regional dialects into popular or regiolectal varieties and from there into the standard language; popular terms create expressive effects in language used in a familiar context; technical terms can be introduced into standardised or informal varieties. Conversely, words belonging to a high-prestige variety can be used to colour a familiar discourse, for instance, to signal the social rank of the speaker. In dialect varieties, borrowings from a standard or informal variety are frequent, as the dialect speaker can introduce them into conversation at any moment in order to heighten linguistic prestige; such borrowing increases when regional dialects are being progressively abandoned. Even the large number of Latinisms in Romance texts, regardless of the period in which they were introduced (cf. 2.1.5), can be attributed to the concept of variation: in learned circles, Latin was seen as a prestigious variety belonging to the same linguistic continuum as the native Romance language.

It should be noted that the direction of lexical borrowings between varieties often depends on the meaning of the terms. Concepts belonging to the sphere of cultural elaboration frequently migrate from the standard variety to a dialect (or from Latin into a Romance language), whereas concepts with strong emotional connotations are easily borrowed into the standard language from popular varieties, thus evoking an expressive and/or dysphemic (pejorative) effect.

Borrowings between varieties result in lexical ‘connotation’: French words such as *bagnole* “car” or *meuf* “woman” thus do not simply have an objective and ‘denotative’ meaning, but also bear the mark of the variety to which they originally belonged, thereby lending an additional nuance – or connotation – to a term (cf. 9.1.4). Consequently, the semantic value of a word should always be distinguished from its variational value: French *merde* “shit” or *con* “bloody idiot” are terms that possess a vulgar denotative meaning, but they belong to both informal and colloquial French (*français familier* and *français populaire*). Their use in informal discourse does not reflect an instance of borrowing from a popular variety. In contrast, terms like *moufflet* or *mioche* for “child”, which are semantically neutral, clearly belong to popular French and retain a connotative value when used in a familiar context.

4.3.3 Defining diastratic and diaphasic variation

Whereas the diatopic dimension is unambiguously defined by the external parameter of geographical area, a definition of the diastratic and diaphasic dimensions, which are not clear-cut, is more difficult to achieve. In the case of diastratic variation, Coseriu claims that sociological differences within the community constitute a determining external parameter (1992, 280), making it a form of sociolectal variation concerning the speaker’s belonging to a specific social rank or group. More abstractly, and perhaps more appropriately, one might conceive of diastratic variation as correlated with and induced by linguistic prestige (cf. 4.6). In both cases, diastratic variation presupposes a stratified society.

For diaphasic variation, Coseriu suggests several concrete parameters: the speaker’s desire to express him- or herself in a specific manner, or factors such as ‘biological’ or professional difference (*ibid.*). It is thus primarily a situational type of variation connected to specific contexts, composed of linguistic forms that are recognisable as belonging to these contexts. The parameter of contextualisation is, however, not easy to pinpoint: there are countless communicative situations, each of which is characterised by specific linguistic choices. Each language takes on a different appearance depending on whether it is being used in a conversation among friends or at a doctor’s appointment, to compose a love letter, a letter of recommendation or a birthday speech, or to referee a football game. These situations rely on partially pre-established models and require specific variational choices to be made. However, variance that is merely the result of the content of a discourse does not constitute a linguistic variety. As an

example, medical texts from the Middle Ages contain a large number of specialised terms, yet they adhere to grammatical and textual models that were widely used in other genres at the time. Conversely, medical texts written in the 20th century make use of specific lexical and grammatical means, including, in particular, a strong tendency towards both nominalisation and the development of artificially created terminology.

It is necessary to distinguish instances of variation that are dictated solely by content from those that presuppose specific elaboration and assume an identity-forming role. The latter first lead to the modelling of 'textual genres'; later on, their linguistic characteristics may become generalised, and one can then speak of true varieties. In earlier periods diaphasic varieties were indeed linked to different styles of literary and semi-literary texts, divided between high, middle and low styles. During the 20th/21st centuries, specialised forms of textual expression have multiplied due to the diversification of professional activities. 'Textual genres' thus extend beyond the diaphasic dimension, and select concrete utterances according to their objectives and their structure (*cf.* 4.5.1 no. 1 and 4.5.1 no. 2).

The 'biological' parameters advanced by Coseriu have proven to be less meaningful. In Western societies characterised by relative gender equality, different varieties for men and women have not developed; while differences in communicative behaviour are apparent, these do not amount to distinct linguistic varieties. Age, on the other hand, does play a role in youth language; however, once adulthood is reached, language use remains comparatively stable.

4.3.4 Linguistic prestige

Linguistic prestige plays a special role with regard to the make-up of diastratic and diaphasic varieties. Day-to-day experience easily confirms the existence, in any given place, of varieties with differing levels of linguistic prestige: divergences between a standard and a familiar variety, a popular sociolect or a regional dialect are easily perceived by speakers. Linguistic prestige is initially associated with the prestige of different social groups: a variety is imbued with high prestige if it is used by individuals who are seen as being important in society, and vice versa. There is nevertheless no absolute correspondence between social groups and varieties. German linguists frequently cite the example of the former Chancellor Helmut Kohl, one of the most powerful men in Europe who governed the country for 16 years but whose oral expression remained that of a simple man. Conversely, individuals of lower social status may use highly sophisticated language.

Descriptive sociolinguistics positions diastratic and diaphasic varieties on a scale ranging from maximum prestige ('high variety': HV) to minimum prestige ('low variety': LV), with varieties of variable prestige situated between these two extremes. Considered from this point of view, the varieties function as distinct languages in a diglossic or polyglossic situation. These terms were introduced by Charles A. Ferguson in 1959 to

account for the difference between modern Classical Arabic (HV) and dialectal Arabic (LV), which, for the *Romània*, may be compared with that which must have existed between Latin and the Tuscan dialects around the 12th century. This distinction encompasses both internal distance between varieties (i.e. the degree of difference displayed by specific linguistic phenomena) and the sociological roles they assume.

For modern languages in the Western world, in addition to the standard (HV) and the substandard or non-standard (LV) varieties, it is necessary to distinguish a ‘supra-standard’, of even higher prestige than the norm, but whose use remains restricted to certain contexts such as poetic language or the cultivated language of official discourse.

Another set of terms was introduced by Derek Bickerton (*Dynamics of a creole system*, 1975) for polyglossic situations involving creole languages. In this type of situation, an ‘acrolect’ (of the HV type) and a ‘basilect’ (LV) are distinguished, in addition to a ‘mesolect’ of intermediate prestige. This terminology is particularly useful in cases where a given prestigious language is not the native language of the population (e.g. French in French-speaking Africa); in such contexts, these terms indicate the degree (on a continuum) to which the prestigious variety is mastered by the population.

Correspondence between the different sets of terminology is easily established:

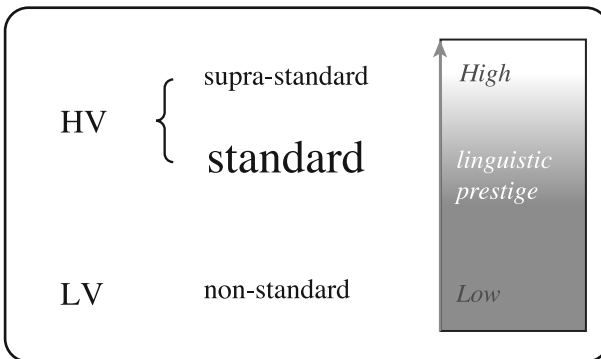


Fig. 17: Terminology used to describe diastratic varieties

Differences in prestige involve both diastratic and diaphasic varieties, in circumstances that are, in part, due to the vagaries of history. In all cases high-prestige varieties should be considered as diaphasic, since individuals who master the standard and/or the supra-standard are also capable of using non-standard forms. Low-prestige varieties, in contrast, are divided between diaphasic varieties (such as the informal varieties of Romance languages) and sociolectal, and thus diastratic, varieties (such as the French spoken in the *banlieues* or popular varieties of Italian and Spanish). The status of a variety depends on its actual use in a given place at a given time.

The development of the diastratic and diaphasic dimensions depends on the existence of a highly diversified and specialised society; thus, the situation in the Western world today is fairly recent. In cultures without writing or in those that are essentially

rural, the only varieties that can acquire differing degrees of linguistic prestige are diatopic ones (i.e. the prestige of the dialect of a rich city or region is higher than that of a dialect spoken in a poor or peripheral town in the countryside). Although it seems natural to us today, the distinction between a standard and a non-standard only arises when a high-prestige variety associated with political and economic power becomes established. A diglossic situation existed in Ancient Rome, but in Europe around the year 1000, only Latin held the role of a prestigious variety and the evolution of vernacular towards a standard resumed later on, during the course of the medieval period, intensifying from the 16th century onwards. For other languages such as Occitan in France or Galician and Catalan in Spain, however, this process resulted in a loss of prestige between the 16th–19th centuries.

In today's society, circumstances are further complicated because social prestige can be linked to political or economic power (in industry or business circles), as well as to mass media or intellectual circles. Consequently, the exact nature of the role played by (socio-)linguistic prestige in language elaboration has become more difficult to ascertain.

Another important consideration is the fact that linguistic prestige has no absolute value. Within any social group there are dominant linguistic norms which create constraints and pressure, independently of its prestige in the context of the entire society. A speaker who wishes to gain acceptance within a group is required to use a defined linguistic variety. The use of a standard variety at a football match, for example, would be equally as inappropriate as the use of a non-standard variety at a high society event. Every linguistic variety has its own internal norms, which its speakers efficiently 'defend' in order to maintain their identity.

In the case of non-standard varieties, the respect of norms is guaranteed by a system of positive and negative reactions on the part of fellow speakers. Linguists speak of implicit 'descriptive' norms. Standard varieties, in contrast, benefit from institutionalised reinforcement; they are protected by the state in conjunction with the education system and supported by dictionaries and usage grammars; these are explicit, codified 'prescriptive' norms. The terms 'standard' and 'non-standard' thus reflect the impact of the two types of norms rather than their nature. In its own way, then, a non-standard variety is just as constraining (or 'prescriptive') as the standard.

The coexistence of different norms relativises the concept of the linguistic mistake (*cf.* 4.1.2). Often, a form considered to be 'wrong' is simply one that is unacceptable in the standard variety, and the discrepancy is caused by interference with a non-standard variety; within the non-standard variety in which it originated, it is not a 'mistake' at all.

- Jones/Parry, Mair/Williams, *Sociolinguistic variation*, OxfGuide 35
 Smith, *Social factors in language change and variation*, CambrHandb 30

4.3.5 Diastratic and diaphasic variation in the *România*

It is significantly more difficult to identify the varieties that exist within a society at any given time than to draw dialect isoglosses to map regional dialects, as the external parameters of social prestige and contextualisation are in a state of constant fluctuation. Recourse to a process of controlled deduction is therefore necessary:

- as a starting point, a social group in a given situation is identified intuitively;
- patterns of language use specific to this group or situation are then described and contrasted with norms that are recognisable according to dictionaries and ‘pre-scriptive’ grammars (a process known as ‘differential description’);
- this allows a variational value to be attributed to the linguistic data: e.g. different forms of negation in French (that is, the use of the adverbs *ne [...] pas* in combination, as opposed to the adverb *pas* in the absence of *ne* or *ne* in the absence of *pas*) convey different levels of prestige;
- surveys are conducted on the basis of these linguistic elements: which speakers use which elements (e.g. which type of negation) in which situations?;
- finally, the results of such studies enable the social and contextual parameters that are pertinent for diastratic and diaphasic variation to be specified.

This process – indispensable in variational linguistics –, may be described as an ‘inductive spiral’. As we progress along this spiral, the precision of our state of knowledge is further increased.

The majority of present-day European languages include informal and popular varieties which coexist with a standard variety. The informal or familiar variety is the habitual means of expression of speakers of the standard in informal situations. The popular variety, in contrast, displays the features of a sociolect; speakers of popular varieties who also master the standard are an exception. Depending on the language, these distinctions are developed to a greater or lesser degree, from an internal point of view as well as in terms of usage.

In Spanish, the distinction between a written standard (*estándar*) and a spoken informal variety (*habla coloquial*) is not pronounced; however, the gap between *español literario* (the supra-standard variety) and *español popular* is greater. Furthermore, diatopic variation in Spanish draws on diastratic and diaphasic variation: standard Spanish and literary Spanish, for example, display a certain degree of regionalisation in Spain and, above all, in the Americas. In turn, informal Spanish and, even more so, popular Spanish strongly diverge from one country to the next.

In Italian, one can distinguish an *italiano standard* (which is more regionalised than French in France and Spanish in Spain), an *italiano dell'uso medio* and a highly diversified *italiano regionale*. The latter tends to replace rural dialects, which neverthe-

less maintain considerable vitality. Regional Italian and Italo-Romance dialects generally fulfil the sociological functions of popular varieties.

Contemporary French encompasses numerous varieties: a *français soigné* (or *littéraire*), a *français standard* (or *commun*) – the corresponding spoken variety of which is known as *français courant* (weakly regionalised) –, a *français familier*, a *français populaire* and finally, a *français des banlieues*. So-called *argotique* words form a separate category, which, rather than representing a separate variety, consists of sets of specific lexical forms used in combination with a grammatical basis belonging to the *familier-populaire* continuum. The original *argot* (slang) developed in the context of the First World War (*argot des Poilus*, based on the slang used by criminals in the second half of the 19th century); but to an even greater extent than the other varieties, its distinguishing features are almost exclusively lexical (with a high number of metaphors and dysphemisms, or vulgar expressions, that make use of animalisation and eroticisation). In this it is similar to forms of student slang and some modern professional jargons.

The other Romance languages, above all Portuguese and Romanian, display yet further internal patterns regarding the make-up of their linguistic architecture. In Romansh, Galician and Catalan, neither diastratic nor diaphasic diversification are particularly highly developed. The structure of the diasystem thus depends on the degree of elaboration undergone by the languages, as well as on their geographical extension.

Concerning the use of the three languages considered more in detail, only *español popular*, *français populaire*, and, more recently, the *français des banlieues* (*et/jou des jeunes*) can be considered as sociolectal varieties in the traditional sense, whereas the others are clearly diaphasic. In addition, the position of *français populaire* seems greatly reduced in today's society compared to its importance in the 19th and the first half of the 20th centuries. This development is even more evident in the former *italiano popolare*, which emerged at the beginning of the 20th century when monolingual dialect speakers first attempted to use some form of Italian, and which has since been replaced by the varieties of *italiano regionale*.

Distinguishing between the diastratic and diaphasic dimensions becomes even more complicated when concrete situations are observed. Consider the following three examples:

1. 'Informal' varieties (*español coloquial*, *italiano dell'uso medio*, *français familier*) are considered to be diaphasic since speakers are able to choose between these varieties and the standard (their use is thus determined by contextual constraints). In France, however, the 'popular' variety has largely merged with *français familier* and many speakers of the latter have not mastered the standard variety. The 'familiar' variety is thus the only form of linguistic expression of which they have full command and it should therefore be classified as a sociolect.

2. The varieties of the *banlieues*, in contrast, are seen as diastratic varieties that are used by the inhabitants of poor neighbourhoods; in reality, however, these varieties mainly serve as youth language, whose use is often determined by situational factors. More precisely, the cultural context of the second half of the 20th century has caused the stage of ‘youth’ to be prolonged and has led to the emergence of a real youth culture (cf. 10.6.4). Linguistically speaking, this culture leans towards models provided by peripheral groups in society. Young middle-class people thus imitate the language used in disadvantaged urban neighbourhoods. It should also be noted that the media directed at the younger generation further contribute towards reinforcing the internal characteristics of this language by accentuating them. Finally, forms of expression such as rap represent the deliberate elaboration of linguistic varieties that are regarded as socially peripheral. Returning to the chain of variation, what was originally a diastratic variety functions as a diaphasic variety; however – and this complicates matters even further – the new diaphasic variety always reflects the original sociolect to some degree.
3. Professional varieties are considered to be discourse traditions rather than varieties (cf. 4.5.1 no. 1); nevertheless, they comply with all the characteristics of situational choice typical for diaphasic varieties, and at the same time, they carry strong sociological implications. The variety used by lawyers, for example, is recognised by all speakers as characteristic of a social group.

These three examples show that current diasystematic theory is only partially helpful as a means by which to describe present-day society. They were more adequate in the 1950s and 1960s, when they were first theorised by Flydal and Coseriu. For this reason, we prefer to speak of a ‘diastratic-diaphasic continuum’, which encompasses different facets that often overlap.

→ RSG 3, sect. XIV, *Les variétés diastratiques et diaphasiques des langues romanes*, art. 201–206
Trumper, *Slang and jargons*, CambrHist 1,14

4.4 Linguistic proximity and linguistic distance, speech and writing

4.4.1 Oral sources in linguistic analysis

Until now, we have scarcely mentioned the parameters of speech and writing, which are nevertheless constitutive of language use, variation and change. Every utterance, without exception, can be classified according to this distinction in medium as it is either oral (phonic) or written (graphic) – and the choice between the two modalities has significant consequences for its physiognomy.

Contrary to what one might expect, linguistics has always favoured written sources for analysis because they are more easily available and accessible than oral ones. The

study of the latter requires direct contact with the respective speakers, whereas a written text can effortlessly transcend space and time, especially in printed form. Moreover, oral texts first have to be laboriously transcribed – i.e. transferred into graphic code – before they can be analysed. Even though language is primarily based on orality, linguistics tends to be based on writing, leaving ample opportunity for misunderstandings to arise.

Systematic collections of oral data were first made in the field of dialectology at the beginning of the 19th century (cf. 2.2.2 no. 3 and 4.2.1 no. 2). Throughout the century such work led to the identification and documentation of a large number of spoken varieties. Since approximately 1900, the collection of sources has been expanded by the creation of dialectal atlases. Although these generally focus on words in isolation, they also reproduce larger units of text, and they allow a more systematic appreciation of geolinguistic variation than the 19th-century repertoires. During the same period, dialectologists were followed by phoneticians, who developed refined phonetic alphabets, enabling the accurate transcription of speech into writing (cf. e.g. the revolutionary recordings *Archives de la parole*, created by Brunot from 1911–14). Sociolinguistics has greatly contributed to methodological advances in the field of dialectology, as has research on oral literature, which focuses on larger texts.

A large number of oral corpora exist for the 20th and 21st centuries, allowing the study of discourse traditions that are specific to spoken language. Examples include:¹⁰

- Pusch *et al.* (eds), *Romanistische Korpuslinguistik I: Korpora und gesprochene Sprache*, 2002 (for a synthetic overview of currently available corpora for different Romance varieties, cf. pp. 245–264).
- Cresti/Moneglia, *C-oral-Rom: Integrated Reference Corpora for Spoken Romance Languages*, 2005.

The incorporation of spoken language into linguistic research is indispensable in enabling general thought on linguistic sources and their material aspects. Whereas research on historical documents must contend with the problem of a limited number of sources that reflect language use in an incomplete manner (cf. 4.5.4 no. 1), studies focusing on present-day language face the opposite problem: today's Romance languages can be observed on four continents and through a potentially unlimited number of sources, both written and oral.

¹⁰ An exemplary balanced (i.e. representative) oral corpus is provided by Biber (*Variation across speech and writing*, 1988), for English. Regrettably, no such resource exists for the Romance languages.

4.4.2 The theorisation of linguistic immediacy and distance

The theorisation of the distinction in medium between oral and written communication was put forward by several German Romance linguists only two decades after that of the diasystem. In 1974, Ludwig Söll was the first to specify the fundamental difference between the linguistic conception and the realisation of a message.

According to this theory, the concrete mode of expression (*phonic code* or *graphic code*) is only a secondary factor, contrary to what might be expected. The most essential component is the conception of a message (*oral* or *written*), since it determines the nature of discourse and the physiognomy of the utterance.

To give some examples, a discussion broadcast on television (oral conception) can be transcribed and published in a magazine (graphic code); the resulting piece of writing will, however, still reflect the initially oral conception of the discussion. The same can be observed with regard to emails, especially those exchanged among young people; these are always written (graphic code) but they reflect a relatively high conceptual proximity to orality. In contrast, a university lecture, although delivered orally (phonic code), is frequently read from a written template; in either case, it corresponds to a conceptually rather complex written text, planned and constructed with a certain amount of effort (written conception). This is even more striking in the case of court rulings which, although conceived in writing, are always read aloud.

The distinction between speech and writing was further clarified by Peter Koch and Wulf Oesterreicher, who precisely identified the differences between linguistic distance (German: *Distanzsprache*) and proximity, or immediacy (German: *Nähesprache*). Communicative distance is thus characteristic both for written production and for public speaking (or interaction between strangers), whereas communicative immediacy is usual for familiar, intimate and spontaneous interactions. Oral language production by its very nature is mostly – but not always – correlated with linguistic immediacy, and written production with linguistic distance, although there are some exceptions here too.

The theory was first developed by the two friends in their fundamental article published in 1985 (*Sprache der Nähe – Sprache der Distanz*) and exemplified shortly afterwards in a textbook encompassing French, Italian and Spanish (*Gesprochene Sprache in der Romania*, 1990, translated into Spanish in 2007 and significantly expanded in a second edition in 2011). Over the course of the following twenty years they disseminated their theory in many different languages. Koch and Oesterreicher portray the continuum between linguistic immediacy and distance in the shape of a parallelogram:

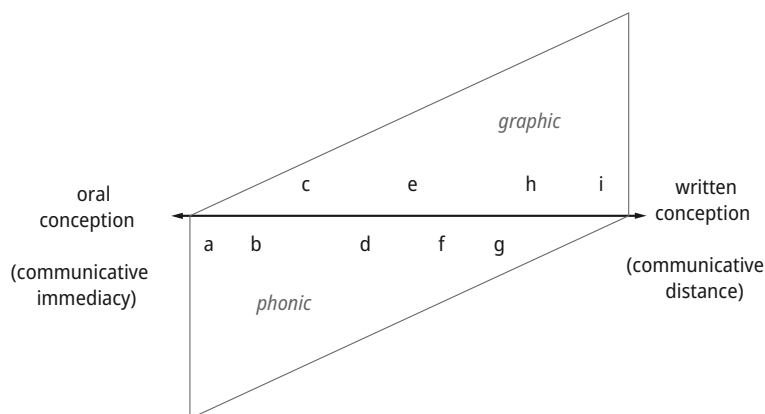


Fig. 18: The continuum of linguistic distance and immediacy

Source: Koch and Oesterreicher (1994: 588; the textual genres illustrated have been slightly adapted here).

Phonic:

a: telephone conversation between friends

b: familiar conversation

d: job interview

f: sermon

g: academic lecture

Graphic:

c: private letter or emails

e: interview published in a magazine

h: newspaper article

i: legal text

Every possible communicative situation thus lies somewhere between the two extremes of immediacy (corresponding to an oral conception) and distance (corresponding to a written conception). The distribution of phonic and graphic utterances in the diagram reflects the strong – but not absolute – affinities between the phonic and graphic modalities and the two corresponding forms of conception.

→ Selig, *Diamesic variation*, CambrHandb 29

4.4.3 The particular characteristics of linguistic immediacy

The distinction between linguistic proximity and distance is a general and anthropological feature. Strategies of immediacy and orality imply a reduction in the degree of discourse planning: they are characterised by a simple, linear discourse structure and simple, cumulative syntactic patterns ('parataxis'). Distance and written conception, on the other hand, exhibit a hierarchical and complex structure and a strong presence of syntactic subordination ('hypotaxis'). In addition, vocabulary is generally less varied in utterances reflecting proximity and includes many repetitions and filler words. There are three universal features that occur in concentrated form in orality:

1. Speech first of all displays a large number of ‘discourse markers’ (or ‘pragmatic markers’) that serve to structure discourses (‘discourse particles’, cf. Remberger, *Discourse and pragmatic markers in the Romance languages*, OxfEnc). These include¹¹:

- markers of discourse structuring (indicating the beginning of a sentence or text: *oui/ouais* “yes/yeah”, *eh bien* “hm, well”, *écoute* “well, listen”, *tu sais* “you know”; or its end: *non, n’est-ce pas* “isn’t it (etc.), you know, right”, *hein, quoi* “huh, eh”, *tu sais* “you know”);
- turn-taking signals (Fr. *écoute* “listen”, *ouais* “yeah”, *n’est-ce pas* “isn’t it (etc.), you know, right”);
- emotive markers signalling interaction on the part of a speech partner (Fr. *hm* “hm”, *oui/ouais* “yes, yeah”, *d’accord* “ok”, *voilà* “there you go”, *tiens* “gosh”, *voyons* “let’s see”, *c’est vrai* “that’s true”, *sans blague, dites, dis donc, hein* “really, gosh, you don’t say, no kidding”, It. *si, ecco* “there you go”) or the speaker (Fr. *hein, non, n’est-ce pas* “isn’t it (etc.), you know, right”, *tu sais* “you know”);
- hesitation and reformulation phenomena (Fr. *je sais pas* “I dunno”, *et puis euh bon ben* “and then, er, well”, *bon alors* “ok, well, then”; *enfin* “anyway”, *je veux dire* “I mean”, *bref* “anyway”, Sp. *o sea, y, pues*);
- interjections (Fr. *ah* “ah”, *aïe* “ow, ouch”, *bah* “ha”, *bof, oh, ouf* “oh”, It. *ehì*);
- modulation (Fr. *t’as qu’à pas t’en faire* “just don’t worry about it”; *mais* “but”, *quand-même* “anyway”, *donc* “so”).

2. In syntax, the general organising principle of spoken discourse is *parataxis*, including some complement and relative clauses, but lacking complex patterns of subordination. Wolfgang Raible has termed this principle ‘aggregation’, as opposed to ‘integration’, the latter being typical of writing (cf. *Linking clauses*, 2001, LangTyp, art. 46).

What is most apparent in all transcriptions of speech is the strong presence of incomplete or grammatically non-coherent sentences, illustrated in the French example below:

elle a une petite amie, finalement, qui s'appelle Y. Y, je pouvais le concevoir, mais enfin Z. Mais, je, je t'assure qu'elle nous a vraiment ... Mais je me suis, on s'est demandé à un moment si, euh, ... elle allait nous, nous autoriser finalement à l'adopter, à l'adopter réellement, enfin, tu vois, à avoir le jugement. Bon, oh, je, je commençais à bouillir. Mais vraiment, on était accusé! Tu vois un peu le ... Et c'était ... Il y avait quoi? il y avait deux petites jeunes, là, alors là, donc, qui s'étaient, qui étaient sûres d'elles, et puis ... (Stark 1997, 341)

Literally: she’s got a girlfriend, finally, called Y. Y, I could imagine that, but anyway Z. But, I, I’m telling you that she really [us]... But I, we wondered at one point if, um, ... she was going to allow us, us in the end to adopt her, to really adopt her, you know, to get the judgement. Well, ooh, I, I was starting to boil. But really, we’d been accused! You can sort of see the... And it was... What was there? There were two young ones, so then in that case, so, who had, who were sure of themselves, and then ...

¹¹ For the examples, cf. Koch/Oesterreicher ²2011 as well as, for French, Blanche Benveniste (*Le français parlé: études grammaticales*, 1990, and *Approches de la langue parlée en français*, ²2010).

The brain easily finishes incomplete sentences and glosses over repeated (and corrected) attempts at formulation. Consequently, awareness of this phenomenon is very low among speakers.

3. The vocabulary or lexis of spoken language is characterised by low diversity. In addition to a propensity for repetition, filler words are frequently inserted; this contrasts with the tendency towards expressiveness (by means of variational borrowing in particular, *cf.* above). Additionally, speech is characterised by a large number of deictic elements (Fr. *ça, comme ça, là, ce/cette, celui-là*) and presentative constructions (Fr. *il y a, y a, y avait*).

In contrast, differences between speech and writing with regard to morphological and syntactic features in simple sentences remain limited. Contrary to common belief, spoken language does not display significant simplification in the domains of morphology and syntax, and the meanings of lexemes always remain stable in the two modes and conceptions of language. In short, discrepancies between speech and writing do not concern the principles of the linguistic system, but merely the concrete realisation of utterances, or their physiognomy.

→ RSG 3, sect. XIV, *Histoire de la langue parlée dans la Romania*, art. 207–211

4.4.4 Immediacy and the diasystem

Koch and Oesterreicher's model has gained acceptance well beyond the field of Romance linguistics over the course of the past thirty years. Nevertheless, one specific question remains, concerning the status of this new dimension of variation: according to Koch and Oesterreicher, it is a fourth dimension of the diasystem, which has been called 'diamesic'. Other linguists propose to view it as a general dimension – of a universal nature – that acts as an axis of orientation for the other three dimensions, which are of a historical nature and relate to specific elements within an individual language.

Assuming this second perspective, it seems reasonable to distinguish the three parameters of diasystematic variation (diatopic, diastratic and diaphasic) from the parallel continuum of linguistic immediacy and distance (including the conceptual and medium-related contrast between oral and written language). Both entities are always co-present in linguistic reality. Each utterance occupies a specific position both on the three axes of the diasystematic parameters and on the continuum of linguistic immediacy and distance.

This interpretation allows the manifold instances of interaction and interdependence of both entities to be explained without inconsistencies:

- from a diastratic point of view, the linguistic features expressing communicative distance correspond to high social prestige, those expressing immediacy to low social prestige;

- the use of a standard variety in speech therefore immediately establishes distance between speakers, whereas an informal or popular variety produces the effect of proximity or immediacy. This opposition, however, directly involves the diaphasic dimension: standard varieties come into play, above all, in formal contexts or monologues, while informal varieties are characteristic for dialogues, often between speech partners who are close to one another and from the same social group;
- from a diatopic point of view, a dialect variety which is strongly marked is, almost by definition, positioned close to the pole of immediacy;
- a standard variety in writing may reflect the same communicative values as a more informal variety in speech, while the use of a standard variety in speech is easily equated with supra-standard marking. In the same manner, some diaphasic varieties remain more or less restricted either to the domain of written conception (such as the registers of literary style) or oral conception (as is the case with youth language).

The most strongly marked diatopic and diastratic varieties are thus always found in more or less the same position on the continuum of distance and immediacy. Diversification is more significant for diaphasic varieties or textual genres, which reflect extremely variable communicative contexts: official letters are close to the pole of distance; letters or emails exchanged among friends are closer to the pole of immediacy. Legal or scientific texts express a maximum amount of communicative distance, as do university lectures. In contrast, a conversation among scientists may contain many elements characteristic of immediacy.

These examples also illustrate the fact that the opposition between communicative distance and immediacy plays an important role in language elaboration, which is, in turn, a decisive parameter with regard to evolutionary trends in the diasystem. In tribal societies, only varieties conveying linguistic immediacy are well established; the linguistic expression of communicative distance can only develop if different social groups coexist.

The linguistic expression of communicative distance becomes highly intensified as a result of writing. Writing allows communication across space and time and creates a strong sociological distinction between literate and illiterate individuals. Writing furthermore enables and catalyses the formation of elaborated textual genres and standard varieties, in contrast to spontaneous written varieties. Only under these circumstances can the continuum between communicative distance and proximity attain a high degree of development in terms of linguistic expression.

Historically speaking, linguistic features expressing communicative distance and the multitude of established textual genres are a relatively recent phenomenon, linked to the emergence of written language and the evolution of complex societies, the latter being responsible for the invention and development of language as a tool without which they cannot function.

4.5 Discourse traditions and pragmatics

4.5.1 Textual genres, discourse traditions and the diasystem

1 Textual genres and discourse traditions

We have seen that any given utterance (i) complies with the rules of a language, (ii) is rooted in the linguistic diasystem and (iii) is positioned on the continuum between linguistic immediacy and distance. At the same time, each utterance (iv) falls within the linguistic models that are provided by textual genres or discourse traditions. These genres – spoken or written – reflect recurrent communicative situations, for which, over time, a set linguistic form becomes established: legal texts, letters between friends, job interviews, sports commentaries, etc.

Textual genres are extremely diverse: novels, poetry, newspaper articles, readers' letters, advertisements, telephone calls, sales discussions, conversations in the lift or around a table at a restaurant. Such genres may become 'traditional' and constitute by this means discourse 'traditions', as illustrated by examples such as legal discourse or epistolary genres, which may partially merge with corresponding diaphasic varieties.

Textual genres and discourse traditions, however, form an independent and very heterogeneous third dimension parallel to that of the diasystem and the immediacy-distance continuum. Consider as an example the textual genre of the official letter: in modern elaborated languages, this genre requires a diatopically neutral or weakly marked variety. From a diastratic and diaphasic point of view, it expresses high social prestige; moreover, the text can be attributed to a clearly defined discourse tradition – administrative language. By means of these different elements, official letters are clearly positioned towards the pole of linguistic distance. At the same time, this type of letter also fits within the rules of its genre: through its layout (address, paragraphs, signature) as well as through the use of stereotypical formulas ('we very much regret that...', 'I am pleased to confirm...'). The rules of this genre, in turn, have a historical tradition that goes back to Latin and Greek Antiquity, through medieval and modern epistolary rhetoric; hence, it is a tradition that transcends individual languages and falls within a cultural or even anthropological framework.

A genre can also make use of different diasystematic varieties at the same time. This applies, for example, to novels, in which various diaphasic or even diastratic varieties may be exploited (everyday or formal language, but also informal or popular varieties, or those spoken in suburban areas). A genre can also include elements belonging to different discourse traditions – such as administrative, legal or medical language, epistolary style or dialogues – for stylistic purposes.

Internal diversity is omnipresent in the sphere of discourse traditions. For example, scientific and technical language is a feature of all modern academic disciplines: medicine and biology, law and economy, philosophy and theology, literary studies and linguistics.

f these disciplines has, in turn, developed its own terminology, syntactic preferences and specific textual models. An extreme case is that of purely formal languages (used in the domains of mathematics and physics in particular).

All these discourse traditions fulfil highly specific communicative functions. At the same time, both the meaning of the terminology they employ and the general construction of the texts to which they give rise (*cf.* 1.2.4 no. 2) form an intelligibility barrier for non-specialists.

Moreover, they all manifest themselves both as marked forms, such as the manuals of the disciplines in question, and as diluted and therefore more accessible forms, such as dialogues between specialists and non-specialists (e.g. doctor/patient, lawyer/client, professor/student). Different degrees of marking add a 'vertical' dimension to scientific languages.

Genres and discourse traditions reflect general conditions of communication and transcend the level of individual languages. The epistolary genre, novels or scientific styles share internal characteristics in all languages, displaying similarities in grammatical features and vocabulary. Science is international and this extends to its means of linguistic expression. This is evident with regard to terminology; consider a term such as *psychosis* in various other languages: Fr. *psychose*, It. *psicosi*, Sp. *psicosis*, Pg. *psicose*, Rom.n *psihoză*, Cat. *psicosis*, German *Psychose*). However, distinctive cross-linguistic peculiarities can also be found in syntax (*cf.* Forner, RSG 1, art. 170). This is all the more pronounced in the Romance languages, which remain close to one another due to their common origin and history.

Discourse traditions, however, in no way reflect linguistic genealogy. While there is a continuum of primary Romance regional dialects and while it may be useful to compare non-standard varieties of the different Romance languages with one another, the continuum of present-day scientific language has no real 'Romance' dimension: it encompasses all widespread standard languages, above all English, French, German and Russian, but excludes less elaborated Romance varieties. The expansion of scientific and technical languages in the modern world thus leads to convergent tendencies, which in some cases go against the 'genetic' origins of the languages in question, whereas in others they contribute towards strengthening an original relationship. This phenomenon is neatly illustrated by derivational patterns in particular (e.g. derivatives ending in Fr. *-iser*, It. *-izzare*; Engl. *“-ize”*, *“-ise”*).

→ Forner, *Prinzipien der Funktionalstilistik*, RSG 1, art. 170

Pountain/Zafiu, *Register, genre, and style in the Romance languages*, CambrHandb 27

2 Textual genres and the diasystem

This brief presentation of textual genres and discourse traditions has already highlighted the rich diversity of textual genres today. Variation is present at all levels: 'sci-

entific works’ are not merely an unvarying product of the various sciences, but are, in turn, instances of diverse textual models within each discipline, from monographs to journal articles, reviews, debates or transcripts of lectures. ‘School textbooks’ cover the subjects taught at a specific educational level (primary or secondary), and take into account the age, level of schooling and type of education of the pupils to whom they are directed. Even though scientific works and educational works may overlap in terms of the subjects they treat and their linguistic form, the two categories are nevertheless immediately recognisable and are easy to distinguish from one another.

Such overlaps are numerous, to the extent that it is nearly impossible to incorporate all textual genres within a single coherent hierarchical structure. Our communicative universe fits entirely within these genres, which causes an immense degree of complexity. A categorisation of textual genres is thus necessarily always partial and approximate. The following figure displays the utterance types that we have already seen categorised according to their position on the continuum of immediacy and distance and their correlation with the parameters of the diasystem, classified in a very approximate manner according to textual genre:

dialogue types	letters	journalistic writing	religious writing	scientific writing
a b d (e)	c	h e	f	g i

Fig. 19: Approximate classification of utterances according to textual genre

a: Telephone conversation among friends	f: Sermon
b: Informal conversation	g: Scientific conference paper
c: Private letter	h: Newspaper article
d: Job interview	i: Legal text
e: Interview in a magazine	

The above categorisation naturally represents a simplification: the mere fact that an interview and an article are both published in a newspaper does not mean that they belong to the same genre (*cf.* Schweickard, *La cronaca calcistica*, 1987, ch. 0.1). Nevertheless, this figure illustrates the fact that the textual foundation of an utterance needs to be distinguished from its variational foundation and its medium: hence, each utterance occupies a defined place within each of the three diasystematic dimensions, on the immediacy-distance continuum, as well as within the conglomerate of existing textual genres.

Though textual genres and diasystematic varieties represent two separate entities, there is nevertheless strong interaction between them. The structural or linguistic ele-

ments that are characteristic of one given textual tradition may find their expression in other traditions or even serve as models. The elaboration of the modern media and the textual traditions that are inherent to them has had considerable consequences for present-day language (cf. Raible, *Medien-Kulturgeschichte*, 2006), including a significant reorganisation of the diasystem.

Generally speaking, diaphasic varieties draw from linguistic features inherent to specific textual genres, as regards both their functioning and their historical development. Register differences in particular rely on well-defined textual traditions: diaphasic varieties thus arise from the generalisation of linguistic elements considered prototypical for specific textual genres.

Major discourse traditions thus bear witness to the most salient diaphasic varieties of language (cf. RSG 2, sect. XIII, *Domaines de la communication, médias et types de textes*):

- literary language (cf. RSG 2, art. 172–178)
- religious language (RSG 2, art. 179–182)
- scientific and technical language (RSG 2, art. 190–183) as well as economic language (RSG 2, art. 187), the development of which was initiated in the Middle Ages, and which have undergone intensive elaboration since 1900 (cf. RSG 2, art. 194, terminologies)
- legal and administrative language

Further varieties emerged in the modern period, including the language of politics (cf. RSG 2, art. 183–186) as well as varieties and genres associated with the media (i.e. the press and, more recently, audiovisual media, art. 195sq.), advertising (art. 188) and, since the 21st century, sports (art. 189; cf. also Pöckl, *History of languages for specific purposes*, and Garriga Escibano, *The language of chemistry in the Romance languages*, OxfEnc).

In all these cases, an extremely close interaction develops between discourse traditions and diaphasic varieties. At the same time, the discourse traditions concretely shape the corresponding diaphasic varieties and are determined by their dynamics and functional purposes. Although in synchrony the two entities exist alongside one another, in diachrony, each contributes to the evolution of the other.

4.5.2 Issues in text linguistics

Text linguistics is a fairly recent development. It began in the 1970s and 1980s as a subdiscipline of linguistics. In part, its aims are those of traditional stylistics, focusing on literary genres. Since its establishment, scientific and technical genres have also been subjected to more in-depth analysis (resulting in a new subdiscipline known as *Fachsprachenforschung*, i.e. research on specialised languages). Parallel to these studies, methods of discourse analysis and conversation analysis have developed within the

field of pragmatics (*cf.* 4.5.3). Thus, several linguistic trends are involved in the analysis of textual genres, each one addressing different questions.

Central issues concern (1) the semantic coherence of a text and (2) the way in which cohesion is achieved through linguistic means, as well as (3) information structure (topic – comment) and (4) textual structure in general. These four areas can be summarised as follows:

1. Coherence is more encyclopedic than linguistic, concerning the logical, temporal and referential links established within a given text, and those established between the text and the knowledge possessed by the reader.
2. Cohesion, in contrast, is a linguistic category, belonging to the domain of macro-syntax; it is based on an array of significant elements that create an explicit order within texts (*cf.* Raible, *¿Qué es un texto?*, 2000), particularly the following:
 - metatextual markers (announcing a topic) and prosodic markers that indicate the beginning and end of a text;
 - anaphoric elements that refer back to previously-mentioned elements (Fr. *il, là-bas*);
 - connective elements that create links between different sentences or parts of a text (Fr. *en effet, toutefois, mais*);
 - temporal markers that indicate the progress of an action (Fr. *ensuite, auparavant*).
3. Information structure focuses above all on the structure of the topic (given information, T) and the comment (new information, C) normally present in explicit or implicit form in a sentence. If the topic has been mentioned beforehand, the structure of the sentence is unmarked, and it begins with the topic:

Mais l'homme [= T] refuse de répondre [= C]
 “But the man [= T] refuses to reply [= C]”

The topic can also be stressed, in which case the subject is ‘topicalised’:

Moi [= T], je suis fatigué [= C]
 “As for me [= T], I am tired [= C]”

If, on the contrary, the topic has not been mentioned before, the structure of the sentence is ‘focalised’:

C'est Paul [= C] qui m'a donné le livre [= T]
 “It was Paul [= C] who gave me the book [= T]”

If the topic is known to the hearer but has not been mentioned in the utterance context, one speaks of a neutral structure:

Tes cousins [= C] s'annoncent pour demain [= T]
 “Your cousins [= C] are coming tomorrow [= T]”

4. Finally, text structure conforms to the model prescribed by the textual genres concerned; it manifests itself through fixed formal elements (e.g. prosodic patterns, rhymes or a defined syntactic structure).

We will not examine these subjects in any depth here, since the characteristics of text linguistics are less specific to a given language or language family than diasystematic variation (for information structure, however, cf. 8.4.4).

4.5.3 Principles of pragmatics

Since pragmatics, like text linguistics, is not Romance-specific, this field will be presented only succinctly here. Pragmatic factors are, nevertheless, fundamental for language: every utterance pursues extralinguistic objectives (of a pragmatic nature), which determine its specific linguistic form. Such pragmatic intentions correspond to the three semiotic functions defined by Bühler: the cognitive function (*Darstellung*, representation of referential data and elaboration of thought), the expressive function (*Ausdruck*) and, above all, the ‘conative’ or social function (*Appell*), i.e. the desire to influence others (cf. 1.2.1).

Generally speaking, the pragmatic foundation of a sentence thus simply links it to the extralinguistic communicative context and to the persons implicated in the utterance act: from this point of view, the act of speaking or writing is subordinate to other forms of human interaction and their respective rules. The pragmatic framework also determines the type of a sentence (assertive, interrogative or injunctive) and the corresponding markers, such as the use of the essential techniques of modalisation (German *Abtönung*), as illustrated below by an example from French:

Marie l'a sans doute rencontré vs. *Marie ne peut pas ne pas l'avoir déjà rencontré, si ?*
 “Mary has probably met him” vs. “Mary can't possibly not have met him already, can she?”

The pragmatic foundation of an utterance also concerns the personal positions assumed by speakers with regard to its content, such as hesitation, reflection or emphasis, which are thus expressed by forms of topicalisation:

neutral: Fr. *Jean vient* “Jean is coming”; Sp. *¿cómo está el agua?* “what is the water like?”
 → with pragmatic salience: *Jean, il vient* or *il y a Jean qui vient*; *el agua ¿cómo está?*

More specifically, pragmatics studies the grammatical or lexical elements that anchor an utterance within a given context and allow its actualisation (for example by means of topicalisation, deictic elements or periphrasis). Pragmatic factors may also lead to phenomena of grammaticalisation.

A well-known example is the origin of the definite article, which plays a decisive role in sentence structure (cf. 2.3.1 no. 3). In the Romance languages, the Latin demonstrative *ILLE* “that one (over there)” developed into a grammatical marker, the definite article (It. *il*, Fr. *le*, etc.). This change is based on the semantic value of individualisation: the given object of which one is speaking is highlighted.

→ Radtke, *Historische Pragmalinguistik*, RSG 2, art. 199

Held, *Schwerpunkte der historischen Pragmalinguistik*, RSG 2, art. 200

1 Speech acts, the cooperative principle and implicature

A fundamental, more specific field of study in pragmatic linguistics concerns utterances that carry strong pragmatic connotations, above all *speech acts* or *illocutionary acts* (German *Sprechakte*), but also dialogues and conversation in general. A performative speech act makes use of utterances in order to accomplish an action that has direct impact on the persons it implicates (German *Sprachhandeln*). This property of language was first identified and described by the language philosopher John L. Austin (1911–1960), particularly in his posthumous book *How to do things with words* (1962).

The prototypical case is that of illocutionary acts, where words themselves have the value of an action (e.g. ‘I now call the meeting to order’, ‘I hereby pronounce...’, ‘I beg your pardon’, ‘I promise that ...’); the same applies to orders (e.g. ‘Have a seat’). In other cases, speech acts remain indirect (e.g. ‘I’m feeling a little peckish ...’, as a discreet reminder that it might be time to eat). This distinction was introduced by the philosopher John Searle (*1932), who contributed greatly to the development of speech act theory (cf. *Speech Acts*, 1969, etc.).

At the same time, there has been interest in the study of dialogue as communicative action. This type of rhetorical pragmatics played an essential role in the development of discourse analysis and conversation analysis, which, in turn, intersect with text linguistics. Examples of marked pragmatic elements are opening and closing formulas in interactions, such as salutations (‘Hello’, ‘Hi’, ‘Dear friends’; for Old French, cf. Lebsanft, *Studien zu einer Linguistik des Grusses*, 1987).

A new contribution towards an understanding of the phenomena of communication was the establishment of the cooperative principle in 1957 by the philosopher Herbert Paul Grice (1913–1988; cf. *Studies in the way of words*, 1991). Formulated paradoxically as ‘we act as if we wanted to cooperate’, this principle can truthfully be considered to be fundamental to all communication (cf. Lebsanft, *Kommunikationsprinzipien*, 2005: 26–28). On this basis, Grice formulated four conversational maxims characteristic of successful utterances: communicative relevance, quality in terms of veracity, quantity in terms of information supplied, and manner; i.e. clarity of expression. He also developed the theory of implicature, above all conversational (‘Get up!’ – ‘But it’s Sunday!’); a

comprehension of such utterances requires additional information supplied by contextual and encyclopedic knowledge shared by speakers ('inference').

2 Pragmatics, the diasystem and textual genres in language functioning

Within the framework of language production, concrete manifestations ('*parole*' according to Saussure; 'performance' in Chomsky's terms) are determined by the abstract principles of '*langue*' (or 'competence'). The tension between these contrasting entities forms a link between variational, textual and pragmatic linguistics and the systemic components of language described in the following chapters.

The pragmatic intentions of a speaker form the starting point of each utterance. In order to realise these intentions, the speaker calls on his or her linguistic knowledge (competence). This very fact, however, presupposes diasystematic choices on the part of the speaker, i.e. the use of a given diatopic, diastratic or diaphasic variety (e.g. everyday French in Belgium, the Italian regiolect of Rome or informal Spanish of Mexico). A further possible choice is the use of a more or less marked discourse tradition (e.g. the French used in academic linguistic circles or youth language in Italy). Diasystematic parameters determine the position of the utterance on the continuum of linguistic immediacy and distance. As an illustration, the five examples mentioned above can be placed in the following order, beginning with varieties closer to the pole of communicative distance: academic French used in linguistics; everyday French spoken in Paris; the regional French spoken in Marseille or Strasbourg; informal French in Quebec; youth language in the suburbs of Lyon.

The communicative situation then dictates the appropriate linguistic model – the genre or the discourse tradition. If one is writing an official letter, elements of the opening section, the salutation, as well as a number of stereotypical formulas are prescribed in advance; the same goes for a sales conversation in a bakery or a supermarket. These elements of enunciation are not generated solely by a general form of language competence, but rather by means of the actualisation of micro-texts that are part of the collective memory of one and the same language community. Speakers model their production on other similar elements of discourse that belong to the same textual genres and draw from a vast inventory of prefabricated elements of language.

The production of '*parole*', then, consists in the reproduction of already existing forms rather than an act of free creation. Lexical and syntactic choices within a textual genre are restricted, thus making the genres recognisable. One sentence is sufficient to identify a legal, religious or medical text, a spontaneous conversation or a planned discourse.

The production of '*parole*' thus displays four logical stages:

1. The individual utterance is conceived according to the requirements of the pragmatic objective pursued.

2. It thereby reproduces the structure of other utterances that have similar purposes, respecting the precise rules that govern the structure of the textual genre in question; e.g. if a cross-examination in court is not recognisable as such, the audience will be surprised or caught off guard (which, in some cases, can be a purpose in itself, *cf.* no. 3 below).
3. The textual genre then exploits the parameters of the diasystem, the rules of which it must, in turn, respect: diatopic, diastratic and diaphasic marking must remain coherent with the knowledge possessed by the listener (unless the intention is to deceive him or her), placing the utterance within a socio-cultural framework.
4. Finally, the utterance must conform to the functional oppositions and rules of the linguistic system (phonological system, grammatical principles, lexical inventory), in order to remain intelligible. As we have seen, this is the most essential condition of communication, as extreme ungrammaticality or an atypical use of lexis may breach the chain of communication (*cf.* 1.2.2).

From a terminological point of view, the diasystem and textual genres within the above chain of logical dependencies can be equated to the concept of ‘usage norms’, situated by Coseriu between the extreme poles of *parole* and *langue*. The term ‘usage norm’ is reserved in particular for recurrent cases of language use (salutation formulas, habitual expressions, verb valency, etc., *cf.* 1.2.3), though it can also be identified with certain aspects of the linguistic diasystem.

Beyond the terminological debate, the identification and description of these four stages allows a better understanding of the dynamic potential inherent in their interaction. The great variability of language allows any utterance to modify the rules of the textual genre or discourse traditions to which it belongs. When modifications become recurrent, they lead to the transformation of a genre over time.

German research has examined various cases of change of this type, exemplified by the genre of ‘judicial decisions’ (*cf.* Krefeld, *Das französische Gerichtsurteil in linguistischer Sicht*, 1987) or by the evolution of specialised texts in economics (*cf.* Kaehlbrandt, *Syntaktische Entwicklungen in der Fachsprache der französischen Wirtschaftswissenschaften*, 1989; Rainer, *Geschichte der Sprache der Wirtschaft in der Romania*, RSG 1, art. 187; *Id.*, *The language of economy and business in the Romance languages*, OxfEnc).

Changes within the sphere of textual genres can, in turn, cause changes in diasystematic varieties. Such transformations end up affecting the linguistic system as well. As an example, the tendency towards nominalisation (both in terms of word formation and syntactic construction), which can be seen in scientific language, has become a generalised characteristic of specialised languages of the beginning of the 21st century.

There are thus reciprocal relationships of dependency between these four parameters of language production: the system and diasystem determine the concrete form of the textual genres, which, in turn, predetermine the form of the utterances (*‘énoncés’*); conversely, variation within utterances can lead to changes in the textual genres, which

then have repercussions on the diasystem, and finally on the configuration of the language.

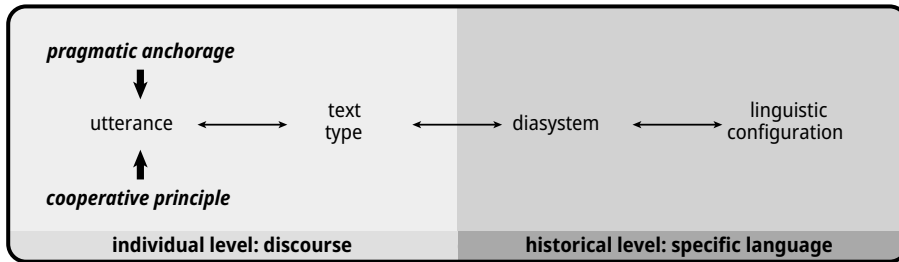


Fig. 20: Interdependence between the four parameters of language production (source: Glessgen, *Diskurs-traditionen*, 2005: 217).

The diagram distinguishes between the individual level of production (utterance, text type) and the historical level of the individual language (diasystem, configuration), which corresponds to the separation between *parole* and *langue*, as well as between language use and the rules governing it (cf. Lebsanft, *ibid.*: 30–33).¹²

There are preferential links between pragmatic goals and the language forms used to achieve them (although naturally, a degree of flexibility remains as regards their choice). Consequently, genres have developed a relatively fixed linguistic structure, motivated by recognisable communicative intentions. A scientific text, for example, should be self-explanatory and comprehensible out of context – writing in this genre thus implies adherence to established conventions which correspond to these aims. A conversation among youths, on the other hand, should be expressive and lively, and necessitates the use of other linguistic means.

4.5.4 The significance of textual genres for linguistic analysis

1 The role of textual genres for the analysis of present-day language and their effects on usage

The concept of textual genres allows the categorisation of a vast quantity of texts, and the categorisation of these texts, in turn, facilitates our access to them. At the same time and for the same reason, textual genres are situated at the interface between recurrent,

¹² The concept of the ‘historical level’ was introduced by Coseriu, who distinguished three levels of language: a ‘universal level’ for linguistic activity in general (‘speech in general’), absent from fig. 20, a ‘historical level’ for specific languages (‘linguistic configuration’) and an ‘individual level’ for specific language production (‘utterance’). In fig. 20, a further differentiation between ‘text type’ and ‘diasystem’ has been introduced.

spontaneous language use and the unit formed by the linguistic system and diasystem, as we have just seen. As we have also seen, these characteristics of textual genres are correlated with their importance for the emergence of language varieties as well as with their impact on language change: with regard to language functioning, textual genres have a tangible influence on the other dimensions of language. At the same time, they are immediately rooted in the pragmatic framework that motivates language use. Taking into account their status and characteristics, text types thus provide a significant contribution to linguistic analysis.

The true extent of the influence of textual genres on the physiognomy, the habitual patterns and the evolution of language is generally underestimated, since the study of present-day languages, which can be directly observed for any specific type of research, is determined by intuitive knowledge:

1. Linguistic intuition enables one to judge the grammaticality of utterances (i.e. their coherence as regards linguistic configuration).
2. On the basis of intuition systematic linguistic tests can be conducted on native speakers in order to evaluate the potential use of a specific language, even if it is non-documented (e.g. does a reader understand certain words or sentences?).
3. Intuition also makes it possible to evaluate the significance of various diasystematic markers in society, and to determine, for instance, whether a given form is endowed with high prestige, whether it is a technical term, etc.
4. Present-day language is observable in all its manifestations, both written and spoken; dictionaries, descriptive grammars and, more recently, digitised corpora have become available for elaborated languages, facilitating their analysis and even allowing quantification.

Consequently, the multiple communicative markers inherent in every utterance can be identified, provided due caution is exercised. For the linguistic analysis of present-day language, the added dimension of textual genres is not essential in order to achieve meaningful results. Often, taking textual genres into account even complicates rather than facilitates research guided by intuition, particularly as the concept of the 'textual genre' is itself rather complex.

However, due to their recurrence, their stereotypical nature and their role as models for the production of (oral or written) utterances, textual genres and discourse traditions considerably increase the reliability of judgement and allow a more accurate understanding of language use. The rules set out by well-balanced grammars such as the *Bon Usage* or the *Grammaire méthodique du français* (GMF), in combination with the definitions provided by an excellent dictionary such as the *Petit Robert*, do not entirely describe current French usage, as they largely appeal to and reflect the intuitive knowledge of the users. Wherever uncertainty prevails, the authors of such works cite concrete examples in order to provide a precise illustration of the rules and restrictions of

use that apply to specific forms. It is in such cases, in particular, that questions relating to language varieties and discourse traditions come into play.

Textual genres play an essential role in all types of linguistic elaboration and they prefigure the current forms of communication: they thus constitute a significant source to be exploited for linguistic description and analysis.

2 The role of textual genres within a historical approach

For research with a historical perspective, the role of textual genres is still more essential and their importance for linguistic evolution and usage thus also becomes more apparent. As Christiane Marchello-Nizia puts it, ‘Comment raisonner sur une langue dont il n’existe plus de locuteur natif pour porter des jugements de grammaticalité ou pour exploiter le sens des énoncés?’¹³ (*Dire le vrai*, 1985). When there are no native speakers, intuition loses its immediate character, tests are impossible, and among the various manifestations of language, only written texts associated with a limited social elite are available for study.

We must therefore make do with a repertoire of textual genres that are very limited in number, whose character is mainly symbolic, and which are intimately linked to defined contexts of communication. The other parameters involved in language production, above all the diasystem and linguistic configuration, must be deduced from our existing knowledge of these textual genres. The internal and external characteristics of these genres thus merit more attention than they do in the context of present-day language. Fortunately – or unfortunately –, due to their restricted number, textual genres from earlier periods are comparatively easier to study.

Early textual evidence reflects only a very limited number of types of communication in comparison to all those that occurred in reality, even within elite social circles (cf. 11.1). As a consequence, these sources are likely to be of a more stereotypical and prototypical nature than written texts today, which are more varied and sometimes closer to spoken language. In textual genres predating the 20th century, recurrent utterances that were linked to specific contexts were transformed into partially fixed linguistic expressions. Hence, what was deemed important enough to be written down must have been of particular significance for the elite circles. Otherwise, the effort required to put it into writing would not have been justifiable.

It is evident that the sector of the diasystem that is reflected in early textual genres is extremely limited. In the first place, the written character of practically all evidence from these periods is invested with high prestige, placing the texts in question towards the far end of the ‘distance’ pole of the language continuum. In addition, strongly diatopically marked forms are excluded from writing. Finally, the pragmatic connotations

13 ‘How can one reflect on a language that no longer has any native speakers to actively express judgements of grammaticality and exploit the meaning inherent in an utterance?’

inherent in these written genres lend them a unique diaphasic status. Due to the vagaries of history, early texts thus came to reflect only the high-prestige variants of the diatopic, diastratic and diaphasic dimensions. As a consequence, the diaphasic dimension of historical diasystems as observed from our perspective today coincides with and can only be deduced from the dimension of textual genres.

Studies on historical language stages may incorporate comparisons with current forms (which are more easily accessible) as well as with dialectal forms – in so far as dialectal varieties of the language in question exist. Nevertheless, researchers are obliged to resort to the study of textual genres and the modalities of their transmission to a greater extent than is necessary for studies on present-day languages. By comparing different genres and identifying their internal characteristics, it is often possible to deduce whether and to what extent given grammatical or lexical features of a language are linked to particular contexts.

The fact remains, however, that literary texts are the only category to have been studied in depth by traditional linguistic historiography. Many philologists of the 19th century and the first half of the 20th century were both linguists and literary scholars (cf. 2.2.2) and their focus was naturally directed towards these elaborated genres with a strong aesthetic dimension and rich in meaning. Even today, historical linguistics relies to a great extent on literary sources. This holds true for more recent periods of language history as well: the number of literary citations in the descriptive grammars and reference dictionaries of all Romance languages is far greater than that of citations taken from all other textual genres. Diversifying the text types considered for analysis thus still holds considerable potential for empirical linguistics.

It may be concluded that textual genres and their corresponding discourse traditions have a place of their own among the abstract entities that govern language use and functioning, and that their consideration by research also has intrinsic value. In the following chapters, however, we will focus on systemic and diasystematic phenomena, returning to textual genres in the section on philology (11.1). Nonetheless, it is important to keep in mind at all times that the elements of language remain anchored within well-defined concrete contexts.

→ Weidenbusch, *Historische Textsortenlinguistik*, RSG 1, art. 197sq.

4.6 Open questions relating to variational theory

As the description of linguistic varieties has shown, the theory of variation is much more complex than one might at first imagine. A large number of questions have no definitive answer. These involve:

- the number and nature of the linguistic features that constitute a ‘variety’ (cf. 4.3.1, Hudson),

- the status of the continuum of proximity and distance (is it a fourth, ‘diamesic’ dimension of variation or not? Cf. 4.4),
- the question as to whether diatopic variation represents a dimension of a different nature (focused on speakers) to diastratic and diaphasic variation (focused on usage, cf. 4.3.1, Halliday),
- the concrete distinction between diastratic and diaphasic varieties, as well as that between diaphasic varieties and discourse traditions (cf. 4.3.3).

Even the fundamental question as to the status of variation in its relationship to the linguistic ‘system’ poses problems (cf. 4.1.1). Dufter and Stark (*La variété des variétés*, 2002: 81) emphasise the intrinsic difference between the supposedly *homogeneous and system(at)ic* character of Saussurean *langue* (and/or Chomskyan *competence*) on the one hand and the irreducibility and omnipresence of variation in natural languages on the other. Fully adopting the tenets of structuralism, Flydal and Coseriu suggested circumventing the problem by assuming the existence of discrete varieties that form linguistic subsystems with their own internal logic (cf. 4.3.1). This model, however, does not take into account the strong interaction that exists between different diasystematic parameters. The heterogeneity generated by variation seems real, both in use and in the brain’s control of language, even if it contradicts the supposed homogeneity of the linguistic system. In the author’s opinion, one could assume that linguistic forms are incorporated into several parallel neural networks at the level of the brain, each network representing a specific type of diasystematic marking; this would explain the ease and the frequency with which overlaps occur (cf. Glessgen, *Le statut du lexème*, 2011: 454).

The model of three (or four) variational dimensions is rife with inherent contradictions, and these are particularly apparent in the description of lexis, as this module displays by far the most intense diatopic, diastratic and diaphasic variation. This fact led us to suggest the potential usefulness of a new view of the diasystem (cf. Glessgen/Schøsler, *Repenser les axes diasystématiques*, 2018), based on the observation made by Coseriu that each utterance is by definition anchored within each of the three variational dimensions. According to our view, however, there are no purely diatopic, diastratic or diaphasic *varieties*, but only variational *marking* of a diatopic, diastratic and diaphasic nature. It thus seems logical to assume that the three variational dimensions principally correspond to three abstract parameters: geographical space, linguistic prestige and situational context. In this scenario, each utterance occupies a specific position within each of these dimensions (strong or weak diatopic marking, high or low linguistic prestige and strong or weak situational specificity).

If an utterance displays strong diatopic marking, it can be assimilated to a predominantly diatopic variety; here, the deviation from the traditional theory of variation is minimal. In the diastratic and diaphasic domains, on the other hand, this view leads to new perspectives: above all, it assigns discourse traditions to the diaphasic dimension, as can be seen in the following diagram. Here, the three diasystematic dimensions are

represented as columns, each corresponding to a continuum, along which the utterance types presented by Koch and Oesterreicher have been distributed in order to illustrate the phenomenon of language mediality (cf. 1.2.4 no. 2):

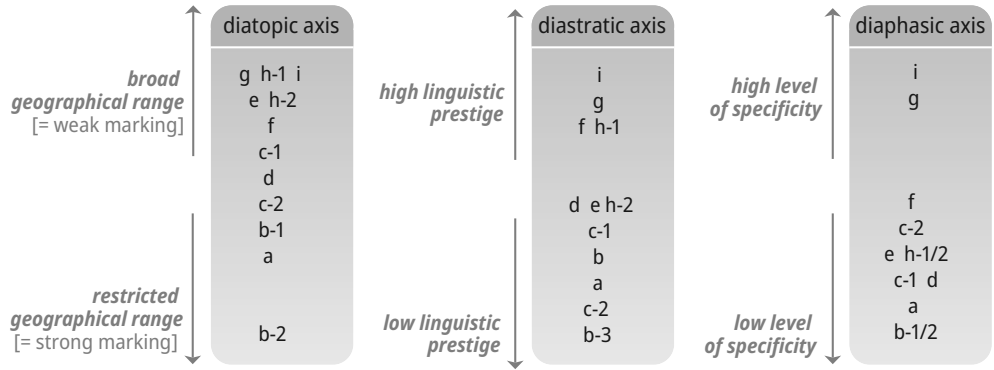


Fig. 21: Positioning of utterances in the diasystem

[linguistic immediacy]		
label	phonic	graphic
a	familiar phone conversation (Quebec, Mexico)	
b-1	familiar conversation (Marseille, Lausanne, Naples)	
b-2	familiar conversation in dialect (Évolène, Labaroche, Gèdre, Altamura)	
c-1		private letter (Lyon, Seville, Bologna)
c-2		personal email (La Chaux-de-Fonds)
d	job interview (Strasbourg, Bologna, Porto)	
e		interview published in a magazine (<i>Elle</i>)
f	sermon (Metz, Valencia)	
g	scientific conference (anywhere)	
h-1		newspaper article (<i>Nice Matin</i>)
h-2		newspaper article (<i>Le Monde</i> , <i>Corriere della Sera</i>)
i		legal text (anywhere)
[linguistic distance]		

Source: adapted after Glessgen/Schøsler, *Repenser les axes diasystématiques*, 2019: 33.

Examples:

- a familiar telephone conversation in Quebec [= a] has (1) strong diatopic marking; at the same time – within the diasystem of French – it has (2) low linguistic prestige as well as (3) a weak degree of specificity or technicality. It thus has the characteristics of a marked diatopic variety.
- a legal text [= i], on the other hand, (1) carries no diatopic marking, (2) displays elevated prestige and (3) a high degree of technicality. It is thus diastratically and diaphasically marked and at the same time correlates to a diastratic variety (the ‘supra-standard’) and a diaphasic variety (‘legal language’), etc.

A direct comparison of the three axes reveals both parallels and differences between them. Thus, legal texts (i) and colloquial conversations in dialect (b-2) occupy the two extreme poles of the three axes: the former has the broadest geographical extension, the highest linguistic prestige and the highest degree of specificity, at least among the selected examples, whereas the latter has the most restricted geographical extension and the lowest prestige and degree of specificity. Classification is more complex in the case of a newspaper article (h): while it generally displays weak diatopic marking, the degree of specificity may depend on its topic. Its prestige is also variable, depending on the quality of the newspaper.

A detailed discussion of each concrete case is beyond the scope of this chapter. Nonetheless, the diagram illustrates the necessity of considering all three axes on an equal footing for all utterances. From an empirical point of view, this results in a clearer picture of the interactions between the different axes.

Our interpretation remains consistent with the variationist tradition: a ‘dialect’ naturally corresponds to a ‘sociolectal variety’ (= characteristic of a group of speakers), with strong diatopic marking (= restricted geographical extension); what is generally called a ‘sociolect’ is also a sociolectal variety (= characteristic of a group of speakers), and while its linguistic prestige is often low, it can also be high (if, for example, it is accepted that academics participate in a sociolect); finally, a ‘style’ or ‘register’ corresponds here to a variety with strong diaphasic marking, but whose linguistic prestige and diatopic marking are variable. The differences between our theorisation and the traditional view are thus to be found on a more abstract level: in our view the concepts of dialect, sociolect or register become secondary, while the diasystematic parameters are primary.

This perspective constitutes an attempt to reconcile the somewhat diverse aspects of the variationist tradition. Its strengths lie in the combination of the abstract parameters of space, prestige and context on the one hand and the systematic and parallel consideration of the diasystematic, medium and genre-specific dimensions on the other. It thus enables a clearer description of the diasystem, significantly reducing the potential for internal contradiction.

Finally, it should be noted that variation neither interrupts nor destroys the cohesiveness of the linguistic diasystem. It is relatively easy to master multiple language registers in parallel, even though the average individual has a limited choice of diatopic and diastratic varieties at his or her disposal. For example, while it is possible to speak Lorrain dialect alongside a different regiolectal variety such as the Alsatian regiolect of French, knowledge of a third diatopic variety is rare. Despite freedom of expression, humans are thus tied to a place and a position in society.

4.7 Variational linguistics in the chapters that follow

Traditionally, the *Romània* provides an exemplary field for research on variation as a whole as well as on individual varieties, both synchronic and diachronic. Romance studies have played an essential role in the theorisation of variational linguistics; this, in turn, explains the importance that has been attributed to these issues here.

When reading the following chapters of this manual, it should be kept in mind that an elaborated language is built around a standard, which is a high-prestige variety that is relatively neutral in diatopic terms and that conveys strong communicative distance. Part 3, dedicated to internal structures and their history, focuses on the main evolutionary tendencies that have led to the formation of today's standard languages. Variational studies, in contrast, mostly focus on non-standard forms, which exhibit greater diversification, also in geographical space.

Although linguistic data relating to variation will not be treated in detail within this framework, they will be incorporated into our discussion, and we will regularly include examples of dialects and other less widely used varieties. In Part 4, devoted to the external history of the Romance languages, we will then attempt to provide a more solid foundation for the explanation of the variation encountered today, taking into account its historical background.

Part 3: **The internal structures of the Romance languages and their history**

5 Domains of language and historical periodisation

5.1 The domains of language

5.1.1 Traditional subdivisions

An ‘internal’ description of multiple Romance languages at different stages of their historical development highlights their shared characteristics, as well as their divergent features. Such a presentation of language change – which is both typological and comparative – is also useful to better discern the internal history of each individual language.

Four ‘domains’ of language, considered here to be fundamental, provide a descriptive framework. These domains are:

- phonetics and phonology (ch. 6)
- inflectional morphology (ch. 7)
- syntax (ch. 8)
- lexis, including proper nouns and word formation (ch. 9)

In addition to defining the domains of language, this ‘segmentation’ outlines the fundamental areas of internal linguistic study. Though it may appear simple at first glance, it raises some epistemological questions (i.e. questions concerning scientific theory) of seminal importance. In the following sections, we will begin by briefly defining the established domains. We will then outline some issues concerning terminology and delimitation, before returning to a discussion of the validity and implications of the proposed subdivisions.

1. Phonetics refers to the measurable acoustic sounds produced by speakers, while phonology concerns the abstract mental representations of these sounds, known as ‘phonemes’. Phonemes are organised on the basis of functional oppositions (so-called ‘minimal pairs’: Fr. /te/ <thé> “tea” vs. /de/ <dé> “dice”). These two domains are concerned with the formal aspect of language.

2. Morphology reflects the internal structure of words, and concerns both inflection (It. *parl-o* “I speak” vs. *parl-iamo* “we speak”) and word formation (Fr. *pays* “country” vs. *pays-age* “landscape”). The latter will, however, be discussed in the chapter on lexis, as it belongs in equal measure to this domain. Morphology presupposes the existence of different parts of speech, grouped primarily into two categories: lexemes (such as nouns or verbs), and function words (such as prepositions and conjunctions).

3. Syntax combines different lexemes and function words into larger formal and functional entities, from phrases to clauses to sentences (e.g. *Paul a vendu sa maison* “Paul has sold his house”).

4. Lexis consists of the lexical forms of a language (e.g. *chien*, *cane*, *perro*) and the semantic features attached to them (for the three forms above: “dog, a carnivorous mammal descended from wolves and domesticated by humans”). A lexeme is the combination of one form and one meaning.

These different domains are highly interdependent, and are in constant interaction in language processing. Although phonetics and phonology are only indirectly linked to syntax and lexical semantics, they provide the foundation for the latter domains: a limited repertoire of phonemes (less than forty in the various Romance languages) forms the basis for the expression of an infinite number of lexemes. These, in turn, convey different lexical meanings and the relationships between them (semantics), which makes it possible to use language as a means of communication (pragmatics).

The distinction between lexical semantics (linked to lexemes) and grammatical semantics (linked to function words and syntax) does not exclude multiple interactions between the two, for instance in the parts of speech, verb valency or, especially, in word formation (cf. 2.2.1 no. 2, 2.3.2 no. 1 and 5, 2.4.4). Moreover, lexical and grammatical semantics interact with each other to create textual meaning.

The suggested distinctions between the different ‘segments’ or ‘modules’ of language are neither absolute nor exclusive:

- the domain of word formation (derivational and compositional morphology) belongs in equal measure to morphology (as it concerns internal word structure) and lexis (in as far as it relates to lexical forms and their meaning);
- phonetics and phonology are treated in the same chapter, even though they involve two radically different and clearly distinct mechanisms of (language) processing in the brain;
- the domains of morphology and syntax are often grouped together and referred to as ‘grammar’.

The term ‘grammar’ must be considered with particular caution since it is used with multiple meanings and implications in linguistics. It may refer to the following concepts:

- in the sense in which it is used throughout this manual: all the features that constitute the internal structure of words (morphology) and their organisation within sentences (syntax); in other words, all morphological and syntactic features of a specific language, including all its varieties;
- in a more restrictive sense: all morphological and syntactic features of the standard variety of a language (i.e. its ‘normative’ grammar);
- with a focus on linguistics rather than language: the linguistic study of these features (i.e. the grammatical study of a language or of its standard variety);

- a work that aims to describe the most important of these features (a work of ‘grammatical analysis or description’, cf. 8.5)
- finally – and this is the most confusing use of the term – the study of grammar is often equated with the study of linguistics in general; ‘grammar’ thus becomes synonymous with ‘linguistics’ and even, by way of metonymy, with ‘language’.

The last of these meanings has proven to be a loaded term: numerous structuralists, generativists, functionalists and other erudite scholars (Guillaume, Tesnière, Chomsky, etc.) have considered, and still consider, aspects of grammar and – more specifically – of syntax to be the essential component of language. In some cases, it is seen as the most specific, vital and essential, or even the most stable part of language. This point of view, referred to as ‘syntactocentrism’, prioritises syntax over phonology, which possesses less semantic value, but also over lexis, owing to its apparently rather unsystematic character and its close connection with memorised concepts, which are considered to be ‘extralinguistic’.

Despite this strong focus on syntax, the domains of phonology, lexis and morphology are constitutive elements of language in their own right, and have equal epistemological validity (for more detailed argumentation on this subject, cf. Glessgen, *Le statut épistémologique du lexème*, 2011). In Romance studies, furthermore, a particularly rich tradition of research exists in each of these three domains. The Romance languages provide particularly fertile ground for the comparative study of the evolution of phonetic, phonological, morphological and lexical variation. A strong tradition of research in these domains had already formed by the mid-20th century when the modern study of syntax began to gather momentum.

Many linguistic models, however, do not conceptualise the domains of language as they are described here. The three-dimensional semiotic model established by Charles Morris in the 1930s (*Foundations of the theory of signs*, 1938) only acknowledges three elements: syntax, semantics and pragmatics. Generative linguistics continues along these same lines, and even now only distinguishes the three domains of phonology, syntax and semantics, as in the following definition:

‘We assume, putting aside the precise mechanisms, that a key component of FLN [*Faculty of language – narrow sense*] is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system’ (Hauser/Chomsky/Fitch, *The faculty of language*, 2002: 1571).

In such models, phonetics, morphology and lexis are not treated as entities in their own right. As for ‘semantics’, it is considered to be a domain outside of language proper, relating to general cognitive capacity. Models proposed in the field of ‘construction grammar’ take an opposite view: here, primacy is attributed to lexis, into which the various grammatical elements are incorporated. The segmentation of language domains thus raises fundamental problems in linguistics.

5.1.2 Contributions from neuro-psychology

Thought on the subdivision of language into structural domains may have something to gain from the consideration of studies on the functioning of the brain in a neuro-psychological (neuro- and psycholinguistic) framework. These are based on three types of observations: (i) the analysis of brain injuries, (ii) psycholinguistic recognition, production or reaction tests and (iii) the observation of neurological and physiological activity in the brain by means of electroencephalography (EEG) or magnetic resonance imaging (MRI; *cf.* 2.3.2 no. 3).

The starting point for these studies, from the end of the 20th century onwards, was the observation of phenomena associated with aphasia: these are serious linguistic deficiencies caused by physical or physiological injuries to the brain. Such injuries do not necessarily affect other types of cognitive competence possessed by the individual.

What is known as ‘Broca’s’ aphasia chiefly affects part of the left inferior frontal lobe (*Broca’s area*, located in centre no. 6, *cf.* fig. 22), and manifests itself in an inability to produce grammatically correct sentences (characteristics are deficient syntactic structures and the absence of grammatical markers, e.g. “boy break mother spanking”. Language comprehension, however, remains relatively unaffected.

‘Wernicke’s’ aphasia is linked to the superior left temporal lobe (*Wernicke’s area*, *cf.* fig. 22, no. 3). It affects lexical semantics and the semantic combination of words, leaving syntax unimpaired. Thus, linguistic production remains fluent and fundamental grammatical rules are respected, but sentences are meaningless; the faculties of comprehension and denomination are affected.

Considerable caution must be exercised regarding the interpretation of these two types of aphasia, and even more so concerning the roles of the areas of Broca and Wernicke. Nonetheless, these are genuine phenomena which point to the existence of a physiological substrate involved in the governing of grammar (inflectional morphology and syntax), and a second substrate involved in the processing of lexico-semantic information (which implicates phonological forms). The different forms of aphasia support the hypothesis of a modular form of language processing, which distinguishes between (i) grammar, (ii) phonological lexical forms and (iii) the concepts linked to them. They also emphasise the opposition (of a morphological nature) between verbs, nouns and function words, or between words and numbers.

Recent research has led to a much more nuanced view of the physiological differentiations within the brain due to the intense interaction between its different parts. The fact that there is some anatomical component in the functional processing of language remains uncontested. Currently, the most widely-recognised anatomic-functional model is that of Hickok and Poeppel (2007), which focuses on the comprehension and production of spoken language: these are actually two distinct processes, even though they take place in parallel with each other, rather than autonomously. The model distinguishes several zones which act as physiological substrates in different linguistic operations. These are mostly located in the left hemisphere of the brain. However, the

lexico-semantic centres also extend to the right hemisphere, as shown in our adaptation of this model:

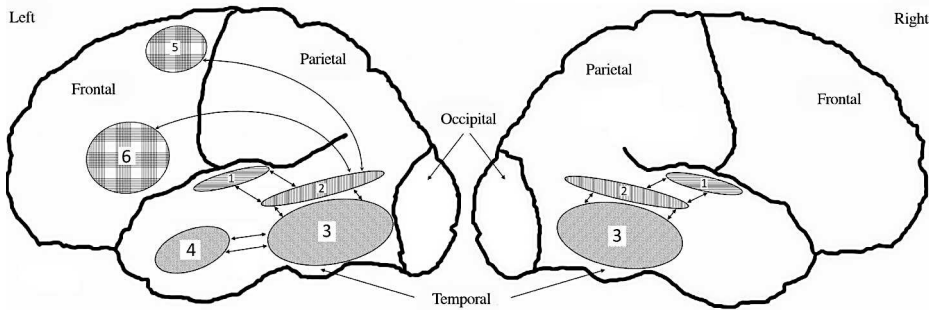


Fig. 22: The model of Hickok/Poeppel

This model of language processing in the brain distinguishes two ‘streams’: one for comprehension (‘ventral stream’), the other for speech production (‘dorsal stream’). It identifies centres responsible for the spectro-temporal analysis of phonetic signals (no. 1) and their association with phonological representations (no. 2). The posterior ventral stream – which is situated in both hemispheres – is used to decode acoustic signals in the representations of lexical forms and/or meaning (no. 3). The anterior ventral stream is involved in the combinatorial network - i.e. morphological and syntactic (no. 4). The dorsal stream is in charge of the articulatory network in speech production (nos. 5 and 6), and is also involved in comprehension.

With all due caution, previous as well as recent neuro-psychological research seems to indicate that language processing in the brain implies some degree of modularity. Everything leads us to suppose that it is necessary to distinguish different ‘modules’ for (i) phonetics and (ii) phonology, (iii) morphology and syntax (which are difficult to separate), as well as for (iv) lexical forms and (v) concepts (which are closely linked to each other).

During comprehension or production these modules are deployed simultaneously, and each one has a comparable role to play in language processing. In other words, they have equal status, they are equally valid and they are of equal importance for linguistics, even though their internal structure and function within the brain may diverge.

There is a fundamental difference between the domains of phonetics/phonology and morphology/syntax on the one hand, and those involving forms and concepts (lexico-semantic) on the other. The former belong to ‘procedural’ memory, which is completely inaccessible to the conscious mind and thus functions ‘automatically’. Forms and concepts, on the other hand, are registered in the semantic memory, which belongs to ‘declarative’ (or ‘explicit’) memory, located, as we have seen, in the median temporal lobe. Like procedural memory, declarative memory is long-term; however, it contains data that can, at least partially, reach the level of consciousness (thus, we can easily enumerate animal species, whilst we are unable to enumerate phonemes without any

prior knowledge of linguistics). Here, once again, it would be erroneous to suppose that lexis is any less a core component of language than phonology or syntax merely because it is partially accessible to consciousness – it is simply situated at a different level of organisation within the brain (cf. 9.1 for the linguistic status of lexis).

The differences between the various domains of language within the brain are only partially reflected by research traditions within the field of linguistics. For the following sections, therefore, we have established the following principles:

- phonetics (i) and phonology (ii) will be treated within the same chapter, as they are particularly difficult to separate from one another when analysed from a diachronic perspective (it is not easy to distinguish sounds and phonemes with regard to previous language stages that have only been transmitted in writing);
- morphology and syntax (iii), which nowadays belong to two quite distinct traditions of research, will be treated separately – the functional difference between the two domains has also been clearly established;
- lexical forms (iv) and meanings as well as concepts (v) will be dealt with together, as these three entities interact closely with each other (cf. 9.2.2 no. 6), and as they all belong to the tradition of lexicological research.

→ Raible, *Language universals and language typology*, LangTyp, art. 1
Dittmann, *Sprachpathologie*, LangTyp, art. 9

5.1.3 The domains of language and grammaticalisation

The functional interaction between the different domains of language immediately becomes apparent when the production of a concrete utterance is described. Using the terminology introduced and defined below, we can say that in the act of utterance, lexemes (= lexis) are assembled with the aid of function words, grammatical affixes and positional rules (= morphology) into syntagmatic groups, which are, in turn, interpreted as constituents forming clauses and sentences (= syntax). This arrangement permits various types of functional equivalence which shed light upon the mutual dependencies between the different domains of language:

- a lexeme can take on the role of a constituent without transformation or the addition of function words or grammatical affixes;
- a clause can function as a sentence in its own right;
- a sentence can function as a text (understood as an independent semantic and pragmatic entity);
- even a single lexeme, accompanied by a determiner, can constitute a sentence (known as a ‘holophrase’) or even a micro-text (e.g. *Un demi!* “a half pint” to order a beer; or *mon œil* “no way ...!”);

- function words can act as the equivalent of constituent positions or prosodic phenomena.

Since lexemes and syntagmatic groups only acquire their specific function within a sentence, there is continual interaction between the syntagmatic axis, which corresponds to the elements of the syntactic chain *in praesentia*, and the paradigmatic axis, which corresponds to the different formal and constitutive elements that can be substituted by others in order to change the meaning of the whole (cf. 1.2.4 no. 1).

Mutual dependence between the different domains or modules of language is particularly apparent when observing the phenomenon of ‘grammaticalisation’, a process which forges close links between lexis and the domains of morphology and syntax. A lexeme may be transformed into a function word as a result of grammaticalisation: e.g. the Old French lexeme *om* “human; male” (< Lat. *HOMO* “human”) is transformed into the pronoun *on* “on” (cf. 9.3.2 no. 2). A function word can undergo a further process known as re-grammaticalisation or ‘secondary grammaticalisation’ to become a more general function word; e.g. in the Romance languages, the Latin demonstrative pronoun *ILLE* “that one (over there)” is transformed into the definite article Fr. *le*, It. *il*, Sp. *el* etc. (cf. 8.2.3).

This type of reinterpretation implies that the forms undergo semantic weakening at the moment of introduction. The new forms may simultaneously fulfil a pragmatic function (e.g. the formation of the article generates an effect of individualisation: an object which is being referred to is emphasised). They may also reinforce the expressivity of discourse, as shown by the evolution from Latin to the Romance languages in cases such as adjectival comparison (*GRANDIOR* → *más grande*), adverbs ending in *-mente*, future and conditional forms (relying on modal expressions) as well as other cases of verbal periphrasis. The redeployment of previously-existing grammatical markers for the purpose of intensification can also be observed, for instance, in the use of prepositions to mark the roles of syntactic ‘actants’ (cf. 8.2.1). Numerous transformations between Latin and the Romance languages reflect this tendency towards expressivity as well as towards the transparency of grammatical information.

The semantic and/or pragmatic connotations of linguistic forms at the beginning of the process of grammaticalisation are subsequently lost and the element is ‘desemanticised’ when the process reaches its conclusion: when using a French verb form in the *passé composé* (*j’ai joué*) or, more particularly, the future tense (*je jouerai*), the speaker is no longer aware of the different lexical meanings of the verb *avoir* (“to possess”; “have to”; etc.).

The phenomenon of grammaticalisation has far-reaching interpretative potential. Furthermore, it highlights the relationship between lexis and grammar. Grammaticalisation means that grammar absorbs recurrent elements of lexis, thereby accelerating the brain’s management of language. At the same time, this implies that grammatical elements are in constant interaction with the lexemes on which they are based, and of which they allow the processing.

5.2 Periodisation of the Romance languages

5.2.1 Periods established for external and internal linguistic analysis

Before examining the different domains of language, it should be pointed out that problems relating to segmentation in linguistics also involve time. Language is always considered from a chronological perspective; even the language of today is determined by time in that it necessarily represents a given moment on a chronological axis. When observing language from a historical point of view, it is essential to distinguish moments of external rupture which may have had a real impact on the evolution of the Romance languages (take, for example, the fall of the Roman Empire) from internal changes, which play a structuring role in language configuration (such as the introduction of the grammatical article).

It is important to remember that the motives that determine an important linguistic change – such as those that result in the establishment of dialect frontiers or in the emergence of a new diastatic or diaphasic variety – always derive from external circumstances, even if their manifestation is always internal (*cf.* 1.2.6). Only three external factors are significant enough on a macroscopic level to be considered as ‘chronological boundaries’, or milestones, in the internal development of the Romance languages:

- the fall of the Western Roman Empire in 476 AD
- the rise of Western Europe in the 11th century
- the Renaissance and discoveries outside Europe around the 16th century

These far-reaching political and socio-cultural changes allow the establishment of four main ‘external’ periods generally recognised by historical linguists:

- the Roman period: before ca. 500 AD
- the High Middle Ages: ca. 500–ca. 1000 AD
- the Low Middle Ages: ca. 1000–ca. 1500 AD
- the Modern era: since ca. 1500 AD

We suggest a further subdivision of the last of these periods from ca. 1880 onwards, after the Industrial Revolution and the formation of the modern nation-states, which will be referred to here as ‘the present-day period’.

From a linguistic point of view, the Roman period is also the Latin period; the beginning of the High Middle Ages witnessed the birth of the *Romània*; the Low Middle Ages saw the systematic scriptural elaboration of numerous Romance languages; the Modern era corresponds to the age of standardisation, and the present-day era to that of generalised literacy. These aspects will be examined in greater detail in chapter 10.

The establishment of periods for ‘internal’ linguistic change is infinitely more complex. Firstly, each individual change needs to be considered as a case in its own

right before it can be determined whether or not it is related to other changes within the same domain – within phonology or lexis, for example –, and whether or not there is any interaction between the different domains of language. If significant changes in phonology, grammar and vocabulary can be ascertained for a given period, a more generalised transformation of language may be assumed to have taken place.

Throughout the linguistic history of the *România*, only one global internal transformation may be said to comply with the above criteria, namely the change from Latin to the Romance languages, which was accomplished between the 6th and the 8th centuries. A valid internal criterion for the completion of this transformation is the introduction of the definite article into the noun phrase around 700 AD (*cf.* the account provided by Chambon, *Sur la date de la fixation de l'article défini*, 2014). There is thus a clear interval between the external moment of rupture (ca. 500 AD) and the full manifestation of internal changes (ca. 700 AD). At the same time, the transformations responsible for the diversification of Latin into a large number of Romance varieties are spread over the whole of the first millennium; this strongly reduces the epistemological validity of any definable moment of rupture.

A second macroscopic transformation, less significant but nevertheless far-reaching, is that of the evolution of the medieval Romance languages into the modern Romance languages. The 16th century witnessed the parallel transition of Middle French to Modern French and Old Spanish to Modern Spanish (both ca. 1480–1630) as well as medieval Tuscan to Modern Italian (ca. 1500–1560). In each of these cases, changes took place in phonology, grammar and lexis (for the respective dates see Vachon, *Le changement linguistique au XVI^e siècle*, 2010, for French; for Spanish, Eberenz, *Castellano antiguo y español moderno*, 1991). The rapidity with which the changes took place in Italian can be explained by the fact that it was essentially a written and not a spoken language (*cf.* 10.5.3 no. 6)¹⁴.

¹⁴ As is the case for the delimitation of dialects or diastatic varieties, historical periods and the intervals between them are defined *a posteriori* on the basis of scientific reasoning; the perception of contemporary speakers, however, is always first and foremost that of the continuity of a language, rather than the transformations taking place within it. Nonetheless, the following citation from an essay of Montaigne's bears witness to his extraordinary consciousness of the ephemerality of the language of his generation, as well as of the timescale of its evolution: 'J'escris mon livre a peu d'hommes et peu d'années. Si c'eust esté une matière de durée, il l'eust fallu commettre a un language plus ferme. Selon la variation continuelle qui a suivy le nostre jusques a cette heure, qui peut espérer que sa forme présente soit en usage, d'icy a cinquante ans ? Il escoule tous les jours de nos mains et depuis que je vis s'est altéré de moitié.' (Montaigne, *Essais*, III, 9) ('I write my book for few men and for a few years. If it had been durable matter, it would have had to be committed to a more stable language [= Latin]. In view of the continual variation that has prevailed in ours up to now, who can hope that its present form will be in use fifty years from now? It slips out of our hands every day, and has halfway changed since I have been alive.' Transl. K. Chenoweth, *Montaigne on Language, The Oxford Handbook of Montaigne*, ed. Desan 2016, p. 377).

In a very general manner, the two major transformations in the linguistic history of the *România* mentioned above create a rough framework for internal periodisation, taking into consideration the slow rate of linguistic change. Based on this, we shall distinguish three extended periods, one Latin and two ‘Romance’, each of which can be further subdivided:

Latin Antiquity	
Archaic, Classical and post-Classical Latin	7 th century BC–4 th century AD (survives in written form throughout later periods)
Late Latin (regionalised in speech)	end of 4 th century–7 th century AD
The medieval România	
Romance languages (spoken; evolved from regionalised spoken Late Latin)	8 th –11 th century (Medieval Latin in written form)
Medieval Romance languages (written; in the process of elaboration)	12 th –15 th century (with intensification in writing during the 14 th century)
The modern and present-day România	
Transitional period	16 th century
Modern Romance languages	17 th century–ca. 1880
Present-day Romance languages	since ca. 1880

The internal periodisation suggested for the *România* may be further differentiated for each individual language, especially with regard to linguistic elaboration. The identification of different periods therefore chiefly depends on textual production, as illustrated below for French (each period being accompanied by the mention of several key texts):

Medieval French	
Pre-textual period	8 th century–ca. 1050, with only few complete texts: <i>Strasbourg Oaths</i> (842) <i>Canticle of Saint Eulalia</i> (end of 9 th century)
Old French	ca. 1050–middle of 14 th century: <i>Life of Saint Alexis</i> (1050), <i>Song of Roland</i> (~ 1100), the romances of Chrétien de Troyes, <i>Romance of the Rose</i> (1270), the tales of Reynard the fox
Middle French	mid 14 th century–ca. 1500: geographical diffusion of the linguistic norm established by the royal chancellery; <i>Ovide moralisé</i> , <i>Renart le contrefait</i> , <i>Guillaume le Maréchal</i> , François Villon, Charles d’Orléans
Modern and present-day French	
Pre-Classical French	ca. 1500–ca. 1630
Classical French	beginning/mid 17 th century–ca. 1880
Pre-Contemporary French	ca. 1880–ca. 1970
Contemporary French	since ca. 1970

Research on historical periodisation in the *Romània* is patchy; each Romance language has its own chronology and the different varieties of a language do not necessarily change at the same rate. The complex reality of the linguistic diasystem thus renders the establishment of clear-cut temporal delimitations impossible (cf. 10.5.3 no. 6).

5.2.2 The concept of Proto-Romance

The complementary concept of ‘Proto-Romance’ forms an essential part of a diachronic approach to the Romance languages for the period before AD 700, during which Latin was the only language used in writing. Even though most Romance lexical forms are attested in Latin, the basis of the early Romance languages is exclusively oral. Thus, a discrepancy exists between forms that have been transmitted in writing and oral forms to which we can only gain access by means of comparative reconstruction (cf. Chambon’s fundamental article, *Remarques sur la grammaire comparée-reconstruction en linguistique romane*, 2007 and the detailed presentation in the three volumes of the DERom). Consider the following example: the written Latin word GŪLAM (note that Romance lexemes are generally based on the Latin accusative) was pronounced /ˈgola/ as early as the 5th century, more or less as in modern Italian.

Reconstructed oral forms are referred to as ‘spoken Latin’ or ‘Proto-Romance’, depending on the author (the term ‘Vulgar Latin’, although widely used, should be avoided). The term ‘Proto-Romance’ also refers to forms dating from the Latin period that are not recorded in writing but that survive in the *Romània*. This includes simple lexemes, derivatives and grammatical innovations, as well as loanwords which entered Latin from contact languages such as Celtic. The following forms are examples of Proto-Romance:

- * /baʃa/ n.f. > Fr. *bave* “drool”: a simple onomatopoeic form, restricted to spoken language
- /mon-te/ → * /monˈtania/ n.f. > Fr. *montagne*, Sp. *montaña* etc. “mountain”: a derivative formed in late Latin, not documented in written texts
- * /li-ga/ n.f. > Fr. *lie* “dregs (e.g. of wine)”: a borrowing from Gaulish into Latin, also unrecorded in written texts

The existence of such forms in the Romance languages constitutes evidence of their prior existence in Latin; the application of the reconstructive method is not only justified, but indispensable. Proto-Romance, then, corresponds to spoken Latin, which is by definition reconstructed on the basis of the Romance languages (and our knowledge of written Latin). Any spoken Latin form that underwent the process of hereditary evolution towards one of the Romance languages could theoretically be labelled as ‘Proto-Romance’. For spoken Latin forms surviving in the *Romània* for which there is written documentation, however, it is conventional to speak of ‘Latin’. The designation ‘Proto-Romance’ should only be used to refer to the etyma (or ‘bases’) of Romance forms which are not documented in written Latin. This terminological distinction highlights

the coexistence of two separate linguistic entities within the Latin diasystem: Latin (written and spoken) and Proto-Romance (spoken).

With regard to historical periodisation, Proto-Romance cannot be defined chronologically; in other words it is neutral, diachronically speaking. Though the number of non-recorded Proto-Romance forms increased during the period of Late Latin (from the end of the 4th–7th century), they began to appear at least as early as the first century AD.

In our treatment of the internal history of the *Romània*, we shall only consider larger periods, in order to provide a general overview as well as a framework within which the individual linguistic changes can be structured chronologically.

→ Baum, *Periodisierung in der romanistischen Sprachgeschichtsschreibung*, RSG 1, art. 5

5.3 Remarks on the following chapters

Taking the ideas expressed above as a starting point, we shall present the fundamental domains of language in the following four chapters. These are phonetics and phonology (ch. 6), morphology (ch. 7), syntax (ch. 8), and lexis (ch. 9). Definitions will be provided for each domain, as well as essential terminology, and the most salient internal characteristics of each will be presented from a diachronic Romance perspective. In order to illustrate and explain the terminology in each case, we shall rely on examples from French and English, if there is no noteworthy intra-Romance variance.

We are also obliged to limit ourselves to a rough outline of the main developments. We will thus emphasise the major changes which took place between Latin and the Romance languages. At the same time, concrete transformations will be examined from functional, comparative and diachronic points of view. We shall further consider some elements of continuity that are characteristic of both Latin and the Romance languages, in contrast to other languages (e.g. the central role of verb inflection in grammar).

Regrettably, the innumerable linguistic forms and paradigms which make up the various Romance varieties, both medieval and modern, cannot be reproduced here. Our goal, then, remains that of familiarising readers with the basic issues, thereby enabling them to make optimal use of the major manuals and specialised bibliographies (cf. 2.4).

6 Phonetics, phonology and graphemics

6.1 Sounds, phonemes and sound change

6.1.1 Sounds, phonemes and allophones

Linguistics operates a clear distinction between speech sounds (or phones) and phonemes. Sounds correspond to the actual acoustic signals produced by speakers. They are created by means of our articulatory apparatus (i.e. principally the vocal cords, tongue and lips), and exhaled air. The resulting sound waves can be analysed and visualised as oscillograms and spectrograms, which reproduce the phonetic chain in graphic form, allowing the various sounds to be segmented. Hence, sounds strictly belong to physical reality.

Phonemes, in contrast, exist only at a mental level, and cannot be directly observed or pronounced. More specifically, phonemes correspond to the idea we have of the sounds we hear; they are the mental representation of sounds, a psychic phenomenon. In other words (concrete) sounds relate to *parole*, whereas (abstract) phonemes relate to *langue* (cf. 1.2.3).

What primarily distinguishes phonemes from speech sounds is that while sounds display high variability, phonemes are homogeneous. A specific phoneme in a given language (e.g. French) may be realised by a wide range of pronunciations, which vary according to speaker (male or female, adult or child), geographical origin (Paris, Marseille, Épinal, Québec, Tunis ...) or social background, or manner of speaking (e.g. rapid or slow enunciation). Hence, what all speakers interpret as one single phoneme is in reality represented by a series of sounds that may display a notable degree of acoustic (and thus physical) variation (cf. the different realisations of the phoneme /s/ in the example below, 6.3, fig. 26).

Phonemes are an abstraction created by our cognitive apparatus: at the level of the brain, we ‘hear’ a phoneme /a/ (conventionally noted between forward slashes) and not one of the many variant forms of the sound [a] (conventionally noted between square brackets). Humans are biologically predisposed to recognise similarities between different phenomena and to reduce an infinite variety of appearances to an acceptable number of entities. In the process of learning a given language, one learns to reduce potential distinctions between concrete sounds, so as to retain only those relevant to that language (i.e. sounds that fall within a whole range of sounds perceived as representing a particular phoneme). The brain goes so far as to correct accidental phonetic substitutions (e.g. /ləf tɔ̃ˈbe/ instead of /ləs tɔ̃ˈbe/ *laisse tomber* “let it go”); these are processes of which we are generally not even aware. The act of listening is always an interpretative act. Paradoxically, then, our perception of language is both unconscious and autoreflexive, as the linguistic data we receive are organised and structured at an unconscious level.

Phonemes serve to distinguish one word from another (i.e. they create a contrast between forms with different meanings in a language), thereby enabling the formation of words and sentences. This distinctive property is based on functional oppositions within a given language. If the commutation of two sounds in an identical sequence results in two words with different meanings, one can deduce that the two sounds represent distinct phonemes. Such word pairs are termed ‘minimal pairs’. In the following example, replacing [p] by [b] results in two distinct words [‘po] Fr. *peau* “skin”, and [‘bo] Fr. *beau* “handsome”; consequently, one can ascertain the existence of two separate phonemes /p/ and /b/ in French. In this respect, phonemes are the smallest distinctive units of a language that are responsible for the constitution of meaning, as illustrated by the following minimal pairs in French and Italian (cf. Pöckl / Rainer / Pöll 2003: 12):

	word-initial position	word-medial position	word-final position
/p/	pas /‘pa/	pépé /pe‘pe/	pompe /‘pɔ̃p/
/b/	bas /‘ba/	bébé /be‘be/	bombe /‘bɔ̃b/
/t/	thé /‘te/	gratin /gra‘tɛ̃/	honte /‘ɔ̃t/
/d/	dé /‘de/	gradin /gra‘dɛ̃/	onde /‘ɔ̃d/
/k/	quai /‘ke/	caca /ka‘ka/	bac /‘bak/
/g/	gai /‘ge/	gaga /ga‘ga/	bague /‘bag/
/s/	sel /‘sɛl/	casser /ka‘se/	fesse /‘fɛs/
/z/	zèle /‘zɛl/	caser /ka‘ze/	fez /‘fɛz/
/tʃ/	It. Cina /‘tʃina/	It. cacio /katʃo/	
/dʒ/	It. Gina /‘dʒina/	It. agio /aɖʒo/	

The minimal pairs listed above allow the French phonemes /p/ ~ /b/, /t/ ~ /d/, /k/ ~ /g/, /s/ ~ /z/ to be defined according to the distinctive feature of being either voiceless or voiced. The table also accounts for their position within the word.

In the case of the Italian phonemes /tʃ/ ~ /dʒ/, which are also distinguished by the opposition voiceless ~ voiced, the word-final position remains unoccupied, since the phonemes only occur in the initial and medial positions:

	word-initial position	word-medial position	word-final position
/tʃ/	Cina /‘tʃina/	cacio /katʃo/	
/dʒ/	Gina /‘dʒina/	agio /aɖʒo/	

In contrast, the Spanish sounds [m] and [ɱ] displayed in the following table do not occur in minimal pairs:

	word-initial position	word-medial position	word-final position
[m]	Sp. mal /'mal/	Sp. animal [ani'mal]	
[ŋ]		Sp. infierno [in'fjerno] Sp. con + [f] [komf(er'βor)]	

Instead, the two sounds occur in complementary distribution: [ŋ] is only found before the labiodental fricatives [f] and [v] (as in [in'fjerno]), while [m] (or [n]) occur in all other positions. Thus, the sounds [m] and [ŋ] do not represent two separate underlying phonemes, but are variant forms, so-called allophones of a single phoneme.

Allophones may be defined by their position in the word ('positional' or 'combinatory' variants) or by the diasystematic context in which they occur; these types of allophony differ from cases of 'free variation', which occurs independently of both syntagmatic and diasystematic context.

As an example of the first type of allophony (i.e. the positional variants), in Modern French the sound [k] is pronounced differently in *qui*, *cage*, *comme* and *cou*: depending on the quality of the following vowel, its articulation is more or less palatal or velar; these variant realisations are thus allophones of the phoneme /k/. In Spanish and Italian, the sound /ŋ/ occurs only before the velar consonants /k/ and /g/ (Sp. *rincón* "angle" [rin'kon], *ninguno* "nobody" [nin'guno]; It. *lingua* "tongue" ['lingwa]); [ŋ] is an allophone of the phoneme /n/. In contrast, the distinction between /n/ and /ŋ/ in English points to two separate phonemes in syllable-final position (compare *sin* /sin/ and *sing* /sin/). This shows that the distinction between phonemes and allophones depends on each individual language and its phonological system.

Allophones determined by a specific diasystematic context are independent of phonetic context. In Modern French and Italian, for example, the phoneme /r/ displays different variants in some regional or sociolectal varieties: it is pronounced as a uvular trill [ʀ] in varieties close to standard or familiar French throughout the former *langue d'oïl* territory, but as an alveolar [r] in familiar southern French, as well as by some speakers of *oïl* dialects and of French in Africa. An opposite tendency can be observed in some northern regional varieties of Italian, where the phoneme /r/ occurs in the form of an allophonic [R]. This type of allophony is often called 'free variation', which is misleading since actual free variation is not diasystematically bound.

The functional role of phonemes emphasises the degree to which languages rely on contrast as a principle of perception. The definition of the phoneme established by the Prague Linguistic Circle in 1939 provided a basis that would allow the development of more effective structural thought, facilitating the emergence of structural semantics and syntax in the following decades (cf. 2.2.4 no. 3).

6.1.2 Phonetics, phonology and sound change

Phonology, dedicated to the study of phonemes and their distribution in different languages, relies on observations from the field of phonetics. The main branches of phonetics focus on the production of sounds (articulatory phonetics), the description of acoustic waves corresponding to different sounds (acoustic phonetics) or the perception of sounds (auditory or perceptive phonetics). Syllables and sentence melody (prosody), on the other hand, are of more interest to phonology, which places greater importance on semantic issues.

The two disciplines, together with their respective subdisciplines, witnessed considerable empirical, technological and theoretical progress during the last decades of the 20th century. This development was particularly significant in the field of Romance studies, as phonological variation is extremely pronounced across the different Romance languages, especially if one includes dialectal varieties, each of which has its own phonological system. Furthermore, the Romance languages provide particularly fertile ground for the observation of phonetic and phonological variation and change.

The discovery of the regularity of sound change was one of the major milestones of linguistics during the 19th century. It was found that in many cases specific Latin sounds consistently resulted in specific corresponding sounds in the Romance languages (for example, Latin stressed /a/ in open syllables resulted in /e/ in French: CASA- > *chez*, -ATA > -ée). Such observations, which were first made with regard to Sanskrit (by William Jones in 1786, and later on by Franz Bopp and Rasmus Rask, who laid the foundations for comparative Indo-European philology), catalysed studies on the history of the German language (such as the pioneering work of Jacob Grimm) as well as the Romance languages (beginning with Friedrich Diez, who is traditionally seen as the founder of Romance philology). The school of German linguists known as the ‘neogrammarians’ (*Junggrammatiker*, cf. Hermann Paul, *Principien der Sprachgeschichte*, 1880) saw in these phonetic correspondences (natural) ‘laws’ that acted on language and to which there were no exceptions (cf. 2.2.3). Though this conception appears exaggerated today, scholars still speak of ‘laws’ of sound change, and the regularities that come to light when one compares the various Romance languages to Latin remain striking.

The development of phonology as a discipline created new theoretical foundations which serve as a basis for the understanding of the mechanisms of phonetic and phonological change (cf. 6.2.4). To begin with, it should be noted that these changes occur unconsciously, slowly and gradually. As we have shown, a change always first appears as a phenomenon of phonetic variation at the level of *parole* – i.e. as an articulatory variant of a previously existing phoneme. If this new variant becomes frequent and regular within a geographically or socially defined group of speakers, it may acquire the status of an allophonic realisation. More precisely, a free variant becomes prototyp-

ical either in a particular positional environment (e.g. in syllable-initial or intervocalic position) or within a specific diasystematic variety.

At this stage, the phoneme in question encompasses both the old and the new sound. Thus, allophonic variation in synchrony is the evidence of linguistic change in progress. The process of phonetic change is accomplished when speakers unconsciously interpret the new sound as being ‘normal’, thereby relegating the original sound to the position of an infrequent free variant, until it disappears completely. When the sound has been phonologised (i.e. established as a phoneme at the systemic level of *langue*), the transformation has reached its conclusion, and can no longer be reversed.

This somewhat simplified description provides a clear explanation as to why such processes cannot be fast-paced; even within a specific social group, different variants generally coexist for quite long periods of time. Studies focusing on the relative chronology of sound changes normally estimate that a time span of three generations or even an entire century is required for the completion of a phonetic change – i.e. the definitive transition from one sound to another (*cf.* the various articles by Elise Richter, *Chronologische Phonetik*, 1934 and Georges Straka [1979]). It is, however, more prudent to assume that such a time span is necessary for the process of phonologisation alone (*cf.* e.g. Michel Banniard, *Viva voce*, 1992), with additional longer or shorter periods of time both before the transition and after it, during which free variants exist.

Even more time is required in order for a sound change to become generalised within a large geographical area and across various social strata than for its normalisation within a specific group. This explains why, for example, regional varieties often retain older variants of a word e.g. [mʏe] for standard Fr. *moi* [mɔa].

The chronology of phonetic changes is difficult to determine owing to the fact that there is always interaction among different phonemes. The evolution of individual sounds overlaps, and different processes influence one another. This makes any attempt to define a strict chronological order unrealistic.

In the French tradition of diachronic linguistics, the theory of ‘relative chronology’ elaborated by Richter and Straka is often interpreted in terms of absolute and minutely detailed chronology (*cf.* e.g. Gaston Zink, *Phonétique historique du français*, 1986). This method has been heavily criticised, and with good reason (*cf.* Jean-Pierre Chambon, *Aspects de l’œuvre linguistique de Georges Straka: chronologie relative*, 1996–1997).

Given the complex nature of Romance phonological systems and the variation and change to which they are subject, we will restrict the following presentation to a series of basic descriptive elements, relating, in particular, to articulatory phonetics, and we will apply them to the four standard languages French, Italian, Spanish and Portuguese (*cf.* 6.2 below). We will then use the terminology introduced in the course of the presentation to provide a summary of the major phonetic evolutions that have taken place in

the Romance languages since Latin (*cf.* 6.5) and to present major changes in morphology (*cf.* 7.3.4; 8.3.3) and lexis (*cf.* 9).

- Baldi/Savoia, *The notion of the phoneme*, CambrHandb 8
- Loporcaro, *Phonological processes*, CambrHist 1, 3
- Schmid, *Segmental phonology*, OxfGuide 25
- Sampson, *Sandhi phenomena*, OxfGuide 44
- Wheeler/O'Neill, *Sandhi phenomena*, CambrHandb 6
- Bonet/Torres-Tamarit, *Typologically exceptional phenomena in Romance phonology*, CambrHandb 9
- Gsell, *Chronologie frühromanischer Sprachwandel*, LRL 2/1, art. 118

6.2 Sounds and phonemes in the Romance languages

6.2.1 Place of articulation

In order to avoid misunderstandings, it should be noted that sections 6.2.1 and 6.2.2 concern sounds (indicated between square brackets [...]), whereas section 6.2.3 concerns phonemes (indicated between forward slashes [...]). The production of speech sounds is dependent on the interaction of the vocal cords and the flow of air, as well as on the articulation of sounds by means of the tongue, lips and soft palate. In the Romance languages, speech sounds are always formed by simple expiration, which is the most usual mode of sound formation in most languages in the world. There are two main parameters that determine the quality of the sounds: their place of articulation and their manner of articulation.

Let us first consider the various places (or points) of articulation. When the tongue touches the hard palate, airflow is obstructed or impeded; this results in the production of different consonants, depending on the place of occlusion. When air passes freely, vowels are produced; their quality depends on the position of the tongue in the buccal cavity. Furthermore, voiced sounds (i.e. vowels and voiced consonants [b, d, g], including trills [r] and laterals [l]) are produced by vibration of the vocal cords, in contrast to voiceless consonants [p, t, k, s], which are produced without vibration of the vocal cords. The following diagram illustrates the precise places of articulation:

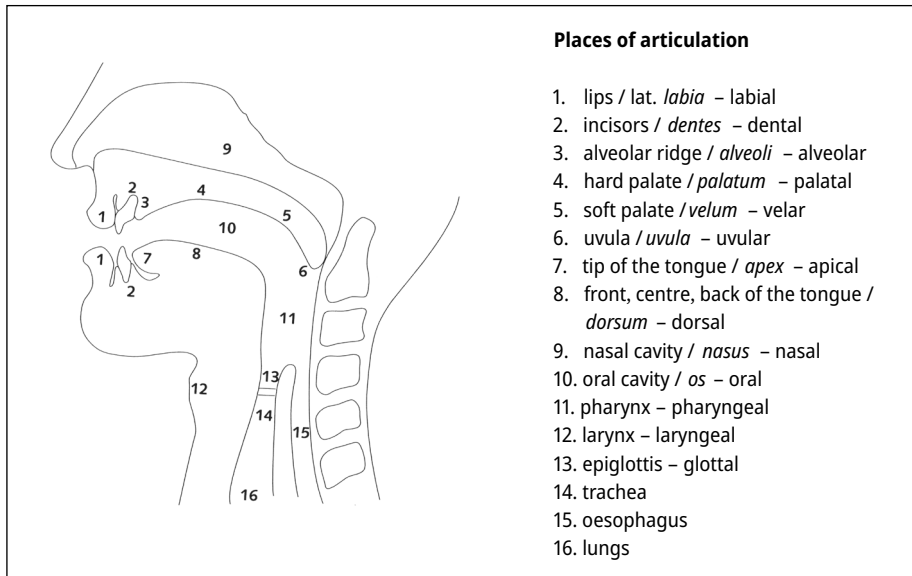


Fig. 23: Articulatory organs and places of articulation

Source: Mangold, *Artikulatorische Phonetik*, LRL 1/1, art. 23a, p. 616.

Remarks:

- The majority of places of articulation have little or no mobility: the upper lips, teeth, alveolar ridge, hard palate, soft palate, uvula and the wall of the pharynx are places of ‘passive’ articulation.
- In contrast, the lower lips and various points of the tongue are highly mobile; these are places of ‘active’ articulation.

The phonetic description of consonants accounts for the following places of articulation (terms are derived from the Latin denominations for the respective places of obstruction; *cf.* no. 2 below for examples):

- bilabial consonants: articulated with both lips [p b β m]
- labiodental consonants: articulated with the lower lip and upper teeth [f v m]
- dental or interdental consonants: articulated with the tongue touching the upper teeth [t d s z ts dz] or positioned between the teeth [θ ð]. The articulation of [t d s z ts dz] varies in different languages: it tends to be alveolar in German or English, whereas it is dental in most Romance languages; in Spanish, however, [s z] are alveolar. The distinction is thus not clear-cut.

- alveolar consonants: the tongue touches the alveolar ridge [n l ʎ], possibly vibrating [r r:]
- palato-alveolar or prepalatal consonants: between the alveolar ridge and the hard palate [ʃ ʒ ʧ ʤ]
- retroflex or cacuminal consonants: *id.* (palato-alveolar), with the tongue curved backward (Sicilian, Calabrian)
- palatal consonants: the tongue touches the hard palate [ɲ ʎ]
- velar consonants: the tongue touches the soft palate [k g x ɣ ŋ]
- uvular consonants: the uvula vibrates against the tongue [R]
- glottal consonants: produced by a narrowing of the glottal cavity (at the level of the vocal cords) Germ. [h]

Only three positions of the tongue are typically distinguished for the classification of vowels:

- front (or palatal) vowels: articulated with the tongue near the front part of the hard palate [i y e ε ø œ]
- back (or velar) vowels: articulated with the tongue near the soft palate [u o ɔ]
- central vowels: articulated with the tongue in an intermediary position [ɨ ə], possibly also [ɐ ɤ] (classified as back vowels by the IPA) and [a] (considered to be a front vowel by the IPA)

The three Romance glides (also known as ‘semivowels’) are classified according to the place of articulation of their consonantal component: [j] and [ɥ] (as in Fr. *huit* ‘eight’) as palatal glides, and [w] as a velar glide.

The terminology used for processes of phonetic change relies in part on the same designations:

- palatalisation: shift of the place of articulation of a consonant toward the hard palate (e.g. [a] > [e]; [k] > [tʃ] > [ʃ] in Lat. *CABALLU* > Fr. *cheval*)
- velarisation: shift of the place of articulation of a consonant toward the soft palate (e.g. [l] > [u] > [au] > [o] in Fr. *cheval* / *chevaux*)
- labialisation: shift of the place of articulation of a consonant or a vowel involving the position of the lips (e.g. [u] > [ɥ] in Lat. *MURU* > Fr. *mur* or [w] > [b] in Lat. *VĚRVĚCARIU* > Fr. *berger*)

There are several different systems of phonetic transcription for sounds and phonemes: the most widely known today is the alphabet of the International Phonetic Association (IPA), which is the system used here. There are other alphabets used specifically in Romance linguistics which have a longer tradition: these include the system used in the *Atlas Linguistique de la France* (ALF) for Gallo-Romance, that of the *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* (AIS) for Italian, and that of the *Atlas lingüístico de*

la Península ibérica (ALPI), which is inspired by the AIS (albeit with subtle differences), for Ibero-Romance. The main differences between these systems and that of the IPA are i) that the former use diacritics to indicate the degree of vowel openness (the ALF uses é for IPA e [= closed] and è for IPA ε [= open], while the AIS and ALPI use ɐ [= closed] and ɛ [= open]), and ii) that they use other symbols for the notation of fricatives and affricates.

The IPA, however, does not rely on diacritic symbols for ‘broad’ transcriptions, with the exception of the tilde for nasals (ẽ, ã, ã). Instead, each individual speech sound is noted by means of a letter or digraph (i.e. a two-letter sequence), in addition to a few symbols belonging to or inspired by the Latin and Greek alphabets [β γ ð ε θ ʎ χ]. Phonetic transcriptions may be broad and approximate, or narrow and, consequently, more precise; in the second case, the IPA introduces diacritic symbols for the additional specification of place and manner of articulation. Broad transcriptions will suffice for most examples provided in this manual.

→ Mangold, *Artikulatorische Phonetik*, LRL 1/1, art. 23a (includes an inventory of IPA symbols on p. 618, illustrated by examples from the Romance languages)

6.2.2 Manner of articulation

1 Romance consonants

The second main parameter that determines the quality of speech sounds is their manner of articulation; this parameter essentially refers to the manner in which the vocal organs obstruct or divert airflow in order to produce a sound in a given place of articulation (this may be a complete obstruction of expired air, as in Fr. *p(ère)*, *t(able)*, or simply the reduction of airflow without obstruction, as in *f(ou)* or *s(oleil)*. Here, too, the classification of consonantal sounds is more fine-tuned than that of vowels:

- plosives (also called stops or occlusives): airflow is blocked entirely, then released abruptly, as in the following examples:

	Fr. [of France]	It.	Sp. [of Spain]	Pg. [of Portugal]
[p]	<i>père</i>	<i>padre</i>	<i>padre</i>	<i>pai</i>
[b]	<i>bon</i>	<i>buono</i>	<i>bueno, vino</i>	<i>bon</i>
[k]	<i>cas</i>	<i>casa</i>	<i>casa</i>	<i>casa</i>
[g]	<i>gars</i>	<i>gatto</i>	<i>gato</i>	<i>gato</i>
[t]	<i>table</i>	<i>tavola</i>	<i>toro</i>	<i>touro</i>
[d]	<i>dame</i>	<i>donna</i>	<i>dama</i>	<i>dama</i>

- fricatives (or spirants): airflow is impeded (but not occluded) and channelled through a narrow passage, thus creating friction; fricatives include sibilants [s z ʃ ʒ]:

	Fr.	It.	Sp.	Pg.
[f]	<i>fou</i>	<i>foglio</i>	<i>fecha</i>	<i>fado</i>
[v]	<i>vous</i>	<i>voglio</i>		<i>vou</i>
[β]			<i>trabajo</i>	<i>abelha</i>
[θ]			<i>cena</i>	
[ð]			<i>cada</i>	<i>pode</i>
[s]	<i>soleil</i>	<i>sole</i>	<i>sol</i>	<i>sol</i>
[z]	<i>rose</i>	<i>sbaglio</i>	<i>misimo</i>	<i>casa</i>
[ʃ]	<i>chien</i>	<i>scena</i>		<i>chave</i>
[ʒ]	<i>jaguar</i>			<i>janela</i>
[x]			<i>ajo</i>	
[χ]			<i>hago</i>	<i>alguem</i>

To be more specific, the Spanish sounds exemplified above are not actual fricatives, but glides; in a narrow phonetic transcription, these are [β ð ɣ].

- affricates: these are formed by the release of a plosive into a fricative (or, more specifically in the Romance languages, a sibilant) at an identical place of articulation:

	Fr.	It.	Sp.	Pg.
[ts]		<i>tazza</i>		
[dz]		<i>mezzo</i>		
[tʃ]		<i>ciao</i>	<i>charla</i>	
[dʒ]		<i>gelato</i>		

- nasals: nasality occurs when the soft palate is lowered, allowing air to escape through the nasal cavity; it can be a complementary feature of any vowel (*cf.* Pg. below) or of some consonants that involve an occlusion of the vocal tract. Nasals become ‘oral(ised)’ when the soft palate is raised (e.g. [m] → [b]):

	Fr.	It.	Sp.	Pg.
[m]	<i>mal</i>	<i>male</i>	<i>mal</i>	<i>mau</i>
[ŋ]		<i>inferno</i>	<i>infierno</i>	<i>inferno</i>
[n]	<i>nez</i>	<i>naso</i>	<i>nariz</i>	<i>nariz</i>
[ɲ]	<i>signe</i>	<i>segno</i>	<i>baño</i>	<i>banho</i>
[ɲ]	<i>parking</i>	<i>anche</i>	<i>banco</i>	<i>banco</i>

- laterals and trills: laterals [l λ ʎ] are produced by means of a central blockage of air by the tongue: consequently, air escapes along the sides of the tongue. Trills involve an intermittent interruption of expiration created by a vibration of the tongue [r r:] or of the uvula [ʀ]. The general term ‘liquids’ is used to refer to laterals and trills, which form a category:

	Fr.	It.	Sp.	Pg.
[l]	<i>lune</i>	<i>luna</i>	<i>luna</i>	<i>lua</i>
[λ]		<i>figlio</i> [ˈfiʎʎo]	<i>calle</i>	<i>ilha</i>
[ʎ]				<i>tal</i>
[r]		<i>caro</i>	<i>caro</i>	<i>caro</i>
[r:]		<i>carro</i>	<i>carro</i>	
[ʀ]	<i>rouge</i>			<i>carro</i>

- glides: articulatory organs approach one another without resulting in actual obstruction. Glides always appear in combination with vowels, forming diphthongs. Hence, diphthongs consist of a vowel and a glide within a single syllable, resulting in a change in sound quality during production. It is necessary to distinguish diphthongs from the phenomenon of ‘hiatus’, which refers to two consecutive vowels that belong to two separate syllables (within an individual word, one may speak of ‘internal hiatus’ or ‘diaeresis’: e.g. Fr. *réunion* “meeting” [Reyn]ø), which involves a hiatus followed by a diphthong; Sp. *Europa* “Europe” when articulated slowly (e.g. in poetic diction) as [europa], with a distinct pronunciation of [e] and [u]; Rom.n *aude* “he hears” [aʊde]. The Romance glides are exemplified below:

	Fr.	It.	Sp.	Pg.
[j]	<i>pied</i>	<i>piede</i>	<i>pie</i>	<i>piedade</i>
[ɥ]	<i>noir</i>	<i>buono</i>	<i>bueno</i>	<i>ceu</i>
[ɥ]	<i>nuît</i>			
[ɛj]				<i>tem</i>
[ãw]				<i>canção</i>
[õj]				<i>canções</i>

Two glides combined with a vowel result in a so-called triphthong; cf. e.g. It. *miei* “mine” [ˈmjei], *tuoi* “yours” [ˈtwɔj], Rom.n *eu* “I” [ˈjew] or O.Fr. *lieit* “bed” [ˈljeɪt].

Other classification parameters for consonants are the previously mentioned distinction between voiced- and voicelessness (with or without vibration of the vocal folds, cf. [b] vs. [p]), and the part of the tongue involved in the production of a sound: its place

of articulation may be ‘apical’ (articulated with the tip of the tongue, *cf.* Lat. *apex*) or ‘dorsal’ (produced using the body of the tongue, *cf.* Lat. *dorsum*).

2 Romance vowels

Three parameters are distinguished for vowel articulation:

- vowel height or openness, referring to the position of the tongue with regard to the roof of the mouth – from high (near the roof of the mouth, producing open vowels) to low (at a distance from the roof of the mouth, producing closed vowels)
- roundedness, referring to the shape of the lips, which can be either rounded or unrounded
- position of the velum, inducing an oral or nasal sound.

Whereas the action of the lips and velum can be described using binary pairs (+ rounded vs. - rounded, oral vs. nasal), up to six positions of vowel height or openness are distinguished.

In vowel description, these three elements are combined with the parameter of ‘backness’ (referring to the position of the tongue with regard to the back of the mouth, which produces ‘front’, ‘back’ and ‘central’ vowels), whereby it is recognised that there is no inherent hierarchy among the various distinctive features; these may therefore be cited in any order.

A basic inventory of Romance oral vowels is provided below. They are categorised according to three degrees of backness (the parameter of roundedness is only decisive for front vowels in the languages exemplified), in order from the most closed to the most open:

- unrounded front vowels:

	Fr.	It.	Sp.	Pg.
[i]	<i>fil</i> s	<i>figlio</i>	<i>hijo</i>	<i>filho</i>
[e]	<i>né</i>	<i>sera</i>	<i>cena</i>	<i>ele</i>
[ɛ]	<i>chaise</i>	<i>bene</i>		<i>ela</i>
[a]	<i>chat</i>	<i>gatto</i>	<i>gato</i>	<i>gato</i>

- rounded front vowels:

	Fr.	It.	Sp.	Pg.
[y]	<i>lune</i>			
[ø]	<i>peu</i>			
[oe]	<i>peur</i>			

– central vowels:

	Fr.	It.	Sp.	Pg.
[ə]	<i>le</i>			<i>verdade</i>
[e]				<i>porta</i>

– back vowels (rounded):

	Fr.	It.	Sp.	Pg.
[u]	<i>vous</i>	<i>luna</i>	<i>luna</i>	<i>tudo</i>
[o]	<i>beau</i>	<i>sole</i>	<i>todo</i>	<i>porto</i>
[ɔ]	<i>porte</i>	<i>cosa</i>		<i>porta</i>
[ɑ]	<i>pâte</i>			

The Romance diphthongs will not be listed here. The majority of the examples given display ‘rising’ diphthongs (with the vowel, of greater volume than the glide, in second place: *pied* [pje], *buono* [bʊno], etc.), as opposed to ‘falling’ diphthongs (with the vowel in first place); e.g. It. *voi* [vɔj], *Paolo* [paʎlo].

The only modern standard Romance languages that have nasal vowels are French and Portuguese. They are presented in the following table, in the same order as those above:

	Fr.	Pg.
[ɪ̃]		<i>sim, vim</i>
[ễ]		<i>entra, somente</i>
[ɛ̃]	<i>chemin, saint, rien</i>	
[œ̃]	<i>lundi, un, brun</i>	
[ɐ̃]		<i>canto, mando</i>
[ũ]		<i>fundo, um</i>
[ɔ̃]		<i>ponto, bom</i>
[õ]	<i>pont, dont, font, son</i>	
[ã]	<i>langue, penser</i>	

Portuguese additionally has three nasalised diphthongs: [ãw] (*canção* “song”), [õj] (*canções* pl. “songs”) and [ẽj] (*vem* “he comes”; *tem* “he holds” or “he has”).

Among the distinctive elements of vowels, backness (place of articulation) and height (openness) – when considered from an abstract point of view – can be represented as a cross-section of the buccal cavity; this representation is often projected onto a ‘vowel trapezium’ diagram. Modern French is exemplified in the following figure:

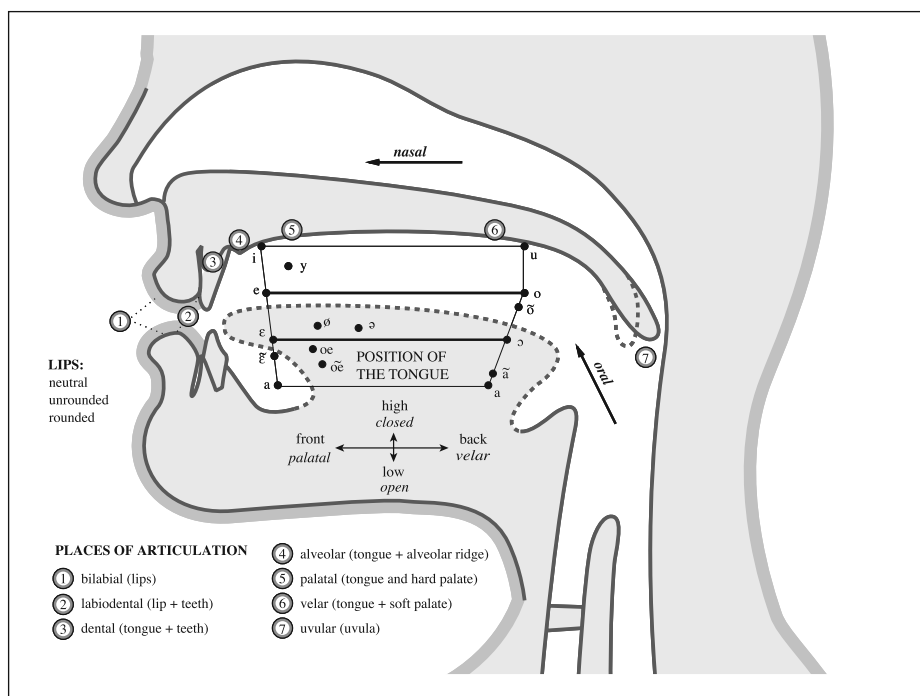


Fig. 24: The 'vowel trapezium'

Source: translated from A. Stein, *Einführung in die französische Sprachwissenschaft*, 32010.

6.2.3 Phonological systems in the *România*

The parameters used to describe phonemes are 'distinctive features', which, at the same time, allow variable classifications. They are typically arranged in rectangular diagrams, the horizontal axis of which is based on the place of articulation; in consonant tables, the vertical axis represents an increasing degree of sonority (in the following order: plosives, fricatives, nasals, liquids, glides – vowels being the speech sounds with the highest degree of sonority).

The sounds presented up to this point are organised in different ways within each individual phonological system: phonological oppositions vary according to the basic phonetic inventory of a language. Furthermore, sounds that have the status of phonemes in one language may be allophones in another.

The following brief description of the phonological systems of the four Romance languages French, Italian, Spanish and Portuguese does not include diphthongs or triphthongs, nor does it include detailed discussions concerning the precise description of different phonemes, which is far from straightforward (cf. Schmid, 2002 [ms.]).

1 French (of France)

- *vowel system*: 16 phonemes

		front	middle	back
	rounded	- +		+
closed		i y		u
		e ø		o
		ɛ ẽ œ œ̃	ə	ɔ ɔ̃
open		a		ɑ ɶ

The vowel repertoire is relatively variable among speakers of Modern French. A number of phonological oppositions are undergoing a general process of neutralisation, disappearing in specific positions:

- /e/ ~ /ɛ/, especially in the distinction between future and conditional / past imperfect forms: *j'irai* ~ *j'irais*
- /o/ ~ /ɔ/, e.g. *saule* “willow” ~ *sol* “ground, soil” (the opposition is absent in southern regional French)
- /ø/ ~ /œ/, e.g. *jeûne* n. “fast” ~ *jeune* “young” (absent in southern regional French)
- /a/ ~ /ɑ/, only present in a limited number of words and now reduced, e.g. *tache* “stain” ~ *tâche* “task” (absent in southern regional French)
- /ẽ/ ~ /œ̃/, present in even fewer words and also reduced in current usage, e.g. *brin* “wisp, sprig” ~ *brun* “brown” (maintained in southern regional French, in contrast to the above)
- *consonant system*: 21 phonemes

	bilabial	labio-dental	dental	alveolar	palato-alveolar	palatal	velar	uvular
voiced	- +	- +	- +	- +	- +	- +	- +	- +
plosives	p b		t d				k g	
fricatives		f v	s z		ʃ ʒ			
affricates								
nasals	m			n		ɲ	(ŋ)	
laterals				l				
trills								ʀ
glides						j ɥ	ɣ	

/ŋ/ only appears in borrowed forms.

2 Italian

- *vowel system*: 7 phonemes

		front	back
	rounded	-	+
closed		i	u
		e	o
		ɛ	ɔ
open		a	

The repertoire of unstressed vowels is restricted to the five phonemes /i e a o u/; moreover, in numerous regional varieties of Italian, the distinction between stressed /e/ and /ɛ/ (*vénti* “twenty” ~ *vènti* “winds”), as well as /o/ and /ɔ/ (*bótte* “barrel” ~ *bòtte* “blows” (n pl.)) is neutralised, thus also leading to a system of five phonemes.

- *consonant system*: 23 phonemes

	bilabial	labiodental	dental	alveolar	palato-alveolar	palatal	velar
voiced	- +	- +	- +	- +	- +	- +	- +
plosives	p b		t d				k g
fricatives		f v	s z		ʃ		
affricates			ts dz		tʃ dʒ		
nasals	m			n		ɲ	
laterals				l		ʎ	
trills				r			
glides						j	ɣ

allophones: [m ~ m̠], [n ~ n̠]

3 Spanish (of Spain)

- *vowel system*: 5 phonemes

		front	back
	rounded	-	+
closed		i	u
		e	o
open		a	

In word-final (unstressed) position, the two ‘extreme’ vowels /i/ and /u/ are reduced to the more central phonemes /e/ and /o/; hence, the only vowels that occur in word-final position are /e a o/.

- *consonant system*: 21 phonemes + 6 allophones

	bilabial	labiodental	dental	alveolar	palato-alveolar	palatal	velar
voiced	- +	- +	- +	- +	- +	- +	- +
plosives	p b		t d				k g
fricatives		f	θ	s	(ʃ)		x
affricates				tʃ			
nasals	m			n		ɲ	
laterals				l		ʎ	
trills				r r:			
glides						j	ɣ

allophones: [b ~ β], [m ~ m̃], [d ~ ð], [n ~ ɲ], [s ~ z], [g ~ ɣ]
/ʃ/ only appears in borrowed forms

4 Portuguese (of Portugal)

- *vowel system*: 8 phonemes + 9 allophones

The distribution of stressed oral vowels is identical to that found in standard Italian, with the minor difference that (European) Portuguese additionally has the phoneme /ɐ/. The unstressed vowel system, however, is quite complex; in particular, it displays reduction of the front vowels /e ε/ and /a/ towards the central vowels /i/ et /ɐ/, which results in a repertoire of four unstressed phonemes or allophones [i i̯ ɐ u].

The allophone system of Portuguese is made up of the nasal realisations of the seven stressed vowels and the two unstressed central phonemes (e.g. *inferno* [i̯fɛɾnu]).

- *consonant system*: 21 phonemes + 6 allophones

	bilabial	labio-dental	dental	alveolar	palato-alveolar	palatal	velar	uvular
voiced	- +	- +	- +	- +	- +	- +	- +	- +
plosives	p b		t d				k g	
fricatives		f v	s z		ʃ ʒ			
affricates								

	bilabial	labio-dental	dental	alveolar	palato-alveolar	palatal	velar	uvular
nasals	m			n		ɲ		
laterals				l		ʎ		
trills				r				R
glides						j	ɣ	

allophones: [b ~ β], [m ~ m̥], [d ~ ð], [n ~ ɲ], [l ~ ʎ], [g ~ ɣ]

6.2.4 Phonological variation and change

The speech sound inventory of all present-day Romance languages and varieties clearly exceeds the elements presented here (for a larger inventory, *cf.* Mangold, LRL 1/1, art. 23a); variation in phonological systems is crucial, since a single difference induces a restructuring of the entire system (*cf.* Martinet, *Économie des changements phonétiques: traité de phonologie diachronique*, 1970 [1955]). The phonological systems of the four standard languages described above have already shown how quite similar phonetic repertoires can give rise to diversification: Portuguese combines 16 vocalic and 27 consonantal allophones (43 in total); French, 16 and 20, respectively (= 36); Italian, 7 and 25 (= 32); Spanish, 5 and 26 (= 31), in distributional patterns that clearly differ from one language to the other.

A historical study – or even a monograph – aiming to provide an exhaustive account of the phenomena of phonological variation across the Romance languages and dialects would be unlikely to be feasible, owing to the large volume of data involved. Nonetheless, studies in phonology, especially those focusing on allophone variation, contribute to an understanding of the evolutionary processes involved. Every phonological change begins at the level of phonetic variation:

1. first, changes in sound production lead to instability: formerly distinct sounds become confused; new sounds are introduced into the language.
2. in a second step, phonetic variation leads to the introduction of new allophones. The new phoneme thus consists of two allophones: e.g. at a particular moment during the change from Latin *CANE* ['kane] to O.Fr. *chien* [tʃjɛn], [k] and [tʃ] became allophones of the phoneme /k/ preceding the vowel /a/;
3. finally, one of these allophones is eventually lost; in the example given above, [k] disappeared, and the allophone [tʃ] became the new phoneme /tʃ/ (*cf.* the examples provided in 6.5).

To describe phonological transformations in synchrony, linguists use specific symbols, which can also be used to refer to diachronic change, although they rarely are. Consider the following three examples:

- allophonic spirantisation (i.e. the formation of a fricative) in Spanish can be described as follows:

/d/ → [ð]/V _ (i.e. /d/ is realised by its allophone [ð] when preceded by a vowel), e.g. /kada/ → ['kaða];

- the conditions of positional schwa-dropping in present-day French can be represented thus:

/ə/ → Ø/VC _ CV (the phoneme /ə/ is dropped when preceded by a vowel-consonant sequence and followed by a consonant-vowel sequence [this means that the schwa is not dropped in cases where its deletion would result in a sequence of three consecutive consonants]), e.g. the phonological sequence /la fənɛtR/ is phonetically realised as [la fnɛtR];

- in diachrony, the aforementioned palatalisation of /k/ can be transcribed as follows: /k/ → /tʃ/ /\$_a (in other words, /k/ is palatalised as /tʃ/ in syllable-initial position when it is followed by /a/; the symbols # and \$ are used to refer to the phonetic boundaries of words and syllables, respectively), e.g. ['kane] → [tʃɛn].

For the sake of clarity and simplicity, phonetic changes will be presented in a more succinct and traditional manner in this section. According to our system, the third example mentioned above would be displayed thus: k(a) > tʃ(a) or [k(a)] > [tʃ(a)] (cf. 6.5.4 no. 2).

The terms used to refer to processes of phonetic change are based on the description of phonological mechanisms in general, the most multifaceted of these processes being assimilation:

- definition: a sound is partially or completely assimilated to a neighbouring sound; e.g. in synchrony, It. /'anke/ → ['aŋke] (partial assimilation), in diachrony, FACTU “deed” > It. *fatto* (total assimilation: /k/ is completely assimilated to /t/);
- basic types: the previous two examples display regressive assimilation (i.e. the second sound influences the first); progressive assimilation is less frequent (e.g. MUNDU > Central-South It. *monno/munnu*); even less common is reciprocal assimilation, where both sounds are transformed (e.g. CLAMO > Sp. *llamo* [ʎ]);
- there are also cases of long-distance assimilation, affecting vowels in particular (phenomena of metaphony: e.g. FECI > *fici* > Fr. *fis*; South It. *quistu*, -i vs. *questa*, -e: a word-final high vowel causes metaphony of the stressed vowel);
- depending on the place of articulation, assimilation may involve velarisation (cf. /'anke/ above), palatalisation (in synchrony, Fr. *qui* [ki] vs. *cou* [ku]; in diachrony, CENTU > It. [tʃɛnto]) or labialisation (e.g. DEBERE > It. *dovere*);
- still with regard to place of articulation, it may involve the following phenomena:
 - nasalisation (where a nasal consonant influences a preceding vowel, e.g. in synchrony, in O.Fr. *bon* ['bõn], *tems* ['tẽns] ~ ['tãns], in diachrony, in Modern French *bon*, *temps* or in Pg. *inferno* [ĩfɛrnu]);

- voicing (especially of consonants influenced by neighbouring vowels, e.g. in synchrony, Sp. /mismo/ → ['mizmo], in diachrony STRATA > It. *strada*);
- spirantisation (by means of which already voiced plosive sounds are transformed into fricatives, e.g. in diachrony, AMICA > Sp. *amiga* [a'miya], PIPERE- > Fr. *poivre*).

Among further processes of sound change that come into play, the following are particularly characteristic of diachronic change in the Romance languages:

- dissimilation (or differentiation): differentiation of a sound from another; e.g. in diachrony, PEREGRINU > Fr. *pélerin* ([r] > [l]), AN(I)MA > Sp. *alma* ([n] > [l]);
- metathesis (or inversion): displacement of a sound (especially a lateral or trill); e.g. in diachrony FORMATICU > Fr. *fromage*, MIRAC(U)LU > Sp. *milagro*;
- syncope: loss of a (generally unstressed) word-medial vowel (cf. the examples of *alma* and *milagro*);
- apocope: loss of a sound in word-final position (e.g. in synchrony, It. *cantan* for *cantano*);
- apheresis: loss of a sound in word-initial position (e.g. in diachrony, Gr. *apothēke* > Fr. *boutique*);
- epenthesis: insertion of a transitional sound (e.g. in diachrony, HUM(I)LE > Fr. *humble*, STELLA > Sp. *estrella* owing to interference with Lat. ASTRU).

Since phonological systems are restructured as soon as a new phoneme appears, it is difficult to pinpoint the exact moment in history at which specific case of phonologisation (i.e. the process by means of which a phonetic variant acquires a phonemic status) took place. Hence, it is not possible to present phonological changes *stricto sensu*, but only cases of phonetic change; when discussing the latter, it is nevertheless necessary to refer to 'archetypal' sounds and to gloss over much of the intense variation that it involves in reality.

Methodologically speaking, elements of articulatory phonetics are indispensable in the field of linguistics. Current research is nevertheless more active in auditory or perceptive phonetics, which studies the realisation of sounds for specific purposes (e.g. in the framework of variational linguistics), or which focuses, for example, on phenomena of anticipation (processes by which we deduce the continuation of what we are hearing before actually having heard it).

→ OxfEnc, sect. *Phonological variation and change (in European French, Armstrong; in Italian, Vietti; in Brazilian Portuguese, Battisti; in Romanian, Chitoran)*

RSG 3, sect. XV, *Histoire interne: systèmes phonique et graphique (roumain, Beltechi, art. 222; Bündnerromanisch, Eichenhofer, art. 229a; italiano, Biffi/Maraschio, art. 230; français, Morin, art. 235; Katalanisch, Meisenburg, art. 241; español, Cano Aguilar, art. 245; Portugiesisch, H. Lüdtkke, art. 251; Phénomènes de convergence et de divergence dans la Romania, Geisler, art. 256)*

6.3 Syllables and prosody

‘Prosody’ does not refer to the study sounds or phonemes in isolation, but of words and sentences as ‘suprasegmental’ units (involving several segments, i.e. several phonemes). Consequently, this field of research is also known to as ‘suprasegmental phonology’. There are close links between phenomena at the level of words and sentences (e.g. stress) and the qualities inherent to specific phonemes; in research, however, the two aspects are generally studied separately.

The central unit of observation in prosody is the syllable. In the Romance languages, utterances are generally composed of syllables which all have approximately the same length (‘syllabic isochrony’), in contrast, for example, to word length. Syllables are composed of phonemes arranged in a specific phonotactic structure: syllable theory distinguishes the ‘onset’ of the syllable from its ‘rhyme’, and, within the latter, a ‘nucleus’ and a ‘coda’ (cf. the diagram in fig. 25 below, which illustrates the phonotactic structure of the Italian word *porta* /pɔrta/ “door”).

With regard to the structure of syllabic constituents, the languages of the world – and the Romance languages in particular – tend to prefer open syllables (or ‘free’ syllables, i.e. ending in vowels). The ‘ideal’ syllable displays a consonant - vowel structure (CV), exemplified by the second syllable in the diagram below.

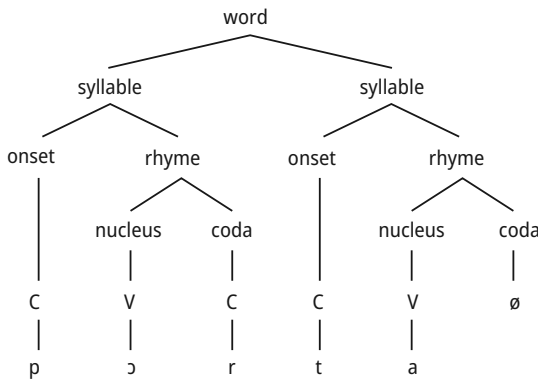


Fig. 25: The phonotactic structure of It. *porta* (source: Schmid, *Einführung in die allgemeine Phonetik*, 2002 (ms.))

Continuing the discussion of universal features, phoneme sonority increases toward the nucleus of the syllable, as in the first syllable of *porta*.

The syllable determines fundamental characteristics of prosody such as vowel quantity and stress (in which vowel length, in turn, comes into play). Whether a Latin syllable is open (e.g. CA-SA) or closed often determines the course of evolution of stressed vowels (cf. 6.5.2).

A syllable is considered to be closed when it ends in a consonant (e.g. BES-TIA, HER-BA); this occurs when two consonants follow one another (the first belonging to the preceding syllable, the second to the following syllable). Consonant clusters composed of a variable consonant followed by a liquid (*l*, *r*, e.g. DUPLEX, MATREM) are an exception; these combinations of consonants, known as *muta cum liquida*, belong entirely to the second syllable; the first syllable thus remains open (hence MA-TREM, DU-PLEX).

More precisely, prosody results from the two elements of vowel length and stress at the level of the syllable, in combination with intonation, which refers to pitch variation throughout the utterance.

It is useful to consider these elements in more detail:

- pitch (in the musical sense) results from the vibration of the vocal cords in the larynx. The individual's physical constitution conditions the average pitch of his or her speech; this so-called 'fundamental' frequency of the individual's speech is essentially linked to the physical properties of his or her vocal cords, which are thinner in women and generally of variable length; this is the only physical difference pertinent to phonetics. Moreover, all human beings are capable of producing the same diversity of speech sounds. Intonation measures variations in pitch with regard to the individual's fundamental frequency;
- intonation thus refers to the change of pitch of sounds in words and sentences (the change in pitch is, in turn, generated by pauses and, in particular, by its independence of semantic and syntactic features). It may be falling (often in final position), rising, or even;
- vowel length is the relative duration of a syllable in an utterance; the absolute duration is not phonologically pertinent since it varies according to pace of speech. Vowel length is of varying importance in different languages. Consider the following examples:
 - in Latin, vowel length is a distinctive phonological feature, cf. the difference between *MĀLUS* "bad" and *MĀLUS* "apple tree", *SILVā* "forest" (nominative case) and *SILVĀ id.* (ablative case), *VĒNIT* "he comes" (present tense) ~ *VĒNIT* "he came" (perfect tense);
 - in present-day French, the distinction tends to be lost, affecting the phonemes /ɛ/ and /a/, cf. *mettre* /mɛtR/ ~ *mātre* /mɛ:tR/ (cf. 6.5.1); *patte* /pat/ ~ *pâte* /pa:t/. These distinctions are absent in southern regional French, as is the opposition between the phonemes /ɔ/ and /o/, cf. *cote* /kɔt/ "classification mark, serial number" ~ *côte* /kɔt/ "rib, chop"; in both cases, the vowel is pronounced as a semi-open variant of /ɔ/;
 - in Italian, vowel length has a purely allophonic value: stressed vowels in open syllables are lengthened, as in *cane* /kane/ → [ˈkaːne], *sole* /sole/ → [ˈsoːle]; as a general rule, stressed vowels are longer than unstressed vowels;
- stress is a complex factor involving three elements, each in a different manner:
 - sound pitch
 - vowel length
 - intensity of phonic volume (known as 'dynamic' accent)

Stress places emphasis on syllables in words or sentences ('word stress' vs. 'sentence stress'). It is of capital importance with regard to the properties of sounds and the identification of the boundaries between words, as well as for sentence structure, especially since the phenomenon of stress is perceived as such by speakers (in contrast to its individual components, which are not). Stress fulfills different functions in different languages: whereas in Latin, it has a defined position in every word (*cf.* below), in French, it always falls on the final syllable of a prosodic group.

One of the challenging issues of prosody resides in the fact that prosodic segments are linked to complex – even contradictory – hierarchies: syllable pitch determines intonation and is simultaneously involved in the production of stress; vowel length is an element in its own right, but which nevertheless also plays a role in stress distribution; rhythm is even more complex as it emerges from the interaction of intonation, stress and speech pauses.

Consequently, systems of prosodic transcription are used far less often than systems of phonetic and phonological transcription. As a general rule, prosodic transcriptions must indicate maximum and minimum pitch and refer to the syllable as a starting point; the latter needs to be described in terms of its pitch and intensity (both of which are difficult to separate from duration); this allows the identification of the distribution of stress in words and sentences as well as patterns of intonation; even rhythm can thus be described.

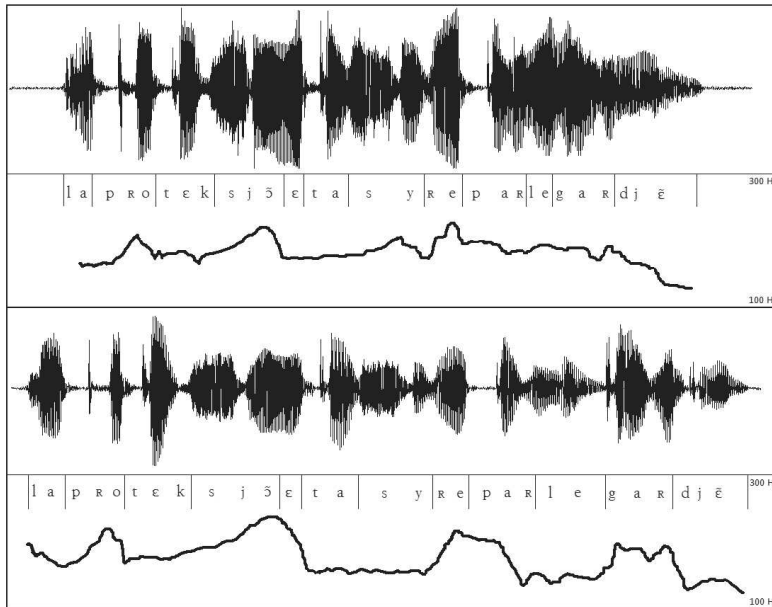


Fig. 26: Visualisation of acoustic waves (oscillogram and prosodic curves)

Source: after Dubois: French sentence '*La protection est assurée par les gardiens*'. The recording was analysed with the computer program Praat, which is a standard application today, as it is accurate, user-friendly (requiring relatively little learning time) and can be downloaded free of charge.

Remarks:

The visualised acoustic patterns are based on the recordings of two female French speakers, approximately 25 years of age (their accents being diatopically neutral in one case and slightly Jurassian in the other):

- the prosodic curves display slight differences in speech pace, including, in particular, the prolonged pronunciation of [le] by the Jurassian speaker;
- in addition, the second curve is slightly more marked in terms of rhythmic stress (*-sion* and *-rée*, an initial accent on *pro-*, and, furthermore, emphatic stress on *gardien*, which is dropped, as the voice is lowered at the end of the sentence);
- idiosyncratic and situational differences can be detected; at the same time, the linguistic differentiation between stressed and unstressed syllables is perfectly recognisable: the brain is able to reconstitute such elements without hesitation through an understanding of opposition;
- in particular, the oscillogram allows one to recognise and distinguish the main types of phonemes, listed below in order of increasing sonority:
 - voiceless stops [p t k], etc. marked by a near-total absence of sonority
 - liquids [r l] and semivowels (recognisable)
 - fricatives (s with consistent, prolonged sonority)
 - oral and nasal vowels ([a o e], etc., which display the highest degree of sonority)

Phonological oppositions are clear, notwithstanding idiosyncratic differences: e.g. there are no differences linked to place of articulation within these different groups [p t k] on the oscillogram.

Difficulties in the description of prosodic phenomena are all the more inconvenient as prosody is of great importance in communication. From a grammatical point of view, it plays a role in delimiting sentences or parts of sentences, determining sentence types (assertive, interrogative, imperative) or placing emphasis on the information structure of utterances (i.e. topic-comment structure, e.g. *this course* (= topic) *is interesting* (= comment). From a pragmatic point of view, prosody may convey the style (or ‘tenor’) of a discourse (formal or informal), and reflect the expression of politeness, irony or specific emotions (different emotional states, such as sadness, boredom or happiness induce strong variation in intonation). It may provide information on the state of health or regional origin of an individual, as well as differences in age, sex, social class or status.

Finally, it may reveal intonational idiosyncrasies (i.e. particularities of intonation linked to the speech of an individual). In order to identify an individual voice, it is necessary to consider a complex interplay of factors, including average pitch, the specific realisation of certain phonemes (vowels in particular, as they may be more or less open), or habitual rhythm patterns.

Prosody also plays a decisive role in phonetic change, since the transformation of individual sounds is easily triggered by suprasegmental habits; however, the study of this type of interaction with regard to the earlier stages of a language is even more difficult than it is for present-day languages.

A fully functional prosodic element that is of particular importance for the phonological history of the Romance languages is Latin stress, since its transformation in Late Latin catalysed numerous subsequent changes. Moreover, the position of stress in Latin words retains its full importance for later evolution, since some sound changes rely on this factor (as an example, stressed Latin /a/ in open syllables becomes /ɛ/ in Modern French; unstressed word-final /a/ is reduced to /ə/, becomes labialised, and then disappears).

The most important rules of word stress in Latin can be summarised as follows:

- in two-syllable words, stress falls on the first syllable, e.g. 'ROSA, 'ROSAE (diphthongs constitute single, long syllables);
- in Latin words of more than two syllables, stress falls on the syllable before last (or the 'penultimate' syllable) if the latter is long by nature (as in A'MĪCA, GENE'RĀLIS) or long by position (i.e. a short vowel preceding two consonants or a geminate, as in CON'TĒTUS or AN'CĪLLA. *Muta cum liquida* combinations are an exception (cf. above), as they do not cause positional lengthening). Such words are referred to as 'paroxytons';
- if the penultimate syllable is short by nature *and* by position, stress moves to the syllable preceding it (the 'antepenultimate', as in 'DŌMINA, 'FABULA). Theoretically, this also applies to short penultimate syllables before *muta cum liquida*: 'TENĒBRAE, 'PATRŌCLUS (such words are known as 'proparoxytons'); in reality, however, the position of stress in these examples is subject to variation (e.g. Sp. *tinieblas* presupposes the Latin form TE'NĒBRAE, the diphthong having evolved from a stressed Ē).

→ Loporcaro, *Syllable, segment and prosody*, CambrHist 1, 2
 Marotta, *Prosodic structure*, OxfGuide 26
 Meinschaefer, *Effects of Stress*, CambrHandb 7
 Bertinetto, *Rhythm in the Romance languages*, OxfEnc

6.4 Writing, graphemics and orthography

6.4.1 Alphabetic writing

The importance of writing in the evolution of languages has already been mentioned (cf. 1.2.5). If a language has been put into writing, research on its pronunciation must take into account its written representation, even in synchrony. Graphemics (i.e. the study of the organisation of graphemes as basic units of writing) becomes crucial to

diachronic phonetics, since it allows the phonological systems of earlier language stages to be reconstructed, and the processes of phonetic change to be retraced.

Three main categories of writing systems are generally distinguished in the context of the history of writing: ideographic, syllabic and alphabetic. The oldest writing systems are ideographic in nature. These include early Sumerian cuneiforms, Egyptian hieroglyphs, Chinese characters or the Meso-American writing of the Maya. These systems rely on a repertoire of several thousand 'ideograms', each of which is used to express one or more lexical meanings. A long learning process is required to master such writing systems; moreover, they are not adapted for the transcription of proper nouns (as the latter are desemanticised by definition). Most ideographic systems ultimately adopt a syllabic system for proper nouns; in this case, instead of representing entire words, the ideograms stand only for their initial syllable. Consequently, ideographic writing systems such as Ancient Egyptian or Chinese simultaneously rely on two different principles – one ideographic and one syllabic (or phonographic). A further step is made when a syllabic system becomes alphabetic: the former ideogram then represents only the initial sound of a word.

In their development, writing systems display a high degree of interconnectedness. Almost every system is based on previous systems, which it adapts and transforms. It is nevertheless likely that writing was invented more than once by different cultures, independently of one another. Cuneiforms (which appeared a little before 3000 BC) and hieroglyphs (around 3000 BC) may have been interrelated; Meso-American writing (600 BC) and Chinese writing (around 1300 BC), however, are independent inventions.

Hieroglyphs served as a model throughout the entire Mediterranean zone until the Latin alphabet appeared; from 1700 BC onwards, Egyptian writing gradually evolved into an alphabetic system, which was adopted by different Semitic languages of the Near East. This purely consonantal system is the basis of the Arabic, Hebrew and Phoenician writing systems. The Phoenician alphabet was later adopted into Greek, and some of its initially consonantal graphemes were modified and attributed a vocalic value, thereby allowing the systematic representation of vowels (in the 8th-century BC). This writing system became known first to the Etruscans and then to the Romans (in the 7th century BC).

The Latin alphabet became the basis for nearly all written Romance varieties, even though there are a number of medieval Romance texts (from Italy) written in the Greek alphabet and some (mostly from the Iberian Peninsula) written in Arabic or Hebrew letters. Only Romanian adopted the Cyrillic alphabet, which would be replaced by the Latin alphabet, but not until much later on (*cf.* 11.2.4). Moreover, from the Middle Ages onwards, a wide range of different scripts were developed to represent the letters of the Latin alphabet (*cursiva*, *libraria*, printed letters, etc.), each with its own repertoire of calligraphic varieties (*cf. ibid.*).

Alphabetic writing, generally thought of as simple and close to spoken pronunciation, is not, however, completely free of ideographic elements: the separation of words at the graphic level (which does not necessarily correspond to phonetic segmenta-

tion), and the use of capital letters (which only became established during the 18th and 19th century) or punctuation (which reflects specific syntactic interpretations) are all aspects which render semantic information visible to the eye. The same applies for the layout of texts (concerning the function of titles, paragraphs, changes of font or typeface, lists, tables, etc.), which transforms the linearity of writing into a hierarchical whole (cf. Raible, *Die Semiotik der Textgestalt*, 1991; Frank, *Die Textgestalt als Zeichen*, 1994: 25 sqq.).

Today, the malleability of alphabetic writing surfaces in text messages, which reflect a particularly high degree of phono-graphemic reinterpretation. Throughout the history of Latin and Romance writing, then, such ‘external’ elements can be seen to have undergone changes that are equally as significant as those affecting the essential relationship between graphemes and phonemes.

→ Finbow, *Writing systems*, OxfGuide 41

6.4.2 The relationship between writing and pronunciation throughout history

Alphabetic writing systems rely on the principle of a direct relationship between graphic signs (<r>, conventionally noted between angle brackets) and sounds ([r]) or phonemes (/r/). In theory, each grapheme represents a specific phoneme, and this phoneme, in turn, is correlated with its corresponding grapheme.

In reality, however, the grapheme-phoneme relationship is never as immediate or as uncomplicated as one might hope. Thus, a phonologically pertinent feature such as vowel length in Latin, for instance, cannot be conveyed by means of the 24 characters of the Latin alphabet. The phonetic diversity of the Romance languages (with repertoires ranging from 29 to 41 speech sounds; cf. the tables in 6.2.3) has led to the introduction of digrams such as <lh, nh> and a series of diacritic symbols:

- the cedilla (<ç>, <ç>, ‘e-caudata’, used to represent [ɛ]), introduced in the medieval period;
- the tilde (<ñ>, e.g. Sp. *año* “year”), which appeared in the 16th century and is still in use in Spanish today;
- various accents, which are especially prevalent in French and Portuguese (<é>, <è>, <â>);
- since the 19th century, subscript diacritics in Romanian (e.g. <ș>, <ț>, alongside <î>, <â>, <ă>).

One of the many ways in which the impact of tradition on writing systems becomes apparent is through resistance to the introduction of completely new characters into an existing system. In order to represent new sounds, existing characters tend to be modified by means of diacritics. Projects to reform and modernise writing systems,

such as that of the 16th-century grammarian Louis Meigret, who devised a ‘reformed orthography’ for the French language (with new graphemes, such as <ɛ> for [ɛ]), could not become established.

The high degree of graphic variation in medieval texts reflects the difficulties that scribes must have experienced in comprehending simple equivalences between writing and pronunciation (cf. 10.4.3 no. 3). It should also be kept in mind that the desire to establish a standardised orthography arose only later. When dealing with medieval texts, then, the present-day reader inevitably faces the daunting task of identifying the sounds underlying a multitude of different variants. As an example, *molher*, *moiller*, *moller*, *molier*, *moillier* and *moilher* are all allographs (i.e. phonologically or phonetically identical graphic representations) of Old Occitan [mɔʎɛr], meaning ‘wife’¹⁵. Due to the influence of modern orthography, editors of medieval texts sometimes have a tendency to reduce the amount of graphic variation present in the sources (cf. 11.5.1). This procedure is detrimental to the authenticity of the text, as pertinent graphemic information is thereby eliminated (cf. 6.4.4).

The use of single stable graphic forms – at least in particular positions or for particular words – only occurs after a language has been consciously reduced to a standardised written form. In European languages, orthography (writing according to norms) was not elaborated until fairly recently, during the 16th and 17th centuries, as a result of the standardisation of modern languages (cf. 10.5.3 no. 3).

The establishment of fixed rules in written language, however, leads to discrepancies between writing and pronunciation, as phonetic change takes place. With the passing of time, pronunciation becomes increasingly distant from the established graphic norms. This is particularly apparent in English, where the correct pronunciation of written forms requires prior knowledge of the words in question, but also in French or Portuguese, where homophones (i.e. words with identical pronunciation) occur frequently. In French, for example, the phonological form /sɛ̃/ may correspond to five different words <saint, sain, sein, ceint, seing>, without counting <cinq> as pronounced in <cinq mille>. Orthographic reforms are catalysed by linguistic inconsistencies such as these; however, they are motivated above all by socio-political circumstances.

As in the case of sounds and phonemes, writing displays positional or allographic variants; e.g. the phoneme /k/ in modern (and medieval) French is represented by two

¹⁵ If one acquires the habit of reading words in medieval texts aloud, and then comparing their pronunciation with the phonological realisation of their corresponding modern forms (which is perfectly legitimate from a theoretical point of view), they become considerably easier to interpret. It is nevertheless important to keep in mind that the meaning of a word may have changed significantly over the span of seven or eight centuries of evolution. Identifying the modern equivalent of a medieval form does not resolve the question of its semantic interpretation, although it facilitates it in some cases, since the evolution of meaning also follows regular patterns.

allographs: <c> before /a o u/ and <qu>, which is obligatory before /e i/ and sometimes appears before /a/ as well (consider the following examples: *camion*, *cour*, *querelle*, *quitter*, but also *quartier*). Theoretically, any given case of homophony (such as the examples cited for /sẽ/) can also be interpreted as allography.

In extreme cases, the dependency of writing on pronunciation may be reversed: if the written form of a word suggests a phonological opposition, it may exert influence on the pronunciation of the word in question. An illustrative example is the case of ‘orthographic pronunciation’ (e.g. in the French place names *Bruxelles* and *Auxerre*, the <x> is pronounced [-ks-], owing to the influence of the written forms, instead of [-s-], in keeping with phonetic history; cf. also the familiar form *schei* [ˈskeɪ] ‘money’, used in the Veneto, after the Italian pronunciation of the written German form *Schei*[*demünze*] [ˈʃaj(dəmyntsə)] stamped on a type of coin formerly issued in Austria).

6.4.3 Orthography and society

The immense variety of graphic forms that coexisted in medieval Romance languages was systematically reduced through the linguistic reforms of the 16th and 17th centuries. The establishment of an orthography played a central role in discussions focusing on language during this period. National academies such as the *Accademia della Crusca*, the *Académie française* or the *Real Academia española* (cf. 10.5.3 no. 4) each envisaged a work on orthography, as well as a dictionary and a grammar for their respective languages¹⁶.

A written language conveys a strong sense of identity, both in the choice of its alphabet and its graphemes and in the visual aspects of its presentation. With regard to the establishment of modern countries, the sense of belonging to a nation (as well as the desire of excluding divergent tendencies) has had strong political impact. One need only think of the transition from the Cyrillic to the Latin alphabet that was implemented in Moldova in 1991. In this respect, it is not surprising that discussions of orthographic reforms continued to occur with remarkable vigour and irrationality at the end of the 20th century.

The national dimension of orthography, moreover, increases the distance between standardised writing and pronunciation, whereby the latter retains a higher degree of individuality. In a country of 50 million inhabitants, there are multiple varieties of a language with varied pronunciations, which the written standard cannot take into

¹⁶ Though the Spanish academy published its work on orthography (*Ortografía de la Real Academia española*, 1713), the French and Italian academies did not (cf. instead Meigret, *Traité touchant le commun usage de l'écriture françoise*, 1542 and Accarisio, *Vocabolario, Grammatica, et ortographia de la lingua volgare [...] con ispositioni di molti luoghi di Dante, del Petrarca et del Boccaccio*, 1543).

account. A direct relationship can thus be established between the size of the communicative space covered by a specific type of writing and the gap between it and the spoken varieties of the language in question. Only a written language that is used within a very small area can reflect the pronunciation of its speakers in a more or less accurate manner.

Hence, the relationship between writing and pronunciation during a given historical period is determined by diverse and partially contradictory factors.

6.4.4 The importance of graphemes for phonetic history

The history of graphemes is fundamental to an understanding of the history of phonemes: it is only through writing that we can acquire an idea of the likely pronunciation of words in periods prior to the invention of systems of phonetic transcription and tape recorders. The consideration of various phenomena can shed light on the earlier phonological stages of a language:

1. Phenomena of hypercorrection (i.e. examples of ‘faulty’ usage that result from the intention to respect linguistic norms – and thus to *avoid* committing errors):

- as an example, writing ‘éthymologique’ instead of ‘étymologique’ in French (because one recognises the word as a Hellenism) provides evidence of the fact that the graphemes <th> and <t> have an equivalent phonetic status in Modern French;
- as a further example, in the 1st century BC, the Latin poet Catullus described the pronunciation habits of one of his contemporaries, noting: ‘*Chommoda*’ dicebat, si quando ‘*commoda*’ vellet dicere, et ‘*insidias*’ Arrius ‘*hinsidias*’ (84, 1 ss) (“Arrius said ‘chommoda’ when he wanted to say ‘commoda’, and ‘hinsidias’ instead of ‘insidias’”). This observation shows that aspirated [h] was still present in spoken Latin at the time, though already considerably weakened.

2. Examples of graphic equivalence, such as those reflected in the following inscriptions discovered at Pompei (all dated before 79 AD):

- *scribenti mii* [for *mihi*] *dictat Amor mostratque* [for *monstratque*] *Cupido* (“Amor dictates to me what I am writing and Cupid shows me”)
- *carminibus Circe socios mutavet Olyxis* [for *mutavit Ulyssis*] (“Through her incantations, Circe transformed the companions of Ulysses (into pigs)”)
- *fonticulus pisciculo suo plurma salut* [for *plurimam salutem*] (“Many greetings from Little Fountain to her Little Fishie”)

These three examples bear witness to the following phonetic changes: loss of aspirated [h], [ns] > [s], [i] > [e], [ü] > [o], loss of post-stressed vowels and loss of word-final [m].

3. Metalinguistic evidence: the phenomena affecting spoken Latin mentioned above are confirmed by the *Appendix Probi*, a list of 227 [220] linguistic observations compiled in the 5th century AD, which endeavours to correct spoken mistakes:

- *tensa non tesa*
- *turma non torma*
- *calida non calda*

The text survives appended to an 8th-century palimpsest manuscript containing a grammar treatise wrongly attributed to the 1st-century grammarian Valerius Probus (hence its title) (*cf.* 11.2.3).

4. Borrowing: e.g. the German terms *Keller* and *Kiste* provide evidence of the fact that the Latin forms *cellariu* and *cista* were pronounced with a [k] at the time they were borrowed; had the Latin words not been pronounced with initial [k-], the German words would have been different.

5. Rhymes in poetry, which allow one to compare and distinguish vowel quality in Old and Middle French or Old Occitan.

In contrast, writing allows only very limited analysis of the prosody of earlier language stages: as a rare example, the punctuation in manuscripts may be correlated to some extent with the distribution of speech pauses. In the Middle Ages, punctuation ‘systems’ were in a state of constant flux due to the lack of orthographic rules; nevertheless, some observations regarding prosodic groups can be made.

This domain remains little studied even today; punctuation loses much of its significance in literary manuscripts that have been transmitted through copies and that have not survived in their ‘original’ form. In addition, editors of medieval texts allow themselves more liberty in modifying the punctuation of a text than its spelling. It is thus necessary to refer to the actual manuscripts in order to adequately address this issue.

6.5 Major trends of phonetic evolution in the *Romània*

6.5.1 The Late Latin vowel systems and their evolution

The most important evolutionary process affecting the vowel system of spoken Latin was the loss of the phonological opposition between long and short vowels (vowel length). This opposition was then replaced by that of open and closed vowels (vowel height and quality), which had existed previously, but without having the status of a distinctive feature: long vowels tended to be closed, and short vowels, open (in allophonic distribution). The distinction of length was thus transformed into a distinction of height

(or quality). This large-scale transformation was catalysed by vowel stress patterns and affected the entire vowel system.

At the same time, the opposition between some vocalic phonemes tended towards neutralisation: in most parts of the future *România*, the opposition between /ē/ and /ī/ disappeared, and the two former phonemes merged into the new phoneme /e/; in parallel, /ō/ and /ū/ evolved into /o/, and /ā/ and /ǣ/, into /a/, as shown in the table below, which provides a synthetic overview of the tonic vowels present in most Romance languages:

ī	ĩ	ē	ě	ā	ǣ	ō	ō	ū	ū
i	e		ε	a		ɔ	o		u

There are other Proto-Romance vowel systems in addition to the common system presented above (the so-called ‘Italic system’, no. 1 in the table below). These include the system that would arise in Romanian and some areas of Basilicata in southern Italy (the so-called ‘Balkan system’, no. 2 below; cf. Rom.n *gură* < GŬLA), the system that would develop in Sardinian as well as in the frontier dialects between Calabria and Lucania, in the (lost) Latin of Africa and in Latin loanwords in Basque (the so-called ‘archaic’ or ‘Sardinian’ system, no. 3; cf. Sard. *pīlu* < PĪLU, *bula* < GŬLA), and the system found in Sicilian, as well as in southern Calabrian and Salentino (southern Apulia) (the ‘Sicilian’ system, no. 4). The following table provides a general overview of the tonic vowels and examples are given for the most salient changes (cf. Lausberg, *Romanische Sprachwissenschaft*, vol. 1, 1956: 95–101 and 6.5.2 below for further details):

		ī	ĩ	ē	ě	ā	ǣ	ō	ō	ū	ū
(1)	Italic system	i	e	ε		a	ɔ		o		u
(2)	Balkan system	i	e		ε	a		o		u	
(3)	Sardinian system	i		e		a		o		u	
(4)	Sicilian system	i			ε	a		ɔ	u		

- ĩ: ‘PĪLU “(body) hair” > Pg. Sp. It. *pelo*, Cat. Occ. *pel*; displaying further evolution: Fr. *poil*, Rom. *peil*; Rom.n *păr*; Sard. Sic. *pīlu*
- ě: ‘STĒL(L)A “star” > Pg. *estrela*, Sp. *estrella*, Cat. Occ. *estela*, It. *stella*; displaying further evolution, as above: Fr. *étoile* and Rom. *steila*; Rom.n *stea*, *steauă*; Sard. *isteddu*; Sic. *stidda*
- ō: ‘VŌCE “voice” > Pg. Sp. *voz*, Cat. *veu*, Occ. *votz*, It. *voce*; displaying further evolution: Fr. *voix*, Rom. *vusch*; Rom.n *boace*; Sard. *boge*; Sic. *vuci*
- ū: ‘GŬLA “throat, mouth” > Sp. (> Pg.) Cat. Occ. It. *gola*, O.Fr. *gole* (> Fr. *gueule*); Rom.n *gură*; Sard. *bula*; Sic. *gula*

The transformation is similar for unstressed vowels, but with an even stronger tendency towards neutralisation; cf. the overview and the examples (again following Lausberg, *ibid.*, pp. 140–141):

		ī	ĩ	ē	ě	ǎ	ǒ	ō	ű	ū
(1)	Italic system	i	e			a	o			u
(2)	Balkan system	i	e			a	u			
(3)	Sardinian system	i		e		a	o		u	
(4)	Sicilian system	i				a	u			

- ī: HĪBERNU “winter” > Pg. *inverno* [interference with *in-*], Sp. *invierno*, Cat. Occ. O.Fr. *iver*, Fr. *hiver*; Rom.n *îarnă*; Sard. *ierru*; Sic. *invernu*
- ĩ: CĪR̄CARE “to search” > Pg. Sp. Cat. Occ. *cercar*, Fr. *chercher*, It. *cercare*; Rom.n *cerca*; Sard. *kirkare*; Sic. *circari*
- ē: SĒCURU “secure” > Pg. Sp. *seguro*, Cat. Occ. *segur*, O.Fr. *seūr* (> Fr. *sûr*), O.Ven. *seguro*, O.Centr.It. *securu* ~ O.Tusc. *sicuro* [= pretonic alternation, typical of Tuscan] (> It. *sicuro*); [Rom.n $\emptyset$]; Sard. *seguru*; Sic. *sicuru*
- ě: NĚPOTE “nephew” > Occ. *nebot*, Fr. *neveu*, It. *nipote*; Rom.n *nepot*; Sard. *nebode*; Sic. *niputi*
- a: LAVARE “to wash” > Pg. Sp. Cat. Occ. *lavar*, Fr. *laver*, It. *lavare*; Rom.n *la* (< **lǎă*); Sic. *lavari*
- ǒ: PŌRTARE “to carry” > Pg. Sp. Cat. Occ. *portar*, Fr. *porter*, It. *portare*; Rom.n *purtă*; Sard. *portare*; Sic. *purtari*
- ō: MŌRU “mulberry; blackberry” (with a change of gender from n.pl. to f.sg. *MORA) > Pg. *amora*, Sp. Cat. Occ. It. *mōra*, O.Fr. *moure*; Rum.n *mură*; Sard. *mura*; cf. Sic. *muredđa*
- ű: FŪR̄CA “fork” > Pg. Cat. Occ. *forca*, Sp. *horca*, O.Fr. *forche* (> Fr. *fourche*), It. *forca*; Rom.n *furcă*; Sard. *furka*; Sic. *furca*
- ū: MŪTARE “to change” > Pg. Sp. Occ. *mudar*, Fr. *muer*, It. *mutare*; Rom.n *mută*; Sard. *mudare*; Sic. *mutari*

The seven tonic and the five unstressed vowels of the common Romance vowel system followed independent patterns of evolution, which, nonetheless, display parallels in the various Romance languages. In addition to its dependence on backness and height, as well as on word- and syllable-internal position, the evolution of Romance vowels has been closely correlated with stress patterns: stressed vowels are generally subject to more diversified transformations than unstressed vowels – which tend to be reduced in number or lost, or to develop free variants (cf. 6.5.3). Similarly, vowels in open syllables are more susceptible to change than those in closed syllables (cf. 6.3).

As vowel length had been lost in Latin, this parameter was not an original feature of the Romance vowel systems, although it arose at certain periods in different languages. In Old French, for instance, long vowels (by position) were distinguished from short

vowels until at least the 16th century: *pâte* /a:/ ~ *patte* /a/, *bête* /ɛ:/ ~ *bette* (ɛ) (< BESTIA ~ BETA, cf. 6.3); cf. also the vowel length distinctions found in present-day Friulian or Emilian dialects (e.g. in Bolognese).

Generally speaking, the evolutionary trends affecting vowels overlapped chronologically; a number of Romance phenomena occurred between the 1st and the 5th centuries before becoming generalised in the 6th to 7th centuries, by which time their geolinguistic patterns of distribution had already become quite distinctive. For the various Romance languages it would therefore theoretically be possible to describe a series of new phonological systems at two-century intervals with some degree of accuracy.

To give one example, the vowel system of Old French around 1100 had nine simple oral vowels /i e ɛ ə a ɔ u y/, five diphthongs /jɛ wɔ ~ wɛ ej ~ oj ew aw/, two triphthongs /eaw jew/, as well as nasalised vowels. The phonematic value of the complex vocalic sounds is uncertain; one may note, however, that whenever they appear in rhymes, they are distinguished from simple vowels. Differences between the phonological system of Old French and those of Latin and modern French are thus clearly apparent.

6.5.2 Stressed vowels in the Romance languages

The principal transformations that have affected Romance vowels are processes of diphthongisation (i.e. the transformation of simple vowels into diphthongs). These processes involved the vowels /a e ɛ ɔ/ (but not the ‘extreme’ vowels /i/ and /u/) and can be divided into three groups:

1 Spontaneous ‘Romance’ diphthongisation

This is the oldest and most generalised type of diphthongisation (which took place as early as the 3rd or 4th century AD). It was not conditioned by the consonantal environment of the vowels affected, but, in some cases, it depended on the type of syllable (open vs. closed) in which the vowel appeared; the process involved the vowels [ɛ] and [ɔ] (i.e. stressed *ɛ* and *ɔ*), which were subject to diphthongisation in many Romance languages (though not in Occitan), for the most part in open syllables. Moreover, in Spanish and Romanian (as well as in Vegliot [= Dalmatian]), the two vowels were also diphthongised in closed syllables, as the following examples show:

- 'ɛ[> jɛ: PĚTRA “stone” > It. *pietra*, Sp. *piedra*, Fr. *pierre*; but: Pg. Cat. *pedra*
- > ja: PĚTRA “id.” > Rom.n *piatră*
- 'ɛ] > jɛ: HĚR-BA “grass” > Sp. *hierba*, but: It. *erba*, Fr. *herbe*
- > ja: HĚR-BA “id.” > Rom.n *iarbă*
- 'ɔ[> ʊɔ > ʊɛ (> œ ~ ø):
- NŌVU “new” > It. *nuovo*, Sp. *nuevo*, O.Fr. *nuef* [ʊɛ], Fr. *neuf* [œ]

PRŌBA “proof” > O.It. *pruova* [> It. *prova*, with simplification after word-initial consonant cluster], Sp. *prueba* [ʎɛ], Fr. *preuve* [œ]

FŌCU “fire” > It. *fuoco* [ʎɔ], Sp. *fuego* [ʎɛ], Fr. *feu* [ø]

RŌTA “wheel” > It. *ruota*, Sp. *rueda*, O.Fr. *ruede* (> Fr. *roue* [u], displaying further evolution), Rom.n *roată* [ɔa]

The French evolution of [ʎɛ] through [ʎø] to [ø] took place towards the end of the 12th century; during the following centuries, the vowel was opened to [œ] before consonants, leading to convergence with the outcome of Latin /o/ > [œ]; the older [ø] (*feu*) and the secondary [œ] before consonants (*preuve*) thus became allophones.

ʎ] > ʎɛ: MŌR-TE “death” > Sp. *muerte* [ʎɛ]; but: It. *morte*, Fr. *mort*

2 Spontaneous ‘French’ diphthongisation (6th to 8th centuries)

The so-called ‘French’ diphthongisation also occurs in Francoprovençal and, to some extent, the dialects of Northern Italy and Romansh. It took place during the transitional period between Latin and Romance, thus constituting evidence for the close connections between French and Francoprovençal at that time. The emerging varieties of Francoprovençal, Northern Italian and Romansh may also have been interconnected. Different stressed vowels are concerned:

- [a] has been maintained in the majority of Romance languages, and palatalised in open syllables in Fr., Frpr. (after [k]) and Rom., but also in Rom.n and in some dialects of southern and north-eastern Italy (e.g. the province of Bari or Parma (pronounced [pærma] in Parmigiano); cf. for Fr.:

ʎa] > aɛ > ɛ ~ e: TALE > fr. *tel* /tɛl/ “such”, GRATU > Fr. *gré* /gre/ “willingness”

The intermediate diphthong [aɛ] is attested sporadically (cf. *maent* < Lat. *manet* “stays”, *Canticle of Saint Eulalia*).

- [e] and [o] have also remained fairly stable in the *România*, once again with the above-mentioned exceptions; in French they are diphthongised in open syllables, and in Romansh closed syllables are also affected by diphthongisation, cf. ʎe] > ej > oj > wɛ > wa ~ ɛ: MĒ > O.Fr. *mei* > (O.)Fr. *moi* “me” (pronounced differently depending on the historical period); TĒLA > O.Fr. *teile*, Fr. *toile*, Rom. *teila*, Piedmont. [tɛjla], Ligurian [tɛj(r)a], Romagnol [tʰajla] “canvas; web”

The last evolution in French (ʎɛ > ʎa ~ e) took place around the 16th to 17th centuries (*François* ~ *Français*; cf. 6.4.3).

- ʎo] > ow > ew > ø ~ œ: FLŌRE “flower” > O.Fr. *flour* [ow] > Fr. *fleur* [ew] ~ [œ], Emilian [ʎjɔwr], Romagnol [ʎjɔwr]

The evolution in French came to a close around the 13th century (cf. ʎ above).

- ʎo] > we: URSU “bear” > Rom. *uers* [ʎɛ] vs. Fr. *ours*

→ Maiden, *Diphthongization*, OxfGuide 38

3 Diphthongisation conditioned by palatal consonants

This is the most complex chapter of diphthongisation, owing to the fact that palatal and palatalised consonants (or clusters of palatalised consonants) also displayed a tendency towards independent evolution. With regard to the history of the Romance languages, this has resulted in an overlap of vocalic and consonantal evolutionary processes. Observations will be limited to a few examples:

- CARU “dear” > O.Fr. *chier* [ʰjɛr] > Fr. *cher* [ʃɛR], CAPRA- “goat” > O.Fr. *chievre* [ʰjɛvrə] (interference between the evolution of the word-initial palatal [k] and the diphthongisation of 'a[= ‘Bartsch’s law’), Fr. *chèvre*, Frpr. *chièvra* (the diphthong has been monophthongised in Mod.Fr. but is maintained in Frpr.)
CANE “dog” > *chien* [jɛ] (represents a special case, as the nasal consonant prevents monophthongisation)
LĒCTU “bed” > O.Fr. *lieit* [ʰjɛjt] (formation of a triphthong where /k/, palatalised before /t/, is vocalised as /j/) > Fr. *lit* [li] (monophthongisation)
 - PRIMARIU “first” > Fr. *premier* (interference between the diphthongisation of 'a[and the evolution of the intervocalic group /rj/)
 - FÖLIA- “leaf” > O.Fr. *fuelle* [ʰfwɛlə] > Fr. *feuille* [ʰfœj]
- The examples *premier* and *feuille* are cases of conditioned (or metaphonic) diphthongisation.

4 The ‘extreme’ vowels [i u]

These vowels have remained stable in stressed positions in the Romance languages, the only exception being the palatalisation of [u] in Gallo-Romance, in numerous dialects of northern Italy (Piedmontese, Ligurian, Lombard) and in Romansh (as well as in Portuguese varieties of the Algarve):

- LŪNA “moon” > Fr. *lune* [ʰlyn], Occ. *luna* [ʰlyno], Rom. (Engad.) *luna* [ʰlyna ~ ʰlyna]

5 Monophthongisation

The major counterparts of diphthongisation in the *Romània* involve processes of monophthongisation (cf. 6.7). Their number is significant: any diphthong at any moment can give way to a simple vowel; this is all the more true for triphthongs, which generally have a short life span, as they can easily be reduced to diphthongs or even monophthongs (e.g. the early change Lat. LĒCTU > O.Fr. *[liej] > [li]).

5.1 Latin or Romance monophthongisation

Monophthongisation was not generalised in Latin and only concerns some Romance languages:

Lat. AU > ō ~ ǫ:

- CAUDA “tail” > CŌDA > It. Sp. *coda*, Fr. *queue* (‘o[was later diphthongised, then monophthongised)
- CAUSA “cause” > CŌSA “thing” > Sp. *cosa*; but: Occ. *cauza* (evolution impeded due to the influence of the learned Latin term), Fr. *chose* (palatalisation of initial [k] presupposes the former presence of a diphthong), Pg. *coisa*
AUDI(T) “hear 3sg” > It. *ode*; but: Rom.n *aude* (form pronounced with (weak) diaeresis / hiatus), Sursilv. (Rom.) *áuda*, Occ. *au*, (O.) Pg. *ouve*
- TAURU “bull” > Sp. It. *toro*, Fr. → *taureau* [tɔˈro]; but: (O.)Pg. *touro*, Occ., Sursilv. *táur*, Sard. *trau*, south.It. *táuru*, Rom.n *taur*
- AE > ε: CAELU “sky” > It. *cielo*, Sp. *cielo*, Fr. *ciel* (‘ε[was normally diphthongised to ‘jε)
- OE > e: POENA “pain, punishment” > It. Sp. *pena*, Fr. *peine*: in French, [e] then opened into [ɛ] in the 16th century according to the ‘rule of position’ (= a closed vowel tends to open if it precedes a final consonant that is pronounced, cf. Revol 2000, 25)

5.2 Later Romance monophthongisation

Cf. nos. 1 and 3 above for the monophthongisation of Romance diphthongs and triphthongs. Some other cases include:

FACTU “deed” > *fait* O.Fr. [faj], Fr. [fɛ]
 MAGIS “more” > *mais* O.Fr. [maj], Fr. [mɛ]
 PLAGA “wound” > *plaie* O.Fr. [ˈplajə] > [plaj] > [plɛ]
 PACARE “pacify, subdue” → Fr. *la paye* [pɛj ~ pɛ]

A vocalic evolution specific to French, Portuguese and some dialects of northern Italy was the nasalisation of vowels preceding the nasal consonants [m, n]. Note that in French:

- nasalisation processes took place between the 11th and 14th century, depending on the vowels in question: MANU “hand” > O.Fr. [ˈmɛ̃n];
- depending on the case, denasalisation processes during the 16th to 17th centuries consisted either in the loss of the nasal consonant in closed syllables (*main*: Old.Fr. [ˈmɛ̃n] > Fr. [ˈmɛ̃]), or the denasalisation of the vowel in open syllables (*an-née*: O.Fr. [ãˈne] > Fr. [aˈne] “year”).

6.5.3 Unstressed vowels in the Romance languages

Unstressed vowels underwent the same loss of the distinction of quantity as stressed vowels, as well as the same processes of monophthongisation (AURICULA > It. *orecchia*, Sp. *oreja*, Fr. *oreille* “ear”); however, they were not subject to diphthongisation.

Since the evolution of vowels depends on stress, divergent variants of a Latin vowel can occur within a single word family in the *România* or even within a single verbal paradigm; cf. Modern French: *il tient*, *nous tenons*; *il reçoit*, *nous recevons*; *Pierre*, *perron* (cf. 7.3.4 and 7.6.1).

As we have seen, there are generally fewer unstressed vowels than stressed vowels in the Romance languages (cf. 6.5.1 and 6.2.3: Italian has five unstressed vs. seven stressed vowels; Portuguese, four unstressed vs. seven stressed, not counting nasal vowels; etc.). The development of these unstressed vowels in the different languages and dialects is varied and sometimes rather irregular, although it often follows a certain logic.

In word-final position – where the morphological significance of the vowels mainly concerns the gender and number of nouns – there are clear differences between the Romance varieties. In French, the descendants of the Latin final unstressed vowels have all been lost, with the exception of [-a]. In contrast, the final vowels are particularly stable in the dialects of central and southern Italy, as well as in Sardinian and Romanian.

The developments depend equally on the three main vowels involved, which are derived from Lat. [-a], [-ĕ] and [-ŭ] (the latter becoming either [-o] or [-u] depending on the Romance vowel system). Lat. [-a] is generally stable, except in French, where it weakened to [-ə] very early, around the 7th century; it was then labialised in the 15th century before being dropped in the 17th century. In the other Romance languages, [-a] has been maintained or – in certain cases – weakened:

RIPA “shore, bank” > O.Fr. *rive* [ʁivə] > Fr. *rive* [ʁiv] vs. Friul. *rive* [rive], Occ. *ribo* [ribɔ], Pg. *riba* [-ə] vs. Sp. Cat. Sard. *riba*, Frpr. Rom. (Engad.) *riva*, It. *ripa* (south), *riva* (north), Rom.n *râpă*

Lat. [-ĕ] has disappeared from most Romance languages and is only stable in Italian, Sardinian and Romanian (= Wartburg’s ‘Eastern *România*’, cf. 3.5.3):

CANE “dog” > Fr. *chien*, Frpr. *tšī*, Rom. (Sursilv.) *tgaun*, Lad. *ćian*, Friul. *tšan*, Sp. *can* (“hound”), Cat. (Mall.) *ca* vs. It. Sard. *cang*, Rom.n *câine* vs. Sic. *cani*; cf. Pg. *cão* (= adaptation to a different inflectional class)

FLORE “flower” > Fr. *fleur*, Frpr. *fjœ*, Pg. Sp. Cat. Occ. Friul. *flor*, Rom. (Sursilv.) *flur* vs. It. *fiore*, Sard. *florē*, Rom.n *floare* vs. Sic. *ciuri*

LEVĀRE “to raise” > Fr. *lever*, Occ. Rom. Pg. *levar*, Cat. Sp. *llevar* vs. It. *levare*, Sard. *leare*, Rom.n *lua(re)* vs. Sic. *livari*

An intermediate position is occupied by Lat. [-ŭ], which is dropped in Gallo-Romance, including Catalan and the Eastern Alps, but is retained as [-o] or [-u] elsewhere:

MANU “hand” > Fr. *main*, Rom. (Engad.) *maun*, Friul. *man*, Occ. Cat. *ma* vs. It. Sp. *mano*, Pg. *mão* vs. Sard. Sic. *manu*; cf. Rom.n *mână* (= change of gender)
 OCULU “eye” > OC’LU > Fr. *œil*, Occ. *ueil*, Frpr. *we(Ń)*, Cat. *ull*, Rom. (Engad.) *œł*, Lad. *uédł*, Friul. *voli* vs. It. *occhio*, Sp. *ojo*, Pg. *olho* vs. Sard. *ogu*, Rom.n *ochi(ul)*, Sic. *occhiu*

As an exception after final consonant clusters, final support vowels are almost generalised in Romance languages (with the exception of Eastern Alpine varieties). The forms are relatively homogeneous for the cognates of Latin nouns ending in [-e]. In the case of Lat. [-ŭ], the languages which conserve this vowel follow the predictable phonetic development (becoming [-o] or [-u]), while the Gallo-Romance varieties mostly display the addition of [-e] or [-ə], with the exception of Francoprovençal, where [-o] is maintained (or perhaps restored for morphological reasons):

PATRE “father” > O.Fr. *pere* [‘perə] > Fr. *père* [‘per] “father”, Occ. *payre*, Frpr. Cat. *pare*, It. Sp. *padre*, Napol. *patre*, Sic. *patrì*; cf. Rom. *pàder* (= descendant of Lat. nominative PATER), Pg. *pai*
 GENERU “son-in-law” > GEN’RU > Fr. Cat. *gendre*, Occ. *gen(d)re* vs. Frpr. *gendro*, It. *genero*, Pg. *genro*, Sp. *yerno* vs. Sard. *bénneru*, Sic. *iennaru* vs. Rom. (Engad.) *gender*, Friul. *dzínar*

The post-tonic vowels in Latin proparoxytons constitute a special case, as they are subject to the phenomenon of syncope (cf. 6.3). The loss of these vowels had already occurred in Late Latin, but had not become generalised, resulting in disparities among the various Romance languages.

VETULU “old” > VET’LU ~ VECLU > *vecchio* [-k-], Rom.n *vechi* [-k-], Sp. *viejo* [-x-], Fr. *vieux* [vʲø]
 cf. OCULU “eye” > OC’LU, above (It. *occhio* [-k-], Rom.n *ochi* [-k-], Sp. *ojo* [-x-], Fr. *œil* [œj])
 cf. GENERU > GEN’RU, above

vs.

HEDERA “ivy” > Sp. *hiedra*, Pg. *hera*, Cat. *eira*, Occ. *elra*, O.Fr. *iere*, *liere* [with agglutination of the article], Fr. *lierre*; but: absence of syncope in It. *èdera*, Rom.n *iederă*
 FRAXINU “ash tree” > Sp. *fresno*, Pg. *freixo*, Cat. *freixe*, Occ. *fraise*, Fr. *frêne*; but: absence of syncope in It. *frassino*, Rom.n *frasîn*, Engad. (Rom.) *fresen*

Another peculiarity concerns the post-tonic Latin vowels ě and ĭ in hiatus. These were neutralised to /j/ at an early stage (VINEA “vineyard” /vinea/ (hiatus) > VINIA /vinja/ (diphthong), ALLEUM “garlic” /alleum/ > ALLIUM /alju(m)/. This development of *yod* later resulted in palatalisation.

Finally, a typical Gallo- and Ibero-Romance development is the introduction of epenthetic vowels [i e] before so-called *s-impurum* (i.e. /s/ followed by a plosive) in Latin:

SCALA “stairs, ladder” > Sp. Cat. Occ. *escala*, O.Fr. *eschele* (> Fr. *échelle*), Frpr. *etsila*, Sard. *iscala* vs. It. *scala*, Rom. (Engad.) *stséla*, Rom.n *scară*
 SPERARE “to hope” > Sp. Pg. Cat. Occ. *esperar*, Fr. *espérer*, Sard. *isperare* vs. It. *sperare*, Rom. (Engad.) *sperer*, Friul. *sperá*, Rom.n *spera*

6.5.4 Consonants in the Romance languages

Major trends of consonantal evolution within Romance languages include the following:

1. voicing, loss and degemination of intervocalic plosives;
2. palatalisation and assibilation (*cf.* no. 2 below) of numerous palatal sounds;
3. loss or devoicing of word-final consonants;
4. simplification of consonant clusters;
5. the majority of remaining changes are linked to specific sounds; these include the velarisation of [l], phenomena of metathesis affecting [r], the loss of [h] in Late Latin, the loss of intervocalic [n] and [l] in Portuguese and the evolution of [f] to [h] in Spanish and Gascon.

1 Voicing, loss and degemination of intervocalic plosives

These changes affect plosives in intervocalic position and before liquids (as in CAPRA, PETRA, LACRIMA “tear”). Intervocalic plosives are generally conserved in Italian and Romanian (with the exception of /b/ in Italian, the evolution of which coincides with that of /v/) and have undergone more or less extreme changes in other languages. The examples are presented in order of increasing degree of modification.

– Voicing and loss:

-p- > -b- > -β- > v	RIPA “shore, bank” CAPRA “goat”	> O.It. <i>ripa</i> , Rom.n <i>râpă</i> > It. <i>capra</i> , Rom.n <i>capră</i> , Sard. (Nuor.) <i>crapa</i>	vs. (O.) Pg. Sard. <i>riba</i> [b], Occ. Gasc. <i>ribo</i> [b], Sp. Cat. <i>riba</i> [β], It. <i>rive</i> , Fr. <i>rive</i> [v] vs. Occ. Pg. <i>cabra</i> [b], Sard. (Logud.) <i>craba</i> [b], Gasc. <i>crabo</i> [b], Sp. Cat. <i>cabra</i> [β], O.Fr. <i>chievre</i> , Fr. <i>chèvre</i> [v]
-t- > -d- > -ð- > Ø	VITA “life” PETRA “stone” -ATA (suffix)	> It. <i>vita</i> , Rom.n <i>vită</i> “live- stock, cattle”, Gasc. <i>bito</i> > It. <i>pietra</i> , Rom.n <i>piatră</i> > It. <i>-ata</i>	vs. Sp. <i>vida</i> [ˈbiða], Pg. <i>vida</i> [ˈvide], Cat. <i>vida</i> [ˈvidə], Sard. (Logud.) <i>vida</i> , Fr. <i>vie</i> [vi] vs. Sp. <i>piedra</i> , Pg. Cat. Rom. Sard. (Logud.) <i>pedra</i> [-d-] vs. Fr. <i>pierre</i> [-θ-], Occ. Gasc. <i>peyro</i> vs. Sp. <i>-ada</i> , Fr. <i>-ée</i> (épée “sword”, <i>journée</i> “day”)
-k- > -g- > -ɣ- > Ø	AMICA “friend” LACRIMA “tear”	> It. <i>amica</i> > It. <i>lacrima</i> , Rom.n <i>lacrimă</i>	vs. Sp. Pg. Cat. Occ. Sard. <i>amiga</i> vs. Fr. <i>amie</i> [aˈmi] vs. Occ. <i>lagremo</i> , Pg. Sard. <i>lâgrima</i> [g], Sp. <i>lágrima</i> [ɣ] vs. Fr. <i>larme</i> , Rom. <i>larma</i> , Gasc. <i>lermo</i>

-b- > -v- / Ø ~ -β-	PROBARE “to try”	> Gasc. <i>proubá</i> [b]	vs. It. <i>provare</i> , Sp. <i>probar</i> [β], Pg. <i>provar</i> , Occ. <i>provar</i> ~ <i>proar</i> , Fr. <i>prouver</i> [-v-]
d- > -ð- ~ -z- > Ø	LAUDARE “to praise”	> Gasc. <i>laudá</i> , Rom.n <i>lăuda</i> , It. <i>lodare</i>	vs. Occ. <i>lauzar</i> [-z-], O.Sp. <i>lodar</i> [ð] > Sp. <i>loar</i> , Cat. <i>lloar</i> , Fr. <i>louer</i> , Pg. <i>louvar</i>
-g- > -ɣ- / -j- > Ø	PLAGA “wound” NEGARE “to deny”	> Rom.n <i>plagă</i> , Occ. Gasc. <i>plago</i> , It. <i>piaga</i> > It. <i>negare</i> , Occ. Gasc. <i>negá</i>	vs. Pg. <i>chaga</i> , Sp. <i>llaga</i> [ɣ], N. Occ. <i>pladzo</i> vs. Fr. <i>plaie</i> [plɛ] vs. Pg. Sp. <i>negar</i> [ɣ], N. Occ. <i>niyá</i> [-j-] vs. Fr. <i>nier</i> [ni'e]

– Degemination:

-pp- > -p-	CUPPA “cask, barrel”	> It. Sard. <i>coppa</i> [-pp-]	vs. Rom.n <i>cupă</i> , Sp. Pg. Cat. <i>copa</i> , Fr. <i>coupe</i> [-p-]
-tt- > -t-	GUTTA “drop” QUATTUOR “four”	> It. <i>gotta</i> [-tt-] > It. <i>quattro</i>	vs. Sp. Pg. Cat. <i>gota</i> , Fr. <i>goutte</i> [ˈgut] vs. Sp. <i>cuatro</i> , Fr. <i>quatre</i> , Sard. <i>batro</i> , Rom.n <i>patru</i>
-kk- > -k- / -f-	BUCCA “cheek”	> It. <i>bocca</i> [-kk-] “mouth”	vs. Rom.n <i>bucă</i> “buttock”, Sp. Pg. Cat. <i>boca</i> [-k-], Gasc. <i>buca</i> , Fr. <i>bouche</i> [ˈbuʃ] “mouth”

The voicing and degemination of intervocalic plosives – along with the presence of sigmatic plurals (cf. 8.3.2 no. 6.3) – were considered by Wartburg to be key phonetic elements distinguishing the Western *România* (innovative) from the Eastern *România* (conservative, cf. 3.5.3). It should be noted, however, that the two phenomena are not evenly distributed: of the languages counted as belonging to the ‘Eastern *România*’ Romanian displays degemination (*cupă*, *patru*) and Sardinian additionally shows voicing (*riba*, *craba*, *vida*). In the ‘Western *România*’, moreover, Gascon and (on the other side of the Pyrenees) Navarrese exhibit degemination but not always voicing (*bito*, *proubá*, *laudá*, *plago*). The geographical distribution of these phenomena thus does not follow a simple bipartite structure.

2 Palatalisation and assibilation

The series of palatalisation and assibilation processes in the Romance languages is extensive and, at first glance, complex; it is nevertheless possible to group these processes coherently. Assibilation involves the development of a sibilant (a dental fricative [s z], a palato-alveolar fricative [ʃ ʒ] or an affricate [tʃ dʒ], which combines the fricative with a preceding plosive).

[k (e/i)] (= Latin *k* (written <c>) before *e* or *i*) was palatalised and assibilated in almost all of the Romance languages (with the exception of Dalmatian and central-northern

Sardinian), and transformed into a palato-alveolar affricate [tʃ], a dental affricate [ts], or simply a dental fricative [s] or an interdental fricative [θ]:

CIVITATE “citizenship” > It. *città* [tʃ-], Rom.n *cetate* [tʃ-], Sp. *ciudad* [ts- > θ- ~ s-], Pg. *cidade* [s-], Cat. *ciutat* [s-], Fr. *cité* [ts- > s-]

CENTU “hundred” > It. *cento* [tʃ-], Sp. *ciento*, Fr. *cent*

[g(e/i)] underwent a largely parallel evolution:

GELU “frost, icy coldness” > It. *gelo* [dʒ-], Rom.n *ger* [dʒ-], Fr. *gel* [dʒ- > ʒ-], Sp. *hielo* [j-]

The results of the evolution of word-initial (and syllable-initial) Latin [dj-] and [j-] coincide, in some cases, with those of [g(e/i)]:

HODIE “today” > It. *oggi* [dʒ-], Sp. *hoy* [j], Fr. (*aujourd’hui*) [j]

MEDIU “centre” > It. *mezzo* [dz-], [Sp. *medio* [-dj-]: Latinism], Pg. *meio* [-j-], Rom.n *miez*, Sard. *meiu*, Fr. *mi* [mi]

IAM “already” > It. *già* [dʒ-], Sp. *ya* [j-], O.Fr. *ja* [dʒ-] → Fr. *déjà* [ʒ-]

Latin [k(a)] and [g(a)] were only palatalised in *Gallo-Romance* (Fr., Fr.pr., south.Occ., Lad.), where they were transformed into [tʃ(a)] > [ʃ(a)] and [dʒ(a)] > [ʒ(a)], respectively:

CASA “house” > **chaese* [ʃʰaes] > Fr. *chez* [ʃe] (except in Normandy and Picardy: *Cambrai* < [CAMARĀCU], *Camembert* [cf. Med. Lat. *Campo Maimberti*])

CANE > Fr. *chien* [ʃ-], Lad. *ćian* [ʃ-]

GAUDIA “joy” > O.Fr. *joie* [dʒ-], Fr. ~[ʒ-] (Occ. *joi* [dʒ-] and It. *gioia* are Gallicisms)

To conclude the description of the evolution of Latin palatal plosives, [kʷ] and [gʷ] show a diversified development in the different languages: they either maintained their labial element or they were delabialised into [k] and [g], or they were even transformed into fully labial consonants:

QUATTUOR “four” > It. *quattro* [kʷ-], Sursilv. *quater* [kʷ-], Sp. *cuatro* [kʷ-] vs. Pg. *quatro* [k-], Cat. Fr. *quatre* [k-] vs. Sard. *báttoro* [b-], Rom.n *patru* [p-]

QUINDECIM “fifteen” > It. *quindici* [kʷ-], Sursilv. *quéndisch* [kʷ-] vs. Sp. *quince* [k-], Cat. Occ. *quince*, Fr. *quinze* [k-] vs. Sard. *bindighi* [b-]

LINGUA “tongue” > It. *lingua* [gʷ-], Sp. *lengua* [gʷ-], Pg. *língua* [gʷ-] vs. Friul. *lenge* [-g-], Cat. *llengua* [-g-], Occ. *lenga* [-g-], Fr. *langue* [-g-] vs. Sard. *limba* [-b-], Rom.n *limbă* [-b-]

The segments [nj, gn] as well as [lj, gʲl, kʲl] were regularly palatalised:

NJ: VINEA “wine” > It. *vigna*, Sp. *viña*, Pg. *vinha*, Fr. *vigne*, Sard. *bindza*

LJ: ALLIU “garlic” > It. *aglio* [-ʎ-], Sp. *ajo* [-x-], (O.Sp. [ʒ]), Fr. *ail* [ʎ] (O.Fr. [aʎ])

GN: LIGNU “gathered wood, firewood” > It. *legno*, Sp. *leño* “log”, Pg. *linho*, O.Fr. *lein*;

but: Rom.n *lemn*, Sard. *linnu*

GʲL: VIGILARE “to (keep) watch” > It. *vegliare* [-ʎ-], Fr. *veiller* [-j-] (O.Fr. [-ʎ-])

K'L: OCULU “eye” > It. *occhio* [-kj-], Sp. *ojo*, Fr. *oeil* [oej], Rom.n *ochi* (cf. 6.5.3 above)
 The apostrophe in [g'l] and [k'l] indicates that these are secondary consonant clusters, i.e. that they resulted from the syncope (in this case) of an unstressed vowel.

[kl-, pl-, fl-, gl-] were often palatalised, though not in French or Catalan:

CLAMARE “to call” > It. *chiamare* [kj-], Sp. *llamar* [ʎ-], Pg. *chamar* [j-] vs. Cat. *clamar*, Fr. *clamer*

[kt] was palatalised in Ibero- and Gallo-Romance:

FACTU “deed” > Sp. *hecho* [-tʃ-], Occ. *fach* [tʃ], Pg. *feito*, O.Fr. [ʔaj(t)] *fait* > fr. [ʔe] *fait* vs. It. *fatto* (by assimilation), Sard. *fattu*, Rom.n *fapt* (by dissimilation)

[nn, ll] were palatalised in Spanish and Catalan:

ANNU “year” > Sp. *año* [-ɲ], Cat. *any* [-ɲ]

CASTELLU “fort” > Sp. *castillo* [-ʎ-], Cat. *castell* [-ʎ]

[pj, bj] underwent affrication, followed by the assimilation of the initial labial [pj > pʃ > ʃ] / [bj > bʒ > ʒ], in French and in most dialects of southern Italy:

SAPIO “know 1SG.PRES.SUBJ.” > Fr. *sache* [-ʃ], Nap. *saccio*

*SAPIUS “wise” > Fr. *sage* [-ʒ]

RABIE “rage” > Fr. *rage* [-ʒ], Nap. *arraggia*

→ Barbato, *The early history of Romance palatalizations*, OxfEnc

Repetti, *Palatalization*, OxfGuide 39

Recasens, *Palatalizations in the Romance languages*, OxfEnc

3 Word-final consonants

Word-final [-m] and [-t] had already been lost in Latin around the 1st century AD; [-t], however, was retained wherever it functioned as an inflectional suffix, and it was also preserved as the final consonant in the conjunctions ET (“and”) and AUT (“or”).

In Medieval French, Occitan and Gascon as well as in the dialects of northern Italy, secondary word-final consonants (i.e. those which became final after the loss of Latin word-final vowels) were devoiced (NOVE- “nine” > Fr. *neuf*, GRANDE- “large” > Fr. *grant*). In French, all final consonants subsequently began to disappear around the 13th century (cf. *clef* [ˈkle] “key”, *porter* [poʁˈte] “to carry”), though – in some cases – they were later restored (*neuf*, *chef* [-f], *finir* [-ʀ]).

More complex is the evolution of word-final [-s], which was maintained in most areas of the *Romània*, with the exception of French (as we have just seen), Italo-Romance and Romanian, where it sometimes developed into a palatal element (Lat. NOS > It. Rom.n *noi*; cf. 8.3.2 no. 6.3 for the plurals ending in [-i]). There are still exceptions, such as the [-s] in 2sg forms, which was retained in certain Italo-Romance dialects (e.g.

Venetian or in the border area between Basilicata and Calabria [the latter including 2pl forms]).

4 Simplification of consonant clusters

Like diphthongs, consonant clusters show little stability in the Romance languages, with processes of assimilation beginning as early as the Late Latin period:

DORSU “back” > DOSSU > It. *dosso* “rise, bump” [alongside the learned form *dorso* “back”], Sard. *dossu* [-ss-], Cat. Occ. Fr. *dos* [> s, in Fr. > Ø]

MENSA “table” > MESA > Sp. Pg. *mesa* [-s-], Rom.n *masă*, O.Fr. *moise* [> -s- > -z-], Dalm. *maisa*

The affricate consonants typical of Old French [ts, tʃ, dz, dʒ] were simplified and became fricatives (or sibilants, to be more precise) from the 12th to 13th centuries onwards:

CAMERA “vault, arched room” > O.Fr. [tʃãbrə] > O.Fr. [ʃãbrə]

Some secondary consonant clusters resulted in the appearance of transitional consonants (phenomena of epenthesis):

VOLERE + HA(BET) “want.INF” + “have. 3SG.PRES.IND” > *VOL’RA > O.Sp. *volrá*, Cat. *voldrà*, Fr. *voudra* vs. It. *vorrà* (with assimilation)

GENERU “son-in-law” > GEN’RU > Fr. Cat. *gendre*, Occ. *gen(d)re*, Frpr. *gendro*, Rom. (Engad.) *gender* vs. Pg. *genro*, Sp. *verno* vs. (without syncope) It. *genero*, vs. Sard. *bénneru*, Sic. *iennaru*, Friul. *dzinar*; cf. 6.5.3

HOMINE “man” > *HOM’NE > Sp. *hombr̥e* (with dissimilation *m’n* > *m’r*) vs. Fr. *homme* (with assimilation) vs. Pg. *homem* (without syncope)

SIMULARE “imitate.INF” > *SEM’LAR > Cat. *semblar*, Fr. *sembler*

5 Specific phenomena

Loss of [h], already in Classical Latin:

HABERE “have.INF” > It. *avere*, Sp. *haber* [a’βer], Fr. *avoir*

Aspirated *h* was reintroduced in O.Fr. in borrowings from Germanic languages: *hardjan* > *hardi*, *hatjan* > *haïr* “hate.INF” (with aspirated *h*); cf. also later borrowings in Rom.n, such as *haină* (< Bulgarian/Serbian).

Velarisation/vocalisation of implosive [ɓ]:

ALTERU “other” > Sp. *otro*, Pg. *outro*, Fr. *autre*

Metathesis of [r]:

*FORMATICU “shaped” > Fr. *fromage* “cheese”; but: It. *formaggio*, Cat. *formatge*

GENERU “son-in-law” > Sp. *yernu*, cf. above

Loss of intervocalic [n, l] in Pg.:

MANU > *mão* (nasalisation of vowel)

LUNA > *lua*, MALU- > *mau*

Aspiration of [f] in Spanish and Gascon (followed by complete loss in Spanish):

FILIUS “son” > Sp. *hijo*

FACTUS “deed” > Sp. *hecho*, Gasc. *hèt*

6.5.5 Patterns of consonantal evolution from Latin to French

As an example of the phonological evolution between Latin and a given Romance language, an overview of the thirty-two consonants of Latin and of (Old and Modern) French is provided below in the form of a table:

- consonants noted in regular type existed in Latin and Old French, and are still present in Modern French: /p b t d k g f w s m n l j/ (= thirteen phonemes of the total thirty-two have remained stable over a period of two millennia)
- consonants noted in italics were present only in Latin: /ŋ k^w g^w h/ (four phonemes); /r/, in italics on a grey background, was found in Old French as well (= five Latin consonantal phonemes out of eighteen disappear)
- consonants represented on a grey background first appeared in Old French; some, underlined, have survived into Modern French: /z ʒ ɲ ɥ v/ (six phonemes). Others, not underlined, have disappeared: /θ ð ts dz ʧ ʤ r:/ (seven phonemes)
- the underlined phoneme /ʀ/, originally an allophonic realisation of /r/, is an innovation of the modern period.

In total, then, thirteen new phonemes were introduced, half of which disappeared again after some centuries. The loss of five Latin consonants is counterbalanced by the emergence of seven French phonemes; hence the total of twenty consonantal phonemes (cf. 6.2.3 no. 1).

	bilabial		labiodental		dental		alveolar		palato-alveolar		palatal		velar		uvular		labiovelar		glottal	
voiced	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
plosives	p	b					t	d					k	g			k ^w	g ^w		
fricatives			f	v	θ	ð	s	z	ʃ	ʒ									h	
affricates							ts	dz	tʃ	dʒ										
nasals		m						n				ɲ		ŋ						
laterals								l												
trills								r												
								r:								R				
glides											j		w							
											ɥ									

The table emphasises the significance of the phonological transformations witnessed by French over the centuries, and illustrates the use of distinctive features to categorise such changes: the strong presence of fricatives and affricates in Old French is a remarkable typological characteristic which distinguishes this language from both Latin and Modern French. In the same vein, although the voicing of plosives involved at least the six phonemes /p t k b d g/, it should nevertheless be viewed as a single evolutionary process.

6.5.6 General observations

There are many different phenomena of phonetic and phonological evolution, some of which have not been mentioned in the present chapter, despite their importance (such as, for example, the restructuring of Spanish affricates and fricatives in the 16th century). Only the most general phenomena, which convey a basic idea of the importance of the changes separating Latin from its various daughter languages are included here.

The phonetic changes described above presuppose continuity with regard to the words in use from one generation to the next over a period of two millennia. Considering that the evolution of a language is constantly affected by contact between different groups of humans, it is not surprising that external factors such as migration, social revolution and the interaction between speech and writing give rise to numerous disruptions in such evolutionary patterns (*cf.* 6.6).

The course of an individual evolutionary process can be altered not only by interference with other processes of evolution, but also by the fact that any word may at

any time deviate from regular, expected patterns of evolution, as a result of analogy or under the influence of Latin words (considering that Latin was until fairly recently the most important prestige language throughout the *Romània*). Consequently, (almost) any Romance word could be subject to ‘re-Latinisation’ – affecting both its phonetic and its graphic form – at any given time.

6.6 Latinisation and re-Latinisation

The importance of written Latin for the history of the Romance languages has already been mentioned several times. As we have seen, the direct ancestor of the Romance languages was not the classical, written language, but a spoken form of Late Latin. Nevertheless, during almost every period of their history, the written and standardised Romance languages incorporated elements of written Latin (in its classical, as well as its medieval and modern forms). Such elements will be referred to here as ‘Latinisms’, ‘learnèd words’ or ‘learnèd forms’. It is essential, though, to keep in mind that, until quite recently, the literate elite of society was versed in Latin (*cf.* 10.4.3 no. 2).

Though at first the influence of Latin was limited to the written texts of a given Romance language, it would later have repercussions on its entire language system. This can be explained by the fact that spoken standard language tends to imitate writing; in a second step, elaborated speech can serve as a model for spontaneous speech.

The impact of Latin throughout the *Romània* affected all domains of language: writing and pronunciation, morphology, morpho-syntax and syntax, and – as far as lexis is concerned – both the semantic properties of words and the mechanisms of derivation and composition. Phenomena of re-Latinisation that simultaneously touch upon several domains can lead to changes on a macroscopic scale: thus, texts written in Old French in the 13th and 14th centuries were closer to Latin than those written in the 12th century.

Today, the frequency of some types of Latinising derivatives with an international distribution in technical languages leads to convergence among different languages, Romance and non-Romance alike (*cf.* the idea of a ‘Euro-morphology’ discussed by Schmitt, 1996).

Latinisms of this type often represent direct borrowings from Latin words for which there is no corresponding Romance form in use; in other words, (grapho-)phonetic Latinisms simultaneously have the status of lexical Latinisms. Nevertheless, there are also cases of hereditary words that are influenced by their original Latin form. Consider the following example: in Old French the learnèd form *table* “table; etc.” coexists alongside the synonymous hereditary variant *taule* (which followed a ‘natural’, gradual course of evolution from Lat. *TABULA*); determining the precise nature of the interaction between the two variants is thus a complex process.

With regard to spelling, Latin elements may be introduced at any time into corresponding Romance words: e.g. O.Fr. *tenz* “time” and *doit* “finger” were re-Latinised

during the Renaissance as *temps* (after TEMPUS) and *doigt* (after DIGITUS). In these specific examples, the new, ‘learnèd’ spelling found its way into common usage; however, spontaneous Latinising variants that did not become established in the long run appeared during all historical periods of the Romance languages.

In some cases, moreover, Latinising spelling coincides with archaic spelling (i.e. graphic forms which perpetuate writing practices reflecting the phonetic stage of a language that is no longer current). The habit of writing word-final consonants that were no longer pronounced in French after the 13th and 14th centuries is an example of this phenomenon.

Hence, a multitude of words of ‘semi-learnèd’ evolution bear witness to interferences between Latin and Romance in their spelling or pronunciation. The French adjective *stable*, for instance, is a Latinism that remains relatively close to its etymon STABILE “stable, firm” (the only change having affected the word being the syncope of unstressed [i]); *étable* (< STABULU “abode, stable”) displays an additional modification (the appearance of epenthetic [e] before *s-impurum*). In contrast, the inherited form *estaule* is attested in Old French for both STABILE and STABULU, but was later abandoned.

Semantic or formal borrowing can also lead to etymological doublets consisting of an inherited form of an etymon and a corresponding learned borrowing, the two normally displaying semantic divergence. The phenomenon is particularly apparent in French, due to the intense phonetic changes that the language underwent in the past, as illustrated below:

cause “cause” / *chose* “thing” (< CAUSA)
captif “captive” / *chétif* “sickly” (< CAPTIVUS)
loyal “loyal” / *légal* “legal” (< LEGALIS)
forge “smith” / *fabrique* “factory” (< FABRICA)
hôtel “hotel” / *hôpital* “hospital” (< HOSPITALIS)
table “table” / *tôle* “sheet metal” (< TABULA)

Moreover, in the modern Romance languages, Latinisms often appear as derivatives within word families of inherited base words:

père “father” → *paternel* “paternal” (< PATERNALIS)
évêque “bishop” → *épiscopal* “episcopal” (< EPISCOPALIS; cf. It. *vescovo* → *vescovile* ~ *episcopale*, Sp. *obispo* → *obispal* ~ *episcopal*)
eau “water” → *aqueux* “watery, water-based” (after AQUA)

6.7 General tendencies of sound change

The phonological make-up of a language is the starting point for every change. The theory of ‘empty categories’ in the vein of ‘natural phonology’ even assumes that when a given phonological slot is vacated (e.g. the voiced dental fricative /z/), it will be newly occupied following the displacement of another sound. More realistically, structural interpretations assume that phonological oppositions that only involve a small number of words tend to disappear (the classic example for modern French being the weakened opposition between *brin* “wisp, sprig” and *brun* “brown” (/ɛ/ ~ /œ/, cf. 6.2.3).

Considering that such changes originate at the level of speech (*parole*) – and not at the level of the language system (*langue*) –, it is nevertheless more likely that variants arise in a contingent, and therefore random manner. The phonological system of a given language during a specific historical period is the basis from which variants emerge; this system, however, does not determine phonetic changes. Among the different languages of the world, the number of phonemes varies between 11 and 141; this fact alone illustrates the instability inherent in phonetic evolution.

This does not preclude the possibility of recognising certain patterns of sound change on a macroscopic level. The significance of diphthongisation, palatalisation and assibilation with regard to changes between Latin and its daughter languages has been noted above. From a typological point of view, these are a characteristic of the Romance languages for which no global theoretical interpretation has been provided to date.

One feature of phonetic change that can nevertheless be considered universal is the tendency to shorten words. In the *România*, this phenomenon is particularly pronounced in French and Portuguese – the two Romance languages displaying the greatest divergence from Latin. On average, French also displays the shortest words (phonetically speaking). As will be observed in the following chapter, one of the effects of morphological derivation is to restore a sufficient phonetic volume in words that have undergone drastic reduction (cf. 9.4.1).

7 Inflectional morphology

7.1 The subject matter of morphology

7.1.1 Inflection vs. word formation

Morphology is concerned with the study of the internal structure of words (cf. Greek *morphē* “form, shape”), including both word inflection and word formation.

By means of inflection, different word-forms of a lexeme (its ‘inflected forms’) can be inserted into actual phrases and sentences. Consider the following example illustrating inflected forms based on the stems of the French lexemes AMI [a'mi] “friend”, PARL(ER) [paʁl(e)] “(to) speak” and FORT [fɔʁ] “strong” (lexemes will appear in small capitals):

inflected form	phonetic realisation	syntactic contextualisation
<i>ami-s</i>	[a'mi] + Ø	<i>les amis de mes amis sont mes amis</i> “the friends of my friends are my friends”
<i>parl-ons</i>	[paʁl] + [ɔ]	<i>Parlons !</i> “Let’s speak!”
<i>fort-e</i>	[fɔʁ] + [tə]	<i>une forte impression</i> “a strong impression”

Word formation enables the formation of complex lexemes from simple lexemes, as illustrated by the French lexemes REMETTRE “(to) put back”, GRAND-MÈRE “grandmother” and AIGRE-DOUX “sweet and sour”:

derivation:	<i>remettre</i>	[ʁə] + [mɛtʁ]
compound (ADJ – N):	<i>grand-mère</i>	[grɑ̃] + [mɛʁ]
compound (ADJ – ADJ):	<i>aigre-doux</i>	[ɛgr] + ([ə]) + [du]

Whereas inflection both occurs and acquires its meaning within a syntactic context, word formation occurs within lexis and will thus be discussed in the relevant chapter (9.4). At the same time, both inflection and word formation deal with the structure as well as with formal aspects of words or lexemes.

Despite a long tradition since the beginnings of linguistics, morphology was ousted from its place among the domains of language by generative grammar (Chomsky 1957) in accordance with the theory that inflection and word formation are entirely dependent on syntax. Recent approaches, however, acknowledge the complexity of this formal dimension of language and recognise its autonomy. As a result, morphology has experienced a revival over the past three decades and can thus be considered as “la discipline de la linguistique qui est à la fois la plus ancienne et la plus récente” (‘the oldest and, at the same time, the most recent linguistic discipline’) (Bernard Fradin *et al.*, *Aperçus de morphologie du français*, 2009: 5).

- Esher/O'Neill, *The autonomy of morphology*, CambrHandb 11
 Cappellaro/Meinschaefer, *Inflexion, derivation, compounding*, CambrHandb 13
 Ledgeway/Smith/Vincent (eds), *Periphrasis and Inflexion in Diachrony: A View from Romance*, 2022

7.1.2 Lexical and grammatical elements in morphology

Defining the concepts of 'word' and 'lexeme' requires caution. Here, we refer to a series of definitions that find their place in the line of research undertaken by Mark Aronoff (*Morphology by Itself*, 1994) and that are compatible with terminology stemming from other traditions (cf. Matthews, *Morphology*, 1974, or, specifically for French, Fradin *et al.*, *Aperçu de morphologie du français*, 2009). It is important to understand that the necessary starting point – the 'word' or rather, the 'lexeme' – cannot be defined in a purely formal manner, that is, without reference to semantic criteria (cf. Glessgen, *Le statut épistémologique du lexème*, 2011).

1 Lexical vs. grammatical word classes

Word forms are organised morphologically in word classes, or parts of speech, which vary from language to language and which may also evolve from one historical stage of a language to another. Nevertheless, all languages distinguish between words that have a lexical function and those that have a grammatical function, as well as between nouns and verbs, which represent the most salient word classes:

- a lexeme is a lexical entity comprising a formal element and a semantic element that refers to a concept (e.g. 'CAT', 'DOG', 'BEAUTY'; cf. 9.1.1)
- a function word is a grammatical entity which also comprises a formal and semantic element, but which expresses an essentially grammatical relation (e.g. the French function word *et* "and" expresses the coordination of two syntactic units of equal status); personal pronouns (*je* "I") or demonstrative pronouns (*ceci* "this") have an additional deictic function (i.e. whose referential meaning is only specified within a given context); a function word is a type of grammeme, parallel to 'bound' grammatical forms which are attached to lexemes (cf. no. 3, *Free vs. bound*, below)
- it is important to note that the number of different function words is very limited in all languages, but that on an individual level they display a high frequency of use.

In the Romance languages, grammatical tradition posits four categories of lexemes (nouns, verbs, adjectives and adverbs) and four categories of function words (prepositions, pronouns, determiners and conjunctions) in addition to interjections such as Fr. *bof!*, It. *ahimè!*, Sp. *¡olé!* which sometimes act as sentences in their own right, and are considered to be a distinct part of speech. Delimitation between these different categories is nevertheless not always clear-cut: uncertainties arise in the distinction of past

participle forms of verbs from the adjectives based on these participles (e.g. Fr. *la porte est ouverte* “the door is open”; *les vêtements usés* “old/worn clothes”) or of adverbial expressions from prepositional expressions (*le prix est indiqué au-dessus* “the price is shown above” vs. *il est monté au-dessus de la colline* “he climbed to the top of the hill”). Pronouns and determiners display a considerable amount of overlap (cf. 7.5), particularly in the medieval Romance languages.

2 Stem vs. grammatical affix

The concepts of stem, conceived as a ‘stem space’ (and thus allowing intrinsic variation), and of grammatical affix can be illustrated by the example of the inflectional forms of the French verb *pêcher* (“to fish”): (*je*) *pêche* ([’pɛʃ] 1PS.SG.PRES.IND. “(I) fish”) is homophonous with the noun (*la*) *pêche* ([’pɛʃ] FEM.SG. “(the) fishing”), but if we consider the plural form (*nous*) *pêch-ONS* ([pɛʃ-ɔ̃] 1PL.PRES.IND.), a slight phonological alternation [ɛ vs. e] within the stem space can be observed.

The stem or ‘stem space’ displays the following characteristics:

- it retains a certain formal and semantic unity: e.g. Fr. *jou-eur* “sb. who plays (an instrument or a sport)” / “gambler”, *jou-et* “toy”, *jou-able* “doable”, *jou-jou* (reduplication) “toy”
- it can, however, include allomorphs, which may be determined by phonetic context (e.g. [’spɔʁ] <sport> “sport” vs. [spɔʁ’t-if] <sport-if> “who plays sport regularly” / “pertaining to sport”), by grammatical parameters (e.g. gender in [’bo] <beau> m. vs. [’bɛl] <belle> f.) or simply by the accident of phonetic change, as in the above-mentioned example ([’pɛʃ] ~ [pɛʃɔ̃])
- allomorphy may take the form of rather pronounced stem alternation: e.g. Fr. *jou-* + *-er* [ʒu-’e] (“(to) play”) and *jeu* [ʒø] (“game”).

Affixes have a grammatical character since they fulfil a relational function with regard to the stem. The same derivational affix can be found in a certain number of words with similar grammatical and semantic properties (ex. Fr. *jou-eur* “sb. who plays (...)”, *chant-eur* “(professional) singer”, *act-eur* “(professional) actor”, *coiff-eur* “(hairdresser)”):

- they are classified according to their position (prefix: Fr. *re-mettre*, suffix: *réalis-able*, infix: Lat. *florēre* “to bloom” vs. *florescere* “to begin to bloom”) and their function (inflectional vs. derivational)
- they can be inflectional (ex. Fr. *ami-s*, It. *amic-i*, Sp. *amigo-s*, Fr. *parl-ons*, It. *parl-iamo*, Sp. *habla-mos*) or derivational (ex. Fr. *chant-eur*, It. *canta-tore*, Sp. *canta-dor*, Fr. *dés-agréable*).

Inflectional affixes express different morphosyntactic properties (gender, tense, mode, etc., cf. 7.3), whereas affixes marking a part of speech (noun vs. verb) are considered to be derivational (cf. 9.4).

3 Free vs. bound (lexemes and grammemes)

A lexeme or a grammeme can occur as a free unit in sentences as an inflected word form or a function word (e.g. Fr. *mourir* “die.INF”, *mourons* 1PS.PL.PRES.IND., *meurt* 3PS.SG.PRES.IND., which correspond to two allomorphic stem forms of the lexeme MOURIR: /mur/ and /mœr/, or the free grammatical forms Fr. *donc* “so”, *quand* “when”, It. *dunque*, *quando*, Sp. *pues*, *cuando*).

A lexeme or, more frequently, a grammatical form, can also be bound and attached to another form (for example, clitics such as It. *-mi* in *dimmi* “tell me”, Fr. *j(e)* in *j’attends* “I’m waiting”, or the English possessive ‘s always occur fused with a lexical base); hence, affixes are by definition bound units.

Certain morphological stems are never realised as inflected forms and are only found in combination with affixes or as part of compounds (e.g. Fr. *somat-ique*, as well as a number of verb stem forms such as *mour-ons*, *pren-ons*).

Recent research in morphology tends to refer to the functional concepts of ‘lexeme’ and ‘grammeme’ as a terminological basis (cf. Aronoff, *op. cit.*, 1994: 9). Traditional structuralist morphology, on the other hand, was originally founded on the idea of segmenting linguistic forms to a maximum, and relies on the concepts of ‘morphs’ and ‘morphemes’ as a starting point¹⁷.

17 ‘Morphemes’ are defined as the abstract entity that reflects the smallest meaningful unit of a language obtained when segmenting a word to the maximum. Morphemes cannot be divided into smaller analysable units (e.g. {in-} {-flex-} {-ible}); in actual utterances, however, they appear in the form of different allomorphs (cf. 7.2.3 no. 1). The actual allomorphic units are defined as ‘morphs’. A ‘morph’ can also be ‘empty’ (i.e. an element devoid of meaning which nevertheless fulfills a morphological function). In the Romance languages, this concerns in particular ‘thematic vowels’ marking different inflectional classes: e.g. the [i] in [dɔʁ’m]+[i]+[R] or in {-ible}.

The concept of morphs and morphemes also implies that of ‘zero’ morphemes, which consist in either the absence or the omission of specific morphemes. This amounts to an implicit, more economical type of marking (in modern Romance languages, for example, the subject is usually zero-marked, whereas the first and second object may be marked: e.g. Sp. *Pablo*_s *ve a Juana*_{o1} “Paul sees Joan”, Fr. *Paul*_s *donne le livre à Jeanne*_{o2} “Paul gives the book to Joan”; consider also the Latin imperative forms *fer* “carry” and *duc* “lead” (2sg) which display zero marking, in contrast to the corresponding indicative forms of the same verbs, *fers*, *ducis* (2sg). A more restrictive conception of the zero morpheme is its application to cases involving omissions, as opposed to the simple absence of markers (e.g. *il est entouré d’ennemis* “he is surrounded by enemies”, where the ‘zero indefinite plural’ article *de* is employed instead of *des*).

7.2 Inflection in Latin and in Romance

7.2.1 ‘Morphosyntactic’ features

The function of inflection is to express a series of grammatical categories and features, which are required in a given language. Typically, nominal inflection or ‘declension’ (for nouns, adjectives, pronouns, articles and determiners) is distinguished from verbal inflection or ‘conjugation’. The inflected forms of lexemes thus fulfil the requirements of phrases and sentences. This implies that the expression of inflection is obligatory both in Latin and in the Romance languages (a noun or verb cannot *not* be inflected).

Inflection involves different properties known as ‘morphosyntactic’ features. The repertoire of obligatory morphosyntactic features varies from language to language and according to the word class to which a lexeme belongs:

- inherent features of inflected forms: for nouns, the expression of number (singular or plural), is merely a result of the speaker’s intention; for verbs, the same goes for person, number and tense (or aspect, as in the Semitic languages);
- contextual features: the inflected forms of Romance adjectives, participles and determiners are determined by other elements in their syntactic context, according to the principle of agreement (or concord) (e.g. the possessive adjective in Lat. *fratrem meum* ‘my brother’ is in agreement with the noun in terms of gender and number; here, *fratrem* is the ‘controller’, and *meum*, the ‘target’).

Agreement is to be distinguished from ‘rektion’ (or ‘government’), which involves a controlling element that does not display the same value as the controlled form (e.g. Lat. *cum fratre meo* ‘with my brother’: here, *cum* governs the case of the noun *fratre* [with which the adjective *meo* is in agreement]; in It. *da me* ‘at my house’ vs. **da io* ‘at + I’, *da* governs an object pronoun, not a subject pronoun).

The above features differ from lexeme-inherent features that are not of inflectional but of derivational nature, or they are lexically determined:

- gender is an inherent and, in principle, a stable feature of nouns; this feature is not conditioned by the syntactic context but rather linked to a given lexeme;
- voice (also known as diathesis) is in some cases an inherent feature of verbs (such as the Latin deponent verbs, which are only conjugated in the passive voice but which have an active meaning).

Inherent, contextual and lexeme-inherent features constitute the main morphosyntactic features which are of particular significance for Latin and the Romance languages. In the following sections, we will examine the remaining phenomena relating to inflectional morphology that are characteristic of Romance. These are: degree of inflected-

ness [inflection vs. agglutination], noun inflection, adjective comparison, the pronoun and determiner system ['D-system'] and verb paradigms.

→ Vincent, *A structural comparison of Latin and Romance*, OxfGuide 4

Ledgeway/Maiden, *Data, theory, and explanation: the view from Romance*, CambrHandb 1

Pescarini/Loporcaro, *Variation in Romance*, CambrHandb 4

7.2.2 Inflectional classes

Within the different word classes, there are groups of lexemes that display identical inflectional patterns and that form 'inflectional classes', which correspond to the traditional concept of inflectional 'paradigms' (an inflectional class or paradigm encompasses all inflected forms of a lexeme; consequently, these forms all share the same lexical meaning).

Inflectional classes vary from one language to another, as well as over time and within the diasystem of an individual language. Thus, the four conjugation classes of Latin verbs (*a'māre*, *te'nēre*, *'legēre*, *ve'nīre*) were neutralised to some extent in the Romance languages, particularly the distinction between the 2nd and 3rd classes, -ĒRE and -ĪRE. The most pronounced difference is still to be found between the 1st class (-ĀRE) and the others, the 1st being the most productive and the most regular (except in the case of Romanian).

In contrast, the five Latin declension classes underwent extensive reduction during their passage to Romance, by way of a complex process that will be described in more detail below (cf. 7.3.1). The presence of different classes remains pertinent above all for the distinction of number and gender, both for nouns (It. *lupo* M.SG – *lupi* M.PL "wolf – wolves" vs. *donna* F.SG – *donne* F.PL "woman – women") and for adjectives (It. *rosso* M.SG / *rossa* F.SG – *rossi* M.PL / *rosse* F.PL "red", Fr. *beau* M.SG / *belle* F.SG – *beaux* M.PL / *belles* F.PL); cf. 7.5 for pronouns.

The various morphosyntactic features and inflectional classes present in Romance will be discussed in the appropriate sections: thus, gender, number, case and declension classes as well as person, number and verb paradigms will be dealt with in the current chapter. Tense, aspect, mood (= TAM) and voice, on the other hand, will be examined in the chapter on syntax (cf. 8.3.2 nos. 1–3), as the interaction of these features with aspects of syntax is of particular interest in the Romance languages.

→ Maiden, *Inflectional morphology*, OxfGuide 27.8

Cardinaletti/Giusti, *Dependency, licensing, and the nature of grammatical relations*, CambrHandb 20

7.2.3 Inflection and agglutination from Latin to Romance

When studying inflection it is also necessary to consider the phenomenon of agglutination, i.e. the expression of morphosyntactic features through the juxtaposition of grammemes and lexemes as opposed to the combination of thematic stems and affixes (Bossong, LangTyp, art. 48: 665). The concept of agglutinative marking follows in the same vein as that of individual morphemes carrying different functions. In its most extreme form, it implies a separate grammeme for each function, which equates to true morphematic isomorphism (i.e. “one meaning > one form”, and vice versa, cf. Aronoff, 1994: 8sq.). Inflection, by contrast, makes efficient use of several principles of asymmetry with regard to the relationships between form and meaning. The difference between inflection and agglutination is illustrated by the following example:

- In Latin *cas-ae* “hut” F.SG.GEN., the suffix *-ae* simultaneously marks number (singular), case (genitive) and gender (feminine). It is therefore an inflectional marker which cumulates three functions.
- In French *à la maison* “in the house; at home” (Sp. *en casa*, It. *a/in casa*) the preposition *à* marks the type of constituent (a function performed by case in Latin, cf. 7.3.4), while the article *la* marks gender (feminine) and number (singular). The marking is thus partially agglutinative – only partially, however, since the article *la* still encodes two features, like the suffix *-ae*.

The example further shows that the degree of inflection underwent extensive reduction during the transition from Latin to Romance, at least in the nominal domain. As far as morphological asymmetry is concerned, it includes cases in which an identical meaning is expressed by different forms (allomorphy, fusion, suppletion, no. 1 below) or in which an identical form expresses multiple meanings (cumulateness, syncretism, nos. 2 and 3 below). In morphology as well as lexis, therefore, the relationship between meaning and form is arbitrary (cf. 1.2). While allomorphy is present in both Latin and Romance, albeit in differing form, the effects of cumulation and syncretism are considerably reduced in the latter.

1 Allomorphy, fusion and suppletion

Allomorphy involves *different forms* that express the *same meaning*. The formal variation can be an effect of the phonological (or graphemic) context or it can be inherent to the paradigm of an individual lexeme (‘lexically determined variation’). Allomorphy is generally an extremely frequent feature of inflection, not only in Latin, but also in the various Romance languages (cf. the numerous examples given in OxfGuide 5, 8 sq., 11–16, 22, 24, 27, 42–44).

Contextual allomorphy involves

- inflectional affixes: Lat. M.GEN.SG *domin-i* “lord” vs. *reg-is* “ruler”, Fr. PL *grand-s* vs. *beau-x* (the latter only apparent in writing), Sp. PL *casa-s* vs. *pec-es* [SG *pez*] and Jesi (central Italy) [kan^ot-a:mo] “we sing” vs. [go^od-e:mo] “we enjoy ourselves” (Loporcaro/Paciaroni, OxfGuide 15)
- derivational affixes: Fr. *impénétrable* [ɛ̃] vs. *inhumain* [in]
- word stems: It. *esco* [ˈesk-o] “I leave” vs. *esci* [ˈɛʃʃ-i] “you leave” vs. *uscire* [uʃʃ-ˈiɾe] INF (Ledgeway, OxfGuide 14), Sp. *siento* “I feel” vs. *sentimos* “we feel” vs. *sintamos* SBJV (Tuten/Pato/Schwarzwald, OxfGuide 22), Fr. NUM: *cheval* [ʃəˈval] (‘horse’) vs. *chevaux* [ʃəˈvo] ‘horses’ as opposed to the dominant agglutinative principle illustrated by *maison* vs. *maison-s*. Here, the variation is conditioned by phonetic history, as the plural form *chevaux* preserves the trace of the vocalisation of *-l-* before word-final inflectional *-s*, which took place in O.Fr.: *chevals* [ʃəˈvals] ~ *chevaux* [ʃəˈvaʊs])

Lexically determined variation mostly involves word stems, as in:

Lat. [case]: *iter* vs. *itineris* “journey, path” and *imperator* vs. *imperatorum* “emperor” (the latter example involving a difference in the position of the accent)

or derivational affixes, as in:

Fr. *inflexible* vs. *réalisable* vs. *insoluble*

The variation of the thematic stem (‘stem alternation’ or ‘root-allomorphy’) constitutes a particular type of allomorphy known as ‘fusion’. It is often considered to be typical of inflection, as opposed to agglutination. While fusion is not particularly frequent in Latin, it nonetheless affects the whole of the 3rd declension. In Romance standard languages, this special feature of inflection mostly occurs in verbal paradigms¹⁸.

The most extreme form of allomorphy is ‘suppletion’, which involves the presence of different word stems within the same inflectional paradigm. It affects a small number of the most frequent lexemes, verbs in particular. Lat. ESSE “to be” (*sum, est, erat ~ erit, fuit ~ fuerit*) and its Romance equivalents display the highest number of stems:

Fr. ÊTRE (*suis ~ sommes, est, fut, sera*, and, as a result of interference of Lat. STARE “to stand”: *été ~ était*)

¹⁸ Fusion features heavily in the Germanic languages, where it is displayed by phenomena such as vowel alternation or ‘Ablaut’: German *Haus* [haʊs] “house” / *Häuser* [ˈhɔɪzɐ] “houses”, English *man* [mæn] / *men* [mɛn], *woman* [ˈwʊmən] / *women* [ˈwɪmɪn]. Moreover, it is an integral feature of the morphological system of other languages such as Arabic (*yaktubu* “he writes”; *kataba* “he wrote”; *maktub* “written”. Morphologised metaphony is also present in the noun system of many Romance varieties: cf. for example *gal* [gal] “rooster” vs. *gel* [gel] “roosters”, etc. in northern Italian dialects.

It. ESSERE (*sono ~ siamo, è, era, fu, sarà; stato*)

Sp. SER (*soy, es, era, fue, será, sido*)

Suppletion also affects function words:

Fr. *lui* PRON.M.SG vs. *eux* PL, It. *lui* PRON.M.SG / *lei* F.PL vs. *loro* M./F.PL (Maiden, OxfGuide 42; it should be noted that the synchronic suppletion in this case is purely an effect of phonetic change; diachronically, all these forms involve the paradigm of Lat. ILLE)

2 Cumulativeness

The opposite of allomorphy is cumulativeness (or ‘cumulation of functions’), where a *single form* fulfils *multiple grammatical functions* at the same time (we sometimes speak of ‘cumulative morphemes’). This phenomenon is frequent in morphosyntactic features relating to inflection and it is particularly developed in Latin, in contrast to the less inflectional Romance languages.

Latin:

[case] + [gender] + [number]: *domin-ī* [= NOM.M.PL], *magis ardu-um* “harder” [= ACC.M.SG], *cas-ae* [= GEN.F.SG]

[person] + [number] + [tense] + [mode] + [voice (ACT)]: *cant-o* [= PRES.IND.1SG]

Italian:

anzian-o “elderly” [= M + SG], *anzian-a* [= F + SG], *anzian-i* [= M + PL], *anzian-e* [= F + PL] (Ledgeway, OxfGuide 14)

canta-v-ano “they were singing” [= PAST (TENSE), IPFV (ASPECT) and IND (MOOD)] (*ibid.*)

The reduction of cumulativeness reflects the major typological difference between Latin, which is more inflectional, and the Romance languages, which are more agglutinative.

3 Syncretism

A third functional principle of inflection is syncretism, where an *identical form* expresses *different grammatical functions* in different contexts. This can be seen as a particular type of ‘polysemy’ (cf. 9.2.2 no. 7 (8)). Syncretism is widespread in Latin and is also present in the Romance languages, albeit to a lesser degree:

in Lat. *domin-i*, the affix *-i* may mark two different sets of morphosyntactic features, namely GEN.M.SG OR NOM.M.PL

in Sp. *cantab-a*, the affix *-a* marks either 1SG or 3SG IMPF.IND (Tuten/Pato/Schwarzwald, OxfGuide 22); It. *giovan-e* is either M.SG or F.SG, *giovan-i* M.PL or F.PL (Ledgeway, OxfGuide 14)

Here too, a reduction of inflectional features may be observed during the passage from Latin to Romance.

The examples provided show that inflection and agglutination are not diametrically opposed to one another; rather, they represent two extreme positions on a continuum. The presence or partial absence of factors such as stem allomorphy or cumulativeness determines the respective degree of inflection (or ‘inflectedness’).

The distinction between inflection and agglutination may be complicated by the phonetic and/or orthographic integratedness of markers. Depending on the theoretical approach, a phonetically integrated (but graphically separated) form, such as that in Fr. *j’ouvre* (vs. It. *apro*) can be considered either as inflectional or agglutinating; to be more precise, the apostrophe was not used in medieval written Romance languages, appearing only in the 16th century (cf. O.Fr. *jaime* for *j’aime* ‘I love’, *lost* for *l’ost* ‘the army’, *diver* for *d’hiver* ‘of winter’).

This also shows that morphological structure cannot be reduced to a concatenation of minimal units of morphological meaning, which explains why recent approaches in morphology are based on the notions of lexemes and grammatical markers, rather than morphemes (cf. 7.1.2). In the ‘Word-and-Paradigm’ model developed by Mark Aronoff (*Morphology by Itself*, 1994), inflected forms are considered as a whole and are categorised according to their morphosyntactic (and morphosemantic) features¹⁹.

As a general rule, we have seen that Latin displays a high degree of inflectedness, particularly in terms of noun marking²⁰. The transition from Latin to Romance brought with it a reduction in inflection, in favour of agglutinating elements, especially in nominal morphology; Romance verbs, on the other hand, remain highly inflectional. In the following section (7.3), we will thus focus on particularly marked changes in nominal inflection between Latin and the Romance languages.

The observations on inflection and agglutination above serve to clarify a traditional interpretation that goes back to August Wilhelm von Schlegel’s *Observations sur la langue et la littérature provençales*, one of the first treatises on Romance linguistics, published in 1818. According to Schlegel, Latin was a ‘synthetic’ language (i.e. a language in which morphosyntactic features are expressed by means of affixes, which are combined with word stems to form individual graphic words such as Lat. *fratr-is* ‘belonging

19 The ‘word-and-paradigm’ approach replaces the traditional categorisation of paradigmatic forms according to roots, morphs and morphemes, replacing it with the distinction between ‘stem space’ (stem) and inflectional affixes, which are of an inherently variable nature (cf. above). The method of describing the data does not differ greatly between the two approaches; nevertheless, the new model better accounts for the complex relationship between morphosyntactic features and morphological form, as well as for the resulting phenomena of asymmetry (allomorphy, syncretism, cf. no. 1 above), thus, allowing a more accurate explanation of the influence of analogy in verb paradigms (cf. the examples in 7.6.1).

20 Within the domain of verb inflection, there are a number of ‘agglutinating islands’ such as the paradigm of the past imperfect, which has separate affixes for tense (*ama-ba-m*) and person (*ama-ba-m* / *-s* / *-t* etc.); however, other verb forms such as the perfect of irregular verbs, are highly inflectional.

to/of the brother”), whereas the Romance languages are ‘analytical’ (i.e. languages in which grammatical features are expressed by grammemes, which are graphically separated from their lexical base, as in Fr. *du frère* and It. *del fratello*). Today, it is preferable to say that Latin displays a high degree of inflectedness, whereas the Romance languages tend more towards agglutination, a tendency which is more explicit in nominal than in verbal morphology.

→ Maiden, *Inflectional morphology*, OxfGuide 27.1

Maiden/Thornton, *Suppletion*, CambrHandb 12;

Hinzelin, *Allomorphy and syncretism in the Romance languages*, OxfEnc; cf. 7.6.2, ‘morphemes’

7.3 The restructuring of the Latin nominal inflection system

Nominal inflection underwent marked changes during the transition from Latin to the Romance languages. The number of inflectional classes was reduced (no. 1), the system of number marking was reorganised (no. 2), the neuter gender was weakened or abandoned (no. 3) and, above all, the Latin case system was subject to intense reduction (no. 4).

In Latin as well as throughout the Romance area, nominal inflection is paralleled for the most part by adjectival inflection. Thus, nouns, as well as adjectives and pronouns, generally display the following morphosyntactic features (cf. 7.2.1 above), upon which the various inflectional classes are based:

- gender is a noun-inherent distinction which has three values M vs. F vs. N; grammatical gender is not to be confused with biological sex
- number has two values: SG vs. PL. However, it is necessary to further distinguish the following categories:
 - ‘count’ (or ‘countable’) nouns (e.g. Fr. *maison*), which typically occur in the sg. and pl.
 - ‘non-count’ (or ‘mass’) nouns (e.g. Fr. *eau*), which are usually sg. but which also occur in the pl. (cf. Fr. *les eaux du Pacifique* “the waters of the Pacific”, *les eaux usées* “waste water”)
 - the so-called *singularia tantum* which, theoretically, are countable but only occur in the singular (e.g. Fr. *amour*, *auditoire* “audience”)
 - the *pluralia tantum*, which occur exclusively in the plural (e.g. Fr. *Pâques* [generally followed by the verb in the singular], *Alpes*, *gens*)
- case is an inherent feature of Latin inflected forms (NOM, GEN, DAT, ACC, ABL, VOC), which has left various traces in the Romance languages.

7.3.1 Inflectional classes

The five inflectional classes of the Latin nominal system were reduced to three in the Romance languages, by way of a progressive general loss of the 4th and 5th declensions (*acus* – *acūs* “needle”, *diēs* – *diēt* “day, appointment”). Only the first (feminine) and second (masculine) classes survived in their entirety, while some elements of the 3rd declension (characterised by consonant stems and i-stems) were maintained and gradually integrated into the first two classes.

This process of change can be traced back to the period of Late Latin; as early as the 5th century, the glosses of the *Appendix Probi* (5th century) criticise declension ‘errors’ that bear witness to language change:

nurus (fem., 4th declension) *non nura* (fem., 1st declension) “daughter-in-law” (i.e. the grammarian is protesting against the non-normative use of the form *nura* as opposed to the classical form *nurus*, thereby proving the existence – and perhaps even a certain frequency of use – of the type *nura*)

tristis (masc. or fem., 3rd declension) *non tristus* (masc., 2nd declension) “sad” (here too, the mere mention of the masculine form *tristus* for the adjective *tristis* implies the existence of the former in spoken Latin)

The integration of nouns belonging to the Latin 3rd declension into the 1st and 2nd declensions is particularly advanced in French; meanwhile, most other Romance languages still contain a great number of reflexes of Latin 3rd declension nouns (e.g. Sp. *classe*, *joven*, *poder*, *dente*, *habitación*, *leche*, etc.):

Sp. *M can* “hound, mutt” (< *CANE*) – *F flor* “flower” (< *FLORE*), Fr. *M chien* “dog” – *F fleur*, Sursilv. *M tgaun* – *F flur*, It. *M cane* – *M fiore*, Vgl. *M kuj* – *M fjaur*, Ro. [= Rom.n] *M câine* ‘dog’ – *F floare* (cf. Maiden, *Inflectional morphology*, OxfGuide 27.8)

Adjectives preserve the trace of the Latin 3rd declension to a greater extent, in the form of the so-called ‘epicene adjectives’, which display identical masculine and feminine forms (in accordance with their Latin equivalents), and which have maintained their vitality in Modern Spanish and Italian:

Sp. *alegre* “cheerful”, *feliz* “happy”, It. *gentile* “kind”, *felice* “happy”

This inflectional subset also survived into Old French:

M and *F grant* “large”, *fort* “strong”, *vert* “green”, *gentil* “kind”, *loial* “loyal”, *solempnel* “solemn”, *vallant* “courageous”

By the 16th century at the latest, analogous feminine forms began to be introduced for these adjectives (*grand* – *grande*, *fort* – *forte*); the old feminine form nevertheless survives in terms such as *grand rue*, *grand-mère* and *vallamment*.

The drastic reduction of the Latin inflectional classes throughout the Romance area subsequently gave rise to a new distinction, essentially based on specific singular – plural combinations, which also involve the differentiation of gender (cf. 7.3.2 and 7.3.3 below). These constitute the only remaining functional parameter for the distinction of inflectional classes in the noun systems of modern Romance languages. Modern French thus distinguishes three main types of plural formation (*main* – *mains* “hand – hands”, *cheval* – *chevaux* “horse – horses”, *bras* – *bras* “arm – arms”, only the second of which is phonetically realised), and Italian exhibits seven different types, five of which are productive (*lupo* – *lupi* “wolf – wolves”, *rosa* – *rose* “rose – roses”, *fiore* – *fiori* “flower – flowers” and, to a lesser extent, *re* – *re* “king – kings” ~ *virtù* – *virtù* “virtue – virtues” and *poeta* – *poeti* “poet – poets”), while the remaining two are residual, but lexicalised (*uomo* – *uomini* “man/human being – men/human beings”, *braccio* – *braccia* “arm – arms”).

7.3.2 Number

The Latin distinction of number – SG vs. PL – was maintained throughout the whole of the Romance area, but was nevertheless subject to intense formal restructuring. The Latin sigmatic plural was almost generalised in Romance, where, however, it underwent significant phonetic change leading to a contrast between the remaining sigmatic forms (-s / -es / -(u)x) and vocalic plural suffixes -i / -e, the latter characteristic of Italian and Romanian (cf. Rohlfs’ map in fig. 9, which illustrates the geolinguistic distribution of dominant plural forms):

Fr. *chèvres* “goats” (phonetically realised only before a small subset of vowel-initial words such as auxiliaries [CambrHandb 6]), *chevaux* “horses”, Cat. *cavalls* “id.”, *filhes* “children”, Sp. *caballos* “horses”, *hijas* “daughters”, *peces* “fish”, Pg. *cavalos* “horses”, *lugares* “places”

vs.

It. *amici* – *amiche* “friends [M.PL – F.PL]”, Rom.n *socri* “fathers-in-law” – *capre* “goats”

In Old French, the sigmatic plural was already present in what could be called a ‘protoform’ stage, but word-final -s functioned simultaneously as a marker of SBJ.M.SG (thus representing a case of syncretism; cf. 7.3.4). The use of sigmatic plurals gradually became generalised in writing during the 14th century after the loss of the two-case system of inflection (cf. *ibid.*). Curiously, in most instances, word-final -s was no longer pronounced at this time (it began to disappear during the 13th century).

The case of Italian and Romanian is particularly interesting as the forms in -i partially survive, but in most cases they are a phonetic effect of the Latin sigmatic plural; cf. Maiden, *Number*, OxfGuide 42.3 [‘Plural’], who ‘argues that all Romance languages have a historically underlying ‘sigmatic’ system. According to this account, the modern

‘vocalic’ feminine plural ending *-e* results from a regular phonological development **-As* > **-aī* > *-e*. (...). Parallel phonological processes applied to reflexes of the 3rd declension accusative plural *-ES* (masculine and feminine), yielding **-eī* and thence *-i*, e.g. *DENTES* > **denteī* > It. *denti*, Ro. *dinți* ‘teeth’, *PELLES* > **pelleī* > **pelli* > It. *pellì*, Ro. *piei* ‘hides.’ It is possible that the archaic and also late spoken Latin plural ending *-īs* (nominative and accusative), well attested in the Italian and Balkan areas for nouns in *-i* (*OVIS* ‘sheep’, *NAVIS* ‘sheep’, etc.), underwent the same development (cf. Faraoni, *L’origine dei plurali italiani in -e e -i*, 2018: 166–167). The vocalisation of the sigmatic plural markers (*-Ās*, *-ES* > *-e*, *-i*) can thus in many cases be explained as the result of the palatalising effect of word-final *-s* – which had itself become [j] – on the preceding vowel (*-Ās* > *-ajz* > *-ez* > *-e*; *-Ēs*, *-īs* > *-iz* > *-i*).

The origin of masculine plurals in *-i* belonging to the Latin 2nd declension is more complex. Given the types *amici* (ending in *[-tʃi]*) and *fuochi* (*[-ki]*) ‘fires’, Maiden (2000) and Faraoni (2018: 109–146) suppose that the plurals developed within the framework of a bicasual system (‘subject case’ vs. ‘oblique case’, cf. 7.3.4) during the pre-textual period (cf. 10.3.5), starting from the antecedents [ami:tʃi]/[ami:kos] (< *AMICI/AMICOS*) and [fɔ:tʃi]/[fɔ:kos] (< *FOCI/FOCOS*). Nouns very high on the animacy scale (cf. 8.4.3 no. 2 below) would have continued the nominative form; the other nouns would have continued the accusative form but would have adopted the nominative ending *-i*, by then frequent within the paradigms, once [fɔ:kos] had passed to pl. [fɔ:ko] and could no longer be distinguished from the sg. form [fɔ:ko].

The above examples highlight the complex nature of the various combinations of singular and plural markers observable in Romance, which may be the product of the combination of stems and affixes, the result of phonetic changes, or the etymological residue of Latin forms. As we have seen (7.2.2), the inflectional paradigms that came into being as a result of these processes constitute the new Romance inflectional classes.

7.3.3 Gender

The restructuring of the Latin gender system in the Romance languages almost exclusively concerns the neuter; masculine and feminine have been maintained throughout the Romance area (with the above-mentioned changes which affected inflectional classes).

A number of specific lexemes were subject to changes of gender in the various Romance languages. Representatives of the Latin 3rd declension display a high degree of variation in this respect:

Lat. *DENS* M > Fr. *la dent* F vs. It. *il dente*, Sp. *el diente*, both M

Lat. *PONS* > Fr. *le pont* M, Sp. *el puente*, It. *il ponte* vs. Gal./Pg. *ponte*, O.Sp. *la puente*, Rom.n *punte* ‘footbridge’ (cf. **pōnt-e/* in DÉRom [M. Andronache]).

As for the neuter, it has been abandoned in the majority of Romance languages, and in all standard varieties except Romanian (where, however, it does not necessarily represent continuity with regard to Latin, but seems to have emerged as a secondary ‘epiphenomenon of agreement behaviour that depends crucially on the inflexional identity of the singular and plural forms of the nouns’, Maiden, *The Romanian alternating gender in diachrony and synchrony*, 2016). The items in question – nouns, as well as adjectives and specific pronouns – have been absorbed into the other two grammatical genders. This does not usually concern words denoting referents with biological sex, as in Latin, too, gender is assigned in accordance with the biological sex of the referent wherever possible. In the majority of cases, however, the latter does not have a sex and this often leads to differing results in the various Romance languages, thereby contributing to lexical divergence:

Lat. *mare* N “the sea” > Fr. *la mer*, Rom.n *marea* F vs. It. *il mare* M; in Sp. the equivalent form occurs with both genders, *el mar* as the general standard form and *la mar* in a series of set phrases such as *hacerse a la mar*, *en alta mar*, *la mar de* + X

The formal distinction of masculine and feminine has been reinforced to some extent through the reduction of the neuter and the restructuring of the inflectional classes. The majority of lexemes in Sp., Pg. and It. thus distinguish the affixes *-o* and *-a* respectively for these two genders in the singular.

Where the neuter value is maintained, it displays reorganisation with regard to Latin; this is the case in Romanian as well as in various Italo- and Ibero-Romance varieties. In Italian, for instance, the type *il braccio* – *le braccia* mentioned above is a formal vestige of the Latin neuter which plays a regular role in the language in marking collective plurals (*braccia*, *dita* “fingers”, *membra* “limbs”, *uova* “eggs”, *ossa* “bones”, etc.) and plurals of plurals (*migliaia* “thousands”, *centinaia* “hundreds”, *paia* “pairs”, etc.).

In general, the strongest persistence of the neuter function is to be found in the pronominal system. Ibero-Romance forms such as *esto*, *ello* and *lo* represent a grammatical neuter (of the type displayed by the Latin or Romanian gender systems) and they contrast with the masculine forms *este* and *él*. Furthermore, forms like French *cela*, *ceci* and *ça* can be considered as a ‘semantic’ neuter corresponding to non-countable forms, but without grammatical gender opposition.

→ Loporcaro, *Gender*, OxfGuide 57

cf. also Rainer, *Sex-denoting patterns of word formation in the Romance languages*, OxfEnc

7.3.4 The reduction of the Latin case system

The most salient change affecting Romance nominal inflection was the marked reduction and eventual loss of the Latin case system. The main significance of this large-scale

transformation lies in its impact on syntax. In Latin, case marking serves to distinguish between subject (nominative), 1st object (accusative) and 2nd object (dative); these ‘constituents’ (cf. 8.4.3) thus acquire a rather extensive freedom of position (*regis filius* vs. *filius regis* “the king’s son”). In some cases, differences of position become associated with particular nuances of meaning (this concerns phenomena such as emphasis or focus marking).

From a typological point of view, such a high degree of inflection is typical of ancient Indo-European languages (such as Sanskrit, ancient Greek or the Old Germanic languages) – and is relatively common among the world’s languages – but it is notably absent from creole varieties and other simplified systems of communication (cf. 8.4.3 no. 6). In this context, therefore, it is interesting to observe the evolution of the Latin system towards the Romance type of encoding, whereby the Latin case system is abandoned and the roles of the constituents are marked by means of position or pre-posed function words (which evolved from prepositions and conjunctions):

position: It. *Sergio_s ha visto Michele_{o1} ieri* “Sergio saw Michele yesterday”

function words: FRATR-IS “of/belonging to the brother” → Fr. *du frère*, It. *del fratello*, Sp. *del hermano*; Sp. *doy un libro a la chica* “I give/I’m giving the girl a book”.

The six cases of Classical Latin (including the vocative case) were first reduced to three – nominative/vocative, accusative and genitive/dative – in Late Latin. In all Romance varieties, with the exception of Romanian, the genitive/dative case was subsequently absorbed by the accusative, which thus became the most frequent case²¹. In Daco-Romance, however, a bicasual system is maintained for feminine nouns and adjectives, in which the nominative and accusative have merged and contrast with the genitive/dative (NOM.-ACC. vs. GEN.-DAT.).

The distinction of a nominative (which also assumes the functions of the vocative) from all other cases (in the form of the Latin accusative) partially survived into Medieval Gallo-Romance. This two-case system (NOM. vs. OBL. [= oblique < ACC.]; in French, *cas sujet* ‘subject case’ vs. *cas régime* ‘oblique case’), however, had already lost its functionality in the earliest Old Occitan texts. In Old French, it began to disintegrate from the 12th century onwards and by the transition to Middle French (ca. 1330), it was no longer functional. Only in Francoprovençal was the opposition entirely maintained until the 16th century, and some relics have survived into modern varieties.

Similarly, in several Ladin varieties, final -s in adjectives was a productive nominative marker for centuries (and it has now become a marker of the adnominal/attributive distinction).

21 The same phenomenon can be observed in dialectal Arabic, for example, where the accusative asserts itself over the nominative as a result of its frequency.

Moreover, the marked formal difference between the nominative and the oblique case in nouns of the 3rd declension sometimes resulted in the independent survival of both forms. In such cases, the lexemes underwent semantic differentiation:

Fr. *pâtre* “shepherd” / *pasteur* “pastor, clergyman” (< ‘PASTOR – PAS’TOREM)
compaing ~ *copain* “friend” / *compa(i)gnon* “companion” (< COM’PANIO – COMPA’NIONEM)
sire “Sire, lord” / *seigneur* “master, lord” (< ‘SE(N)IOR – SE’NIOREM)
pute (< PUTIDUS, literally “stinking”) / *putain* (by analogy), both meaning “whore” and used as common swear words.

Very exceptionally, only the nominative form of a French lexeme has survived:

Fr. *traître* (< TRADITOR (nom.), “traitor”), instead of the oblique form **traiteur* (< TRADITOREM), owing to its potentially confusing homophony with *traiteur* (which is derived from the verb *traiter* (< TRACTARE “to draw, to handle”) and whose meaning is “delicatessen” or “caterer” in Modern French.

The example of the two-case inflectional system of Old French illustrates the convoluted history of the reduction of inflectedness. Thus, the distinction of a marked subject case from an unmarked oblique case was transparent only in the masculine nominal paradigms of Old French:

2 nd decl. Lat. M:	NOM.SG (<i>li</i>) <i>murs</i> (< MURUS)	OBL (<i>le</i>) <i>mur</i> (< MURUM)
	NOM.PL (<i>li</i>) <i>mur</i> (< MURI)	OBL (<i>les</i>) <i>murs</i> (< MUROS)
3 rd decl. Lat. M:	NOM.SG (<i>li</i>) <i>pere(s)</i> (< PATER)	OBL (<i>le</i>) <i>pere</i> (< PATREM)
	NOM.PL (<i>li</i>) <i>pere</i> (< PATRES)	OBL (<i>les</i>) <i>peres</i> (< PATRES)

In the nominative singular of nouns derived from the Latin 3rd declension, an -s was introduced by analogy with the 2nd declension (-s thus became an inflectional affix of the masculine singular nominative form); in the same manner, also by analogy with the 2nd declension, word-final -s was dropped from the nominative plural of the 3rd declension.

Case distinction was lost completely in the feminine nominal paradigms, which were only marked for number:

1 st decl. Lat. F:	NOM.SG (<i>la</i>) <i>rose</i> (< ROSA)	OBL (<i>la</i>) <i>rose</i> (< ROSAM)
	NOM.PL (<i>les</i>) <i>roses</i> (< ROSAE)	OBL (<i>les</i>) <i>roses</i> (< ROSAS)

The -s was introduced as a marker of the pl. subj. case by analogy with the oblique pl.

3 rd decl. Lat. F:	NOM.SG (<i>la</i>) <i>flours</i> (< *FLORIS < FLOS)	OBL (<i>la</i>) <i>flour</i> (< FLOREM)
	NOM.PL (<i>les</i>) <i>flours</i> (< FLORES)	OBL (<i>les</i>) <i>flours</i> (< FLORES)

Paradigms became especially complicated for nouns derived from the Latin 3rd declension, as the position of the accent differed according to whether the form was nominative or accusative. Hereditary phonetic evolution thus gave rise to ‘stem alternation’:

M:	NOM.SG <i>emperére</i> (< IMPE’RATOR)	OBL <i>empereór</i> (< IMPERA’TOREM “emperor”)
	NOM.PL <i>empereor</i> (< *IMPERA’TORI < IMPERA’TORES)	OBL.PL <i>empereors</i> (< IMPERA’TORES)
F:	NOM.SG <i>suer</i> (< ‘SOROR)	OBL <i>sorour</i> (< SO’ROREM “sister”)
	NOM.PL <i>sorours</i> (< SO’RORES)	OBL.PL <i>sorours</i> (< SO’RORES)

Other examples include: *niés* – *neveu* “nephew” (< ‘NEPOS – NE’POTEM), *énfes* – *enfant* “infant” (< ‘INFANS – IN’FANTEM), *lerre* – *larron* “thief” (‘LATRO – LA’TRONEM), *traí(s)tre* – *traitor* “traitor” (< *TRA’DITOR – TRADI’TOREM), *uem* – *ome* (‘HOMO – ‘HOMINEM), *mielldre* – *meillor* “better” (< ‘MELIOR – ME’LIOREM), *graindre* – *graignor* “bigger” (< ‘GRANDIOR – GRAN’DI-OREM).

The two- (or three-) case inflectional system of Romanian opposes an unmarked case (nom./acc.) *cas-ă* “house” / *cas-a*, “the house” to a marked case (gen./dat.) (*unei*) *cas-e* “of/to (a) house” / *cas-ei* “of/to the house”. Romanian also possesses a form that functions as a vocative (masc. *domnule* [for *domn*] “Oh Lord!”; fem. *Mario* [for *Maria*] “Oh Maria!”). The distinction between the unmarked and the marked case only fully applies to the feminine singular for nouns and adjectives; for determiners, however, it is general. Thus, for the plural and the masculine, the case opposition is marked by the article (and most determiners); since the definite article in Romanian is ‘inflectional’, all masculine nouns also show case inflections (e.g. *bărbat-ul* “man” NOM/ACC.M.SG vs. *bărbat-ului* GEN/DAT.M.SG and *bărbați-i* NOM/ACC.M.PL vs. *bărbați-lor* GEN/DAT.M.PL).

The loss or reduction of Latin case in Romance had far-reaching consequences for the system of nominal determination as well as for sentence structure. It represented the transition from an inflectional to an agglutinative principle, which implies, above all, a partial loss of the cumulative linguistic encoding present in Latin.

→ Dragomirescu/Nicolae, *Case*, OxfGuide 56
Nicolae, *Case-marking in the Romance languages*, OxfEnc

7.4 Elements of adjectival inflection

7.4.1 Comparison

Romance adjectival paradigms are aligned with the nominal paradigms previously discussed (cf. 7.3). Adjectives thus display the same inflectional features as the nouns with which they are in agreement (gender, number, case), with the essential difference that gender is not inherent to adjectives, but is determined by the corresponding nouns. The only inflection inherent to adjectives – and absent in nouns – concerns the degrees

of comparison. This feature is particularly apparent in Latin, where it is essentially produced by inflectional means, whereas the Romance languages prefer periphrastic constructions.

While Latin distinguishes three degrees of comparison, Romance distinguishes four:

- positive: Lat. FORTIS, Fr. *fort* “strong”
- comparative: Lat. FORTIOR ~ FORTIUS, Fr. *plus fort* “stronger”
- relative superlative: Lat. FORTISSIMUS, Fr. *le plus fort* “the strongest”
- absolute superlative: Lat. FORTISSIMUS, Fr. *très fort, peu fort* “very strong, not very strong”

In addition to the regular series, Latin also possessed a series of suppletive adjectival paradigms (e.g. *bonus* “good” – *melior* “better” – *optimus* “best”).

In the Romance languages, most regular inflectional comparative paradigms were replaced by constructions following an agglutinating principle consisting of the positive form of the adjective preceded by a grammaticalised adverb (Fr. *plus*, Sp. *más*, It. *più*). The adjective thus remains inflected:

FORTIOR → Fr. *plus fort* M / *plus forte* F / *plus forts* M.PL / *plus fortes* F.PL
 Sp. *más fuerte* M/F / *más fuertes* M/F.PL
 It. *più forte* M/F / *più forti* M/F.PL

Remnants of the suppletive paradigms survived in some Romance languages, but by no means in all of them:

MELIORE > Fr. *meilleur*, Sp. *mejor*, It. *migliore* vs. Rom.n *mai bun*
 MINUS > Fr. *moins*, Sp. *menos*, It. *meno* vs. Rom.n *mai puțin*

In traditional terms, these inflectional forms are referred to as ‘synthetic’ comparatives; however, in all Romance languages, some comparatives of this type adopted patterns of regular declension formed on the basis of an agglutinative (or ‘analytical’) principle:

MAJOR > Fr. *plus grand*, Sp. *más grande*, It. *più grande*.

The Latin superlative was greatly weakened in the Romance languages; in French, it is marked by the definite article (*le plus fort*); in many cases, it has been replaced by paraphrases:

It. *ricco ricco* “extremely wealthy”, Pg. *horridamente feio* “hideously ugly”.

→ Van Peteghem, *Comparatives and superlatives in the Romance languages*, OxfEnc

7.4.2 Adverb formation

Adverb formation also underwent changes between Latin and the Romance languages. In Latin, adverbs were formed from adjectives by means of a series of inflectional affixes (*dign-us* “worthy” → *dign-e* “worthily”, *fals-us* “false” → *fals-o* “falsely”, *acer* “sharp” → *acr-iter* “sharply”, with stem alternation in the last example). The Romance languages developed various patterns of adverb formation:

- In Romanian, an inflectional type of adverb formation, similar to the Latin model, has been preserved sporadically, as in adj. *firesc* “natural” → adv. *firește* (*firesc* is a derivative of the noun *fire* n. “being”, itself a deverbal formation based on Lat. *FIERI*).
- Most Romance languages display the same suffix (*-men(t(e))*), which is added to the feminine singular form of adjectives to form adverbs: Lat. *LENTE* “slowly” gives rise to Fr. *lent-e-ment*, It./Sp./Pg. *lent-a-mente*, Cat. *lent-a-ment*. This affix originally derived from a lexeme (*MENTE*), which was integrated into the inflectional system and therefore grammaticalised. The grammaticalisation process is complete in French (e.g. *antérieurement*), whereas in other languages, the composition has only undergone partial grammaticalisation, such that the ellipsis of the suffix may occur in pairs of adverbs: OOcc. *francamen e corteza*, Cat. *probrament i honesta*, Sp. *pura y simplemente*, Pg. *pure e simplesmente*. This type of adverb formation may be regarded as part of derivational morphology, as it involves a change in word class (cf. 9.4.2 no. 2).
- In the dialects of central and southern Italy, the neighbouring languages (Sardinian, Friulian, Dalmatian), as well as in Romanian in general, the forms of adverbs and adjectives often coincide with the masculine singular form of the adverb.
- Finally, in some cases, adverbs of Latin origin have survived in a lexicalised form (Fr. *bien* < *BENE* vs. *bon* < *BONU-* [beside *bon-n-e-ment*]); the formal relationship with the corresponding adjective has thus been reduced. A small number of Latin adverbs have even survived in the absence of a corresponding adjective in the Romance languages (*hier* ADV < *HERĪ* ADV “yesterday”).

7.5 The pronoun system and the determiners ('D-system')

7.5.1 Form and function: full pronouns, clitics and determiners

Pronouns and determiners together form a system that can be considered as highly ‘morphosyntactic’. Its constituent elements are of central interest to morphology, as they are inflectional and their inflected forms are organised into paradigms, but also to syntax, as they fulfil important functions at the level of phrases and sentences. Due to

the high degree of interdependence between formal and functional elements, the morphological and syntactic aspects of these items will be examined together.

The Latin pronominal system included personal, possessive, demonstrative, interrogative, relative and indefinite pronouns, which could all function as noun substitutes in their own right. The Romance languages maintained the Latin forms, but broadened their range of functions: in addition to their role (i) as full pronouns, they can also function (ii) as determiners of nominal elements or (iii) as subject and object clitics attached to the verb phrase (VP).

personal pronouns and determiners: (i) Fr. *lui*, It. *lui*, Sp. *él* vs. (iii) subject clitic Fr. *il*, object clitic Fr. *le*, It., Sp. *lo*

possessive pronouns and determiners: (i) Fr. *le mien* “mine”, Sp. *el mío* vs. (ii) *mon livre* “my book”, *mi libro*; in It. the two types are formally identical: *il mio* and *il mio libro*

Stressed pronouns (i) occupy the same syntactic position as the noun phrase to which they refer – hence the term *pronoun*. They thus represent a specific type of particularly heterogeneous noun phrase (NP). Determiners can assume functions similar to those of adjectives (ii), which is why, in traditional approaches, they are often included in the class of adjectives.

Like nouns and adjectives, pronouns and determiners may be marked for gender and number and, where applicable, case. Like verbs, moreover, the personal and possessive series are marked for person.

- gender (only applicable to the 3rd person): (i) Fr. *lui* – *elle*, It. *egli* – *ella*, Sp. *él* – *ella*
- number: (i) Fr. *lui* – *eux*, *moi* – *nous*, It. *lui* – *loro*, *io* – *noi*, Sp. *él* – *ellos*, *yo* – *nosotros*
- case (only applicable to personal pronouns and determiners): (iii) 1^p: Fr. *je* [subject] – *me* [object], It. *io* – *me*, Sp. *yo* – *mí*, 3^p: Fr. *il* – *le* – *lui*, It. *lui* – *lo* – *gli*, Sp. *él* – *lo* – *le*
- person: Fr. (i) *moi* – *toi* – *soi*, *mien* – *tien* – *sien*, (ii) *mon* – *ton* – *son*, (iii) *je* – *tu* – *il*, *me* – *te* – *se*

Semantically speaking, at the time of their formation, the emergence of the ‘unstressed’ forms – determiners and clitics (ii and iii) – reflected a tendency towards salience and transparency (known as ‘topicalisation’ and ‘focalisation’). As in other cases, the initial semantic and pragmatic connotations were later lost as a result of the process of grammaticalisation.

For example, O.Fr. displays the opposition *il joue* vs. *Ø joue*, as in Modern It. *lui gioca* vs. *Ø gioca* or Sp. *él juega* vs. *juega*; in Modern French, the subject clitic is obligatory and emphasis (or topicalisation) must be expressed by means of a complementary full pronoun: *lui, il joue* “he’s the one who is playing”.

The fact that three paradigms with different functions developed from a single Latin paradigm considerably complexifies their grammatical description. Only the series of ‘stressed’ forms (i: *moi*, *toi*, *soi*, ...) represent full pronouns that display some degree of

syntactic autonomy and are capable of standing alone as noun phrases. In contrast, the two groups of 'unstressed' forms (ii: *mon, ton, son*, iii: ...*je, tu, il, ..., me, te, se, ...*) have lost their syntactic independence and have been integrated into the noun phrase (ii: as determiners cf. 8.2.3) and the verb phrase (iii: as clitics functioning as agreement markers). At the same time, as is evident from the above examples, the three groups are interdependent as regards their function, and their forms also display an overlap in diachrony.

Consequently, despite their highly diverse functions, the three series of 'stressed' and 'unstressed' forms based on Latin pronouns are to be considered as inseparable from a historical perspective; synchronically, however, it is necessary to distinguish fully-fledged *pronouns* from *clitics* (or *clitic pronouns*) on the one hand, and *determiners* on the other.

- Pescarini, *Pronoun systems in the Romance languages*, OxfEnc
- Cappellaro, *Tonic pronominal system: morphophonology*, OxfGuide 44
- Pescarini, *Clitic pronominal system: morphophonology*, OxfGuide 45
- Poletto/Tortora, *Subject clitics: syntax*, OxfGuide 47
- Roberts, *Object clitics*, OxfGuide 48

7.5.2 Personal pronouns and determiners

Within the category of personal pronouns and determiners, the following types may be distinguished:

- stressed forms of personal pronouns: Lat. ME > Fr. *moi*, It. *me*, MIHI > Sp. *mí*
- clitic forms of personal pronouns:
 - subject clitics: EGO > **eo* > Fr. *je*, Sp. *yo*, It *io*, etc. (cf. 8.4.3 no. 3 for the introduction of obligatory unstressed forms)
 - object clitics: Fr. *me*, Sp. *me*, It. *mi*
- with regard to the 3rd person, distinct forms exist for the object and the '2nd object' (typically an indirect object): *il – le – lui*, Sp. *él – lo – le*, It. *lui – lo – gli*

The three values of person distinguished by each of the three series correspond to speaker (1st person), addressee (2nd person) and non-participant referent (3rd person).

In the Romance languages, pronouns display full functionality with regard to the replacement of nouns denoting animate beings; however, this is not the case for nouns denoting (inanimate) objects. In Sp., Pg. and Rom.n, pronominalisation is possible in prepositional constructions (Sp. *el libro* "the book" / *sin él* "without it"), whereas in the other Romance languages, equivalent constructions are ungrammatical (cf. in Fr.: *le livre* / **sans lui*). Some languages possess a type of 'zero-pronoun' construction, which is diastystematically marked as a low-prestige variant (Fr. *il vient avec* "he's coming with",

It. *non può fare senza* “he/she can’t do without”, Rom.n *Ai văzut casa?* → *Da, am văzut* “Have you seen the house? – Yes, I have” [= ‘I have seen’]), as well as ‘adverbial’ pronouns (Fr. *y/en*, It. *ci/ne*).

Reflexive pronouns display an even closer connection with the verb phrase (VP). Most belong to the category of clitics (and are thus unstressed), although there are also a number of stressed reflexive pronouns, particularly in Spanish and Italian:

Sp. unstressed	<i>me – te – se</i>	vs.	stressed	<i>mí – ti – sí</i>
It. unstressed	<i>mi – ti – si</i>	vs.	stressed	<i>me – te – sé</i>
Fr. unstressed	<i>me – te – se</i>	vs.	stressed	<i>moi – toi</i>

The French stressed form *soi*, in contrast, is on the verge of disappearing from usage (except when used to refer to an indefinite subject such as *on* or *chacun*).

7.5.3 Personal pronouns and verbal politeness

For a long time, Latin only possessed two forms of address for the second person: singular for one addressee and plural for two or more. From a functional point of view, this system displayed complete coherence as regards communicative practice and the linguistic structure of the verbal paradigms. The Romance languages, however, developed a more complex system which allowed for the expression of politeness, thus creating a contrast between social or personal distance and intimacy.

In the case of a single addressee, politeness is expressed in Modern French by means of the 2P.PL (*je vous salue* “I greet you”), and in Italian or Peninsular Spanish by the 3P.SG (It. *La saluto*, Sp. *le saludo*). Romanian distinguishes two levels: the less formal *dumneata* “your (sg.) lordship” – often used between equals – and the highly formal *dumneavoastra* “your (pl.) lordship”.

For two or more addressees, French also uses the 2P.PL; the same form is also now used in Italian (*Vi saluto*), replacing the former 3P.PL, which has become obsolete. Spanish uses the 3P.PL (*les saludo*) – which in American (and Canarian) Spanish, however, expresses both intimacy and distance. In Romanian, *dumneavoastra* is the only polite form. In French, Italian and American/Canarian Spanish, there is thus no distinction between intimacy and distance for two or more addressees, as in Latin.

The use of a plural form for one addressee (rather than the expected singular form) can be observed in Latin from the 3rd century onwards, in the form of address used for the Roman emperor (*pluralis maiestatis*), although there is little evidence to suggest that this is the direct source of politeness strategies in the spoken Romance languages. In Old French texts, the early occurrences of the plural form of address employed for a single person marks personal, rather than social, distance. This type of diaphasic distinction persists in modern colloquial Spanish, despite the fact that in the Middle Ages the plural

form evidently carried social (= diastratic) implications in both Spanish and Italian (in the latter case, from the 14th century onwards). In 16th- to 18th-century writings, the diastratic, diaphasic and expressive implications of the forms of address used to convey distance or intimacy are generally co-present, and are thus difficult to distinguish from one another. It is a well-known fact that the attempt to abolish the *vous* form during the French Revolution (1793) was unsuccessful and the law was repealed three years later.

Finally, verbal politeness also implicates nominal forms of address (Fr. *Monsieur / Madame*, Sp. *Señor / Señora*, Pg. *O Senhor / A Senhora* or – more formal – *Vossa Senhoria*) as well as forms originating from possessive pronouns by contraction (e.g. Pg. *você* “you” < *vossa merced* “Your Grace”).

→ Ashdowne, *Address systems*, OxfGuide 55

Da Milano/Jungbluth, *Address systems and social markers*, CambrHandb 25

7.5.4 Possessive pronouns and determiners

Initially, the Romance languages also had a double series of possessives consisting of stressed pronouns and ‘adjective-like’ nominal determiners (cf. Fr. *mien* vs. *mon*, Sp. *mío* vs. *mi*).

When examined in detail, however, the historical relationship between formal and functional aspects is far from simple: in O.Fr., for instance, the series of stressed possessives could function both as pronouns and as determiners (e.g. as a determiner in the formulaic expression *mien escientre* “as far as I know”), alongside an unstressed series. Stressed and unstressed determiners seem to have been used interchangeably to some extent, depending on the intended degree of expressivity. Only later was the binary distinction observable in Modern French established.

In other Romance languages, the two series have once again been reduced to a single series of stressed forms that are used as both determiners and pronouns (cf. It. *è mio* “it’s mine” vs. *il mio libro* “my book”).

7.5.5 Demonstrative pronouns and determiners

The formal distinction between a stressed and an unstressed series of demonstratives has been greatly reduced. In most cases, identical forms can be used as pronouns and noun determiners (It. *questo non lo credo* “I don’t believe this” vs. *hai visto questo ragazzo?* “have you seen this boy?”).

In contrast, various Romance languages and dialects have maintained the semantic distinction of three degrees of proximity, which was already present in Latin: near the speaker (‘proximal’), near the addressee (‘medial’), and far from both speaker and addressee (‘distal’), as illustrated by the following examples:

Cl.Lat.:	HIC/IPSE	ISTE	ILLE
Late Lat.:	ISTE	IPSE	ILLE
Sp.:	<i>este</i>	<i>ese</i>	<i>aquel</i>
Pg.:	<i>este</i>	<i>esse</i>	<i>aquele</i>
It.:	<i>questo</i>		<i>quello</i>
Fr.:	<i>celui(-ci)</i>		<i>celui(-là)</i>

In Modern Italian, *codesto* for an intermediate degree of distance only appears in highly formal administrative contexts, except in Tuscany, where its use is more general; like Catalan or Romanian, Italian therefore has a binary system today. In French a further reduction to a single system took place in the 14th/15th centuries. The semantic reduction is correlated with a grammatical distinction between the pronoun *celui* and the determiner *ce*. The difference in proximity is expressed in a secondary manner by the addition of *-ci* vs. *-là*.

In O.Fr. the corresponding forms *cist* (which expressed proximity) and *cil* (which expressed distance) were used as both pronouns and determiners – in the Gallo-Romance area, the distinction of three degrees of proximity present in Latin had already been reduced to two in Late Latin.

As shown by the above example, the evolution of the demonstratives implies a strong element of expressivity, which is linked to the deictic nature of these forms. Thus, Old French *cil* and *cist* evolved from the combination of two Latin demonstratives (ECCE + ILLE (“behold” + “that”) and ECCE + ISTE (“behold” + “this”)), as did the Modern French clitic *-ci*, which evolved from ECCE + HIC. The full pronoun *celui-ci* originates from at least four Latin function words.

→ Ledgeway/Smith, *Deixis*, OxfGuide 54

7.5.6 Relative pronouns

The functions of the Latin relative pronouns remained intact, although their inflectional properties were mostly lost in Romance; the relationship between relative pronouns and the elements to which they refer is commonly marked by position as well as by the use of prepositions.

However, a case distinction still exists in French (*qui* NOM vs. *que* ACC) and Romanian (*care* NOM/ACC vs. *cărui(a)/cărei(a)* GEN/DAT), and can further be observed in various early dialects of northern Italy (cf. Parry, *The interaction ... in the spread of relative che ...*, 2007). In addition, substitute forms have developed in various languages (e.g. Fr. *lequel*, It. *il quale*, Sp. *el cual*), although these are relatively infrequent and highly formal and their use is limited to particular written registers.

Spoken varieties of present-day Romance languages have developed a relative pronoun used exclusively to express subordination without distinguishing case. This

dismantling of functions results in the restructuring of subordinate clauses, which thus acquire the character of attributive adjectives (cf. GME, p. 475):

non-standard Fr. *le livre que je t'en avais parlé*

familiar It. *uno che ero amico* or *nella città che ci sto io* vs. standard It. *un uomo del quale ero amico* or *nella città in cui/dove sto io*

7.5.7 Indefinite pronouns

Indefinite pronouns represent a particularly heterogeneous category consisting of personal, possessive, demonstrative, relative, and interrogative pronouns (i.e. all pronouns that do not belong to any of the four main classes previously mentioned). Indefinite pronouns can nevertheless be divided into two distinct categories: *quantifiers* and *identifiers*.

Quantifying items can express negation, universality, singularity or non-universal plurality:

negation:	Fr. <i>nul, aucun, pas un, personne, rien</i> ; It. <i>nessuno</i> , etc.
universality:	Fr. <i>tout, tous/toutes, chacun</i> ; It. <i>tutto, tutti/tutte</i> , <i>ognuno</i> ; Sp. <i>todo</i> , etc.
singularity:	Fr. <i>quelqu'un, quelque chose, n'importe qui/quoi</i>
non-universal plurality:	Fr. <i>certains, la plupart, plusieurs</i>

Identifying items, on the other hand, function anaphorically (i.e. they refer to previously mentioned elements). They either identify referents (Fr. (*le*) *même*, It. *lo stesso*, Sp. *el mismo*) or distinguish between referents (*un/des/quelques* *autre(s)*) and they are preceded by definite or indefinite determiners.

The evolution of these different types of pronouns and determiners has given rise to a great deal of variation in the Romance languages; they are thus of particular interest for historical and comparative research.

7.6 The restructuring of Latin verb paradigms

Throughout all the stages of their development from Latin, verbs displayed the most complex form of inflection, and they have retained this characteristic in modern Romance varieties. The total number of inflected forms for individual verbs in the Romance languages can amount to over one hundred, in addition to uninflected infinitives, present and past participles, gerunds as well as, for Romanian, the supine. Verb morphology involves the following six morphosyntactic features:

- person: speaker (1st person), addressee (2nd person), non-participant referent (3rd person)
- number: SG vs. PL
- tense: PRS, FUT, PAST etc., whereby simple forms (such as It. *cantavamo* sing.1PL. IMPERF.) are to be distinguished from periphrastic forms, which employ an auxiliary verb (It. *abbiamo cantato* sing.1PL.PFV.)
- mood: IND, SBJV, COND, IMP(ERATIVE)
- aspect: PRF vs. IPF
- voice / diathesis: ACT vs. PASS vs. MED (middle) (the latter only existing in Latin and in some Romance dialect varieties, Italo-Romance in particular).

Person and number are universal features, which thus display a high degree of continuity between Latin and the Romance languages (*cf.* the detailed overview of the formal changes in person and number marking by Maiden, *Inflectional morphology*, OxfGuide 27.3). In Latin, as well as in Romance, they normally agree with the subject (*cf.* 8.4.3 no. 3). The only significant change is the introduction of politeness strategies (*cf.* 7.5.3).

The other morphosyntactic categories underwent fundamental changes in the Romance languages. As mentioned above, they will be discussed in the following chapter on syntax, given their implications for sentence structure (for tense, aspect, mood [TAM], *cf.* 8.3.2 nos. 1–3, for voice / diathesis, no. 5). An overview of the inflectional categories retained, abandoned or newly introduced between Latin and the Romance languages will also be provided, *cf.* 2.3.3). In the present section, the complexity of the formal restructuring of verb paradigms from Latin to Romance will be illustrated.

Verb paradigms underwent significant formal change in the transition from Latin to Romance, much of which was unpredictable. Different verb forms were not only subject to the usual patterns of phonetic change, but also to the strong influence of analogy, which played a role both within and among paradigms: the most frequent forms tended to exert influence on the others, and the most frequent lexemes often acted as models. Consequently, the individual forms of many verbs underwent a completely separate development. A detailed account of the evolution of a specific verb paradigm in a given Romance language thus requires considerable effort, even though, at least from a structural point of view, the corresponding model already existed in Latin.

Though the same phenomena can be observed for other parts of speech, they are more marked in verbs, owing to the large number of forms displayed by each individual paradigm, as well as to the fact that the inflectional system of Latin was mainly retained, in contrast to the nominal system.

In order to identify the main principles underlying the restructuring of the paradigms, a brief commentary on the forms of the present indicative, the past imperfect and the future paradigms in French, Spanish and Italian will be provided, with particular attention to the marking of number and person.

7.6.1 Present indicative

The first example, which shows the Romance issues of the present indicative of a Latin first conjugation verb, serves to illustrate the fact that the inflectional affixes of even this relatively simple and stable paradigm underwent a number of analogical changes, as will be shown in the commentary below.

Lat.	Fr.	Sp.	It.
'CANT-O	(je) chant- <i>e</i> ['ʃāt]	cant- <i>o</i>	cant- <i>o</i>
'CANT-A-S	(tu) chant- <i>e-s</i> ['ʃāt]	cant- <i>a-s</i>	cant- <i>i</i>
'CANT-A-T	(il) chant- <i>e</i> ['ʃāt]	cant- <i>a</i>	cant- <i>a</i>
CAN'T-Ā-MUS	(nous) chant- <i>ons</i> [ʃā'tɔ̃]	cant- <i>a-mos</i>	cant- <i>i-amo</i>
CAN'T-Ā-TIS	(vous) chant- <i>ez</i> [ʃā'te]	cant- <i>á-is</i>	cant- <i>a-te</i>
'CANT-A-NT	(ils) chant- <i>ent</i> ['ʃāt]	cant- <i>an</i>	cant- <i>a-no</i>

Remarks:

- in O.Fr., 1SG was aligned with 3SG, adopting word-final *-e* (although inherited forms without *-e* are also frequent in textual sources: *plour* “I cry”). Owing to the loss of word-final sounds (both unstressed vowels and consonants), the 1SG, 2SG and 3SG, as well as 3PL became homophonous; the neutralisation of the three singular forms and, therefore, the loss of distinction with regard to person (syncretism) was later counterbalanced by the introduction of obligatory subject clitic pronouns (cf. 8.4.3 no. 3). The feature of number nevertheless remained indistinct for 3SG and 3PL; these forms are still homophonous in French today
- in French, the 1PL ending *-ons* was formed by way of interference with the paradigm of the highly frequent Latin verb *ESSE* “to be” (reanalysing the form *sŭmus* 1PL and generalising the sequence, **-ŭmus* as a suffix). In a second step, the French plural affixes *-ons*, *-ez*, *-ent* / *-ont* were extended to all conjugation classes and all tenses
- Spanish forms generally conform to the expected result of regular phonetic evolution (with the exception of the disappearance of *-t-* from the 2PL form, which should not have occurred according to the regular patterns of sound change)
- in Italian, an analogical *-i* was introduced for the 2SG, probably as a result of the phonetic change *-es > -i*; the generalisation of the ending *-iamo* for 1PL seems to be an analogical effect modelled on verbs such as *siamo*, *abbiamo* or *sappiamo* for which *-iamo* represents the historically regular outcome.

As illustrated by the above example, it is essential to compare the different tenses and moods with one another to gain an accurate historical understanding of the internal

structure of Romance paradigms, as well as of the nature and function of individual affixes (*cf.* the concept of the ‘morpheme’ discussed in the following subchapter).

7.6.2 Morphemes

Romance present indicative paradigms often display a striking form of stem space variation resulting in patterns known as ‘morphemes’, following Aronoff. From an etymological point of view, morphemes are often the result of regular phonetic change, but they can also arise through suppletion (e.g. DARE vs. DONARE “to give”, CADERE vs. CASCARE “to fall” etc.). The following example illustrates how two variants of a verb stem (Italian /leg:/ vs. /ledʒ:/) display specific patterns of distribution within a given inflectional class:

	1SG	2SG	3SG	1PL	2PL	3PL
/leg:/						
PRS.IND	<i>leggo</i>					<i>leggono</i>
PRS.SBJV	<i>legga</i>	<i>legga</i>	<i>legga</i>			<i>leggano</i>
/ledʒ:/						
PRS.IND		<i>leggi</i>	<i>legge</i>	<i>leggiamo</i>	<i>leggete</i>	
PRS.SBJV				<i>leggiamo</i>	<i>leggiate</i>	

Parallel distributions can be found in other inflectional paradigms: for instance, patterns analogous to those present in *leggere* are found in the present tense of the Italian verbs *crescere* “to grow” (/kresk/ vs. /krefʃ/: *cresco*, *cresci*), *cogliere* “to gather” (/kɔlg/ vs. /kɔλ/: *colgo*, *cogli*) and *dire* “to say” (/dik/ vs. /diʃ/: *dico*, *dici*). The ‘morpheme’ is neither syntactically nor semantically, nor phonologically defined; it is a pure morphological form (*cf.* the detailed diachronic description provided by Maiden, *Morphophonological Innovation*, CambrHist, 5).

One of the major sources of morphomic variation is the differing stress patterns in Latin 1PL and 2PL forms, whereby the action of regular phonetic evolution led to an alternation in the form of the word stem. This was a prominent feature of Old French:

- O.Fr. *deveir* (Fr. *devoir* “to have to”): *dei*, *deis*, *deit* ‘deivent vs. *de’vons*, *de’vez*
= the 1SG, 2SG, 3SG and 3PL display the diphthongisation of stressed -e- (< Lat. ‘DEBEO, ‘DEBES, ‘DEBET), while 1PL and 2PL remain undiphthongised since -e- is unstressed (< DE‘BEMUS, DE‘BETIS)
- O.Fr. *laver* “to wash”: *lef*, ‘*leves*, ‘*levet*, ‘*levent* vs. *la’vons*, *la’vez*
= the same contrast between forms is seen in the diphthongisation of stressed a- (< ‘LAVO, ‘LAVAS, ‘LAVAT, ‘LAVANT) and those without diphthongisation (< LA‘VAMUS, LA‘VATIS)

- O.Fr. *amer* “to love”: *aim*, 'aimes, 'aimet, 'aiment vs. *a'mons*, *a'mez* ('AMO, 'AMAS, 'AMAT, 'AMANT vs. A'MAMUS, A'MATIS)
= once again, the same contrast may be observed, with the difference that the diphthongisation of stressed *a*- is hindered by the presence of the nasal consonant; the graphic residue <ai> bears witness to this phenomenon.

This alternation, which was regular in Old French, was maintained in verbs of frequent use. These thus display morphomic variation in Modern French (traditional grammar speaks of ‘irregular’ verbs): *je dois* – *nous devons* or *je viens* – *nous venons*. Since the medieval period, however, the paradigms of a certain number of verbs have undergone realignment by analogy with either the stressed stem (*j'aime* – *nous aimons*, *je pleure* – *nous pleurons*) or the unstressed stem (*je lave* – *nous lavons*, *je trouve* – *nous trouvons*).

Morphomic variation forms a striking parallel with the suppletive paradigms characteristic of high-frequency Romance verbs (“to be” or “to go”, cf. 2.2.2 no. 3.1). In the latter case, the different stems are in fact the reflexes of the stem variation already present in Latin, or even of stems originating from other verbs (Fr. *je vais* < Lat. *VADO* / Fr. *nous allons* < Lat. *AMBŪLAMUS*).

→ O'Neill, *Morphologically 'autonomous' structures in the Romance languages*, OxfEnc

7.6.3 Past imperfect indicative

The past imperfect is traditionally referred to by the ellipsis ‘imperfect’ (Fr. *imparfait*, Sp. *imperfecto*, It. *imperfetto* etc.); this, however, leads to confusion between the tense category ‘past’ and the aspectual category ‘imperfect’ (cf. 8.3.2: TAM). We therefore favour the unambiguous label ‘past imperfect’.

The table below illustrates the development of the Romance past imperfect paradigms from the Latin first conjugation. As before, interference (or analogy) with a parallel paradigm may be observed: during the course of the evolution of the first conjugation towards French, the inflectional suffixes of the Latin second conjugation (Ē-BA-M) were adopted:

Lat.	O.Fr.	Mod.Fr.	Sp.	It.
CAN'T-Ā-BA-M	<i>chant-ei-e</i> > -oie	(je) <i>chant-ai-s</i>	<i>cant-a-b-a</i>	<i>cant-a-v-o</i>
CAN'T-Ā-BA-S	<i>chant-ei-es</i> > -oies	(tu) <i>chant-ai-s</i>	<i>cant-a-b-as</i>	<i>cant-a-v-i</i>
CAN'T-Ā-BA-T	<i>chant-ei-t</i> > -oit	(il) <i>chant-ai-t</i>	<i>cant-a-b-a</i>	<i>cant-a-v-a</i>
CANT-Ā-'BĀ-MUS	<i>chant-ii-ens</i>	(nous) <i>chant-i-ons</i>	<i>cant-á-b-amos</i>	<i>cant-a-v-amo</i>
CANT-Ā-'BĀ-TIS	<i>chant-ii-ez</i>	(vous) <i>chant-i-ez</i>	<i>cant-á-b-ais</i>	<i>cant-a-v-ate</i>
CAN'T-Ā-BA-NT	<i>chant-ei-ent</i>	(ils) <i>chant-ai-ent</i>	<i>cant-a-b-an</i>	<i>cant-a-v-ano</i>

In fact, the entire paradigm of the French past imperfect is based on regional Proto-Romance *(CANT-)Ē-BAM rather than on Classical Latin (CANT-)Ā-BAM; in contrast, the latter has survived in Spanish, Italian and Romanian.

Remarks:

- the Spanish forms fully comply with natural patterns of phonetic evolution (with the exception of the form ending in *-ais*, as mentioned above, 7.6.1); the temporal affix *-b-* is stable, and number and person are distinct (with the exception of 1SG and 3SG, for which the distinction has been neutralised). Spanish thus constitutes an example of stability with regard to the evolution of a Latin paradigm towards the Romance languages. Nevertheless, it may be observed that the past imperfect marker is polymorphic depending on inflectional class (*cant-a-b-a* vs. *ven-t-a*)
- Italian also remains close to the Latin paradigm, displaying a regular affix (*-v-*) which marks the past imperfect. From an evolutionary point of view, 1SG and 2SG display an analogical inflectional suffix (*-o* / *-i*), which maintains the distinction between all three persons in the singular
- by way of contrast, the marking of tense is less apparent in French, though it remains present (*-ai-* ~ *-i-*). The problem of number and person distinction characteristic of the IND PRES can also be observed here
- from a phonetic point of view, French 1SG, 2SG, 3SG and 3PL forms display an unusual simplification of the diphthong [we] > [ɛ], already present in the medieval period; from an orthographic point of view, 1SG adopted the form of the 2SG, and both were subsequently partially remodelled on the 3SG form *-ait* after the loss of final *-e* in the spoken language (*-oie/-oies* > *-ais*).

This example shows that the Romance past imperfect has maintained a strong continuity with regard to Latin and consequently displays a high degree of inflectedness. Bossong characterises it as an ‘inflecting island in a sea of agglutination’, with the latter principle being represented by the Romance verb paradigms for the other tenses, many of which are periphrastic (cf. LangTyp, art. 48: 665).

7.6.4 Future

The future is the most salient example of a tense that originated in a periphrastic construction (formed by means of an auxiliary verb, cf. 2.3.2 no. 3 and 2.3.3 no. 2); an initial Latin periphrasis of the type ‘CANTARE HABĒO’ 1SG (“I have to sing”) eventually fused to form new, single forms in the Romance languages (e.g. *chanterai*, Sp. *cantaré*, It. *canterò*). Thus, the non-periphrastic Latin future tense (e.g. CANTABO 1SG) was replaced by periphrastic constructions consisting of the infinitive of the verb in question and an auxiliary.

In the majority of the Romance languages, the infinitive of the main verb is followed by the reflexes of the auxiliary HABĒRE conjugated in the present tense; however, during the (Late) Latin period, this type of periphrasis was in competition with other types, which, in some cases also survived into the Romance languages: VOL(E)Ō CANTARE (“I want to sing”) (→ Romanian), DĒBEŌ CANTARE (“I must sing”) (→ Sardinian), VENIŌ AD CANTARE (“I’m coming to sing”) (→ Romansh). From a functional point of view, the future tense in today’s Romance languages essentially corresponds to its Latin counterpart, although the initial semantic and pragmatic connotations inherent in the original periphrastic construction are no longer present (cf. 5.1.3 and, once again, 2.3.3 no. 2).

The following table displays the most widespread type of Romance periphrastic future forms:

Late Latin	Proto-Romance	Old / Modern French	Spanish	Italian
CANTAR(E) + (H)ABEŌ	*cantar-aio	(je) chanter-ai	cantar-é	canter-ò
CANTAR(E) + (H)ABĒS	*cantar-as	(tu) chanter-as	cantar-ás	canter-ai
CANTAR(E) + (H)ABET	*cantar-at	(il) chanter-a	cantar-á	canter-à
CANTAR(E) + (H)ABĒMUS	*cantar-emos	(nous) chanter-ons	cantar-emos	canter-emo
CANTAR(E) + (H)ABĒTIS	*cantar-etes	(vous) chanter-e(i)z	cantar-éis	canter-ete
CANTAR(E) + (H)ABENT	*cantar-aunt	(ils) chanter-ont	cantar-án	canter-anno

An easily recognisable Latin periphrasis provides the starting point. This is distinguished from a series of so-called ‘Proto-Romance’ forms (cf. 5.2.2), which have been established by reconstruction methods, and which have undergone a certain degree of phonetic evolution: for example, the plural forms have been weakened (e.g. *cantar-av-emos has become *cantar-emos). The degree of phonetic evolution which occurred between the forms of the Proto-Romance paradigm and the modern Romance forms appears to be relatively weak; at the same time, the pervasive influence of analogy can be observed:

- in Fr., 1PL is always formed on the basis of the suffix *-ŪMUS; moreover, 2SG and 3SG are homophonous, as are 1PL and 3PL;
- the Sp. forms once again correspond to the tendencies of regular phonetic evolution, retaining the distinction of both number and person; however, the grammaticalised auxiliary could still be separated from the main verb in O.Sp. and O.Occ. (in the latter, a clitic pronoun could be intercalated between the auxiliary and the main verb: e.g. *pedir vos ha* “he will ask you” (*Poema del Cid*); in modern European Pg. this separation still occurs in higher registers: *queixarei* vs. *queixar-me-ei(de)*).

The future paradigm is similar to that of the conditional, which was formed by means of the auxiliary HABĒRE in the past imperfect tense (cf. 2.3.3 no. 3). A number of further periphrastic forms have been introduced in Romance; these employ the same auxiliary,

but it has not fused with the main verb as it has in the case of the future and conditional (consider the French forms *j'ai chanté*, *j'avais chanté*, *j'aie chanté*, *j'eusse chanté*, *j'aurai chanté*, cf. 8.3.3).

- Hinzelin, *Inflection classes in verbs in the Romance languages*, OxfEnc
 Esher/Florici/Maiden, *Finite verb morphology in the Romance languages*, OxfEnc

8 Syntax

8.1 The study of syntax

8.1.1 Major approaches to syntax

Whereas morphology deals with the internal structure and formation of words, syntax is concerned with the combination of lexical and grammatical elements into larger formal and functional units, called ‘phrases’, which, in turn, are the building blocks of clauses and sentences.

Syntax is not an easily accessible module of language. Students often consider it particularly complex and difficult to comprehend, and with good reason: it is at the level of syntax that thoughts organised in the brain are transposed into temporal-linear utterances in speech. Information that is stored in parallel or hierarchically in the human cognitive apparatus must be expressed as a sequence of elements, a process that necessarily leads to strong constraints (cf. 8.4.1).

The summary of the principles governing syntax presented here is intended to reflect a certain degree of consensus established within the discipline. However, from structuralism to generative linguistics and beyond, the proponents of numerous frameworks have proffered their particular interpretation of grammatical phenomena.

Situated at the margins of structuralism, ‘dependency grammar’ (cf. Tesnière, *Éléments de syntaxe structurale*, 1959) relies on the assumption that the individual word is always in a relationship of dependency with another word in the sentence (‘head-dependent’ relationship) and puts the main emphasis on the relationships between predicates and their arguments. Generative linguistics, on the other hand, is based on a type of ‘constituency grammar’ (cf. Chomsky’s ‘Phrase-Structure Grammar’, *Syntactic Structures*, 1957) and is founded on the principle of hierarchical sentence structures that can be broken down into smaller units known as ‘constituents’. More recently, the term ‘construction grammar’ has been used to refer to a series of more interpretative (as opposed to descriptive) syntactic theories (cf. e.g. Fillmore *et al.*, *Regularity and idiomativity in grammatical constructions*, 1988). The holistic perspective adopted by these approaches, which exceeds the strictly syntactic framework of linguistic constructions, puts them at odds with generative grammar, whose sole focus is syntax, excluding lexis from the ‘faculty of language in a narrow sense’ (cf. 5.1.1; cf. also Croft, *Radical Construction Grammar*, 2012).

This chapter is essentially based on functional theories of syntax, but also introduces a number of concepts from dependency grammar and generative linguistics. The argumentation and examples thus come from a variety of methodological approaches, beginning with the classic contributions of Oesterreicher (LRL), Raible and Bossong (LangTyp), the articles on grammar in the RSG and La Fauci’s *Compendio di sintassi italiana* (2009). For more recent approaches, the CambrHist and, to a greater degree, the

OxfGuide, the CambrHandb and the OxfEnc), were consulted, and references to articles from these volumes are systematically provided in the suggestions for further reading at the end of each subsection.

8.1.2 Observations relating to sentence structure

The sentence is an autonomous syntactic unit. It forms a hierarchical whole, made up of so-called ‘constituents’, which can be defined as meaningful entities that possess a relative functional autonomy within the sentence structure. Whereas lexemes verbalise discrete entities of meaning (cf. 9.2), syntax establishes the links between them.

The sentence hereby generates a statement (or describes a situation), which establishes a relationship between a predicate and its arguments (noun phrases, in particular), thus creating a ‘verb phrase’. Other terminology employed for the syntactic statement are ‘denotative content’, ‘assertive content’ or ‘assertion’ (the latter is usual in French, but in English implies a speech act by which the speaker commits him-/herself to the truth of his/her statement). In other words, the establishment of an (assertive) relationship between a predicate and a number of arguments expresses an extralinguistic ‘state of affairs’ (cf. the German concept of *Sachverhaltsausdruck* developed by Wittgenstein).

Linguistic philosophy assumes that assertive sentences express contents that may be true or false. This implies that ‘true’ and ‘false’ are predicates of sentential contents, just as ‘red’ or ‘blue’ are predicates of objects that can be seen, and ‘sick’ is a predicate of animate beings. The sentence thus constructs the category of ‘truth’; in other words, the concepts of ‘true’ and ‘false’ cannot exist outside of the sentence. Following this logic, the syntactic statement, predicate and arguments become ‘tools with which we perceive the world, comparable to the retina and the optic nerve’ (Bossong, *Ausdrucksmöglichkeiten für grammatische Relationen*, LangTyp, art. 48: 658).

The syntactic relationship between the predicate (P) and its arguments (A1, A2, A3) can thus be conceived of as a ‘fundamental relationship’ (cf. Raible, LangTyp, art. 1: 15). More specifically, the establishment of such a relationship creates a ‘proposition’, which, at the same time, introduces a temporal dimension:

present: *Notre chat*_{A1} *dort*_P / *Our cat*_{A1} *is sleeping*_P.

vs.

past: *Paul*_{A1} *a vendu*_P *sa maison*_{A2} / *Paul*_{A1} *sold*_P *his house*_{A2}.

In actual sentences, complementary information is often added to the ‘bare’ proposition, allowing the latter to be integrated into its syntagmatic and pragmatic context. Thus, the propositions in the two examples given above have the ability to function as autonomous sentences; in order to be linked to a concrete context, however, they require specification by means of circumstantial adjuncts (C):

Chaque jour_C, notre chat_{A1} dort_P dans les endroits les plus bizarres_C pendant des heures_C.

“Every day_C, our cat_{A1} sleeps_P in the strangest places_C for hours and hours_C.”

Malgré les objections de ses parents_C, Paul_{A1} a vendu_P sa maison_{A2} la semaine passée_C.

“Despite his parents’ objections_C, Paul_{A1} sold_P his house_{A2} last week_C.”

In terms of formal syntactic categories the semantic-logical proposition, formed by a predicate (e.g. *vendre*), its arguments (*Paul*; *sa maison*) and circumstantial adjuncts (*malgré les objections* ...) corresponds to a ‘clause’, which is made up of different types of constituents (in this case, a verb phrase, noun phrases and a prepositional phrase).

In other words, clauses are the smallest units of syntax that are capable of expressing complete propositions, apart from pro-sentence forms such as *no!* or *o.k.!*. Clauses may stand alone as simple sentences (in which case one speaks of ‘independent clauses’), as in the examples given above. However, several clauses may also be combined within a sentence: they may be coordinated in order to form a compound sentence or subordinated to form a ‘complex sentence’ (cf. 8.4.5).

Finally, the number of arguments required by a specific predicate is predefined by its ‘valency’ (cf. 8.3.2 no. 6.2); circumstantial adjuncts, on the other hand, are subsidiary elements which can be easily added, deleted and displaced since they do not form part of the predicative core of the sentence. In some theoretical approaches, these elements are nevertheless included in a ‘broad’ valency frame.

The syntactic definition of the sentence according to the model of assertion has a long tradition in linguistic thought, extending all the way back to Aristotle. Parallel, rather than alternative, definitions place greater emphasis on the aspects of ‘predication’ or on ‘utterance modalities’. The former corresponds to a semantic interpretation of the proposition: by predication, a sentence typically attributes properties to a subject (in the sentence *Paul sold his house*, information is given about the subject, *Paul*); from this perspective, the ‘predicate’ (*sold his house last week*) encompasses not only the verb (*sold*) but also the arguments and circumstantial adjuncts linked to the verb (*his house, last week*), with the exception of the subject.

Moreover, every sentence incorporates a type of ‘utterance modality’, corresponding to a pragmatic dimension. The utterance modality may be assertive (or ‘declarative’), interrogative, exclamatory or injunctive (*help me!*).

The subject matter treated in this chapter is based for the most part on the assertive model (cf. 8.4.1). It encompasses, first and foremost, the two most salient types of constituents which occur in the propositional framework (i.e. in the syntactic clause): noun phrases (NP), corresponding to arguments, and verb phrases (VP), corresponding to

predicates. A number of fundamental questions more directly linked to the sentence will then be addressed.

- Ramat/Ricca, *Romance. A typological approach*, OxfGuide 5
 Ledgeway, *Functional categories*, OxfGuide 46
 Ledgeway/Maiden, *Data, theory, and explanation: the view from Romance*, CambrHandb 1
 Corr/Munaro, *Speech acts, discourse, and clause type*, CambrHandb 24
 Munaro, *Interrogatives in the Romance languages*, OxfEnc
 RSG 3, sect. XV, *Histoire interne: morphosyntaxe et syntaxe (roumain*, Livescu, art. 223; *Bündner-romanisch*, Liver, art. 229b; *italiano*, Renzi, art. 231; *français*, Marchello-Nizia, art. 236; *catalán*, Pérez Saldanya, art. 242; *Spanisch*, Stein, art. 246; *Portugiesisch*, Wesch, art. 252; *phénomènes de convergence et de divergence dans la Romania*, Iliescu, art. 257)

8.2 The noun phrase (NP) in the Romance languages

8.2.1 Semantic and syntactic roles

The structuring of arguments within a proposition (or syntactic clause) combines two orders of logic: that of syntactic roles, which operates a primary distinction between subject and object, which may be expressed by means of the linear order of these elements or through additional grammatical markers (*cf.* 8.4.2); and that of hierarchical semantic roles, which primarily distinguishes an ‘agent’ (AGT), or the instigator of an action, and a ‘patient’ (PAT), or the entity affected by an action, as illustrated by the proposition *L’homme_{AGT} a soulevé une pierre_{PAT}* ‘The man lifted a stone’.

For syntactic roles, a further distinction is made between ‘first object’ (O1) and ‘second object’ (O2), which, in many cases, correspond to ‘direct’ object and ‘indirect’ object respectively (consider the following example: *Pierre_s a donné un cadeau_{O1} à Marie_{O2}* ‘Pierre gave a gift to Mary’).

The mapping of semantic and syntactic roles is complex; both can be combined fairly freely, although certain tendencies and restrictions may be observed. Thus, the syntactic subject often expresses the semantic agent, the 1st object often corresponds to the patient, and the 2nd object to the ‘recipient’ or ‘beneficiary’ of an action (*cf.* 8.4.3 no. 2). On the other hand, some role combinations appear incompatible: the semantic agent, for example, is not usually expressed by the syntactic role of the 1st object.

The encoding of syntactic roles underwent a radical change from Latin to the Romance languages: whereas in Latin arguments are marked by means of inflection, in the Romance languages, they are mainly encoded by agglutinative grammatical markers, as well as by their position in the sentence (*cf.* 7.1.2 no. 1). This transformation was accompanied by other phenomena relating to the restructuring of the noun phrase, as well as to the general shape of the sentence.

8.2.2 The noun phrase (NP) and its components

The noun phrase (NP) is structured around a central nominal element (i.e. a noun or pronoun), or ‘head’, which can be supplemented with elements of determination and modification. These elements form a continuum, which is further expanded by qualifying adjectives, relative clauses, and, above all, determiners. The function of the latter is to specify the reference of the head noun in the discourse and in relation to the extralinguistic context (this is known as ‘actualisation’); qualifying adjectives and relative clauses, on the other hand, fulfil the primary function of modifying (or ‘characterising’) the head noun and its semantic content. Thus, whereas the main elements of determination are grammatical in nature, modification is based on lexemes as well as on lexeme-based constructions.

Within the individual clause, the NP may fulfil the function of an argument (e.g. as the subject: Fr. *Les enfants dorment* “The children are sleeping”), of a predicate (e.g. as an attribute: Fr. *Ce sont des enfants* “These are children”), or it may act as a combination of the two in so-called ‘existential’ constructions (e.g. It. *Ci sono bambini*, Rom.n *Sunt copii* and, less usual, Fr. *Il y a des enfants* “There are children”).

→ Giusti, *The structure of the nominal group*, OxfGuide 30

8.2.3 Determination

In most cases, the nominal head of an NP is accompanied by elements of determination, which are also responsible for its specification (the argument *ses deux chiens* “his/her two dogs”, for example, encompasses the noun *chiens*, the possessive determiner *ses* and the cardinal number *deux*)²².

In Latin, nominal determination involves the parameter of number (*gladius imperatoris* SG “the/a sword of the/an emperor” vs. *tres infantes* PL “three infants”); it does not, however, include a grammaticalised distinction between definite and indefinite referential determiners (such as that displayed by Fr. *le* vs. *un*) or a system of partitive determination (compare Lat. *Bibit vinum* and Fr. *Il boit du vin* “He drinks wine” vs. *Il boit le vin* “He drinks the wine”). These forms of determination were introduced into the

²² Proper nouns adhere to a different logic, since they are intrinsically referential (cf. 9.1.2 and 9.7.1); consequently in many Romance languages and dialects they do not require an article nor any other type of determiner (cf. La Fauci, *Compendio*, 2009: 139, who speaks of ‘*nomi nudi*’, i.e. ‘bare nouns’); however the article is obligatory in others, such as in Romanian (e.g. *Mari-a*, where *-a* is the f. sg. definite article), European Portuguese (e.g. *o Carlos / a Ana*) or in dialects of southern Italy (e.g. Salentino *Lu Luca / La Rita*); cf. also the ‘personal’ article in Balearic Catalan (e.g. *en Joan* “John” / *na Joana* “Johanna”, < DOMINUS IOHANNES / DOMINA IOHANNA).

Romance languages by means of the article, an innovation that can also be observed in most other Indo-European languages:

SG Fr. <i>le loup</i> , Sp. <i>el lobo</i> , It. <i>il lupo</i>	vs. <i>un loup, un lobo, un lupo</i> ;
PL Fr. <i>les loups</i> , Sp. <i>los lobos</i> , It. <i>i lupi</i>	vs. <i>des loups</i> (single form for both indef. and part.), <i>lobos / unos lobos, lupi / dei lupi</i> ;

In Romanian, the definite article (but not the indefinite article) is both postposed and inflectional (*cf.* below): *lupul, lupii*.

Romance definite articles evolved from Latin demonstrative pronouns (ILLE in the case of Fr. and the majority of the other Romance languages), while indefinite articles evolved from the Latin numeral UNUS. For a long period of time – until their respective grammaticalisation processes were completed – these parallel forms were employed to specific semantic and pragmatic ends; they served, above all, to express specificity and individuation.

The definite article emerged around 700 AD and gradually became established (*cf.* 5.2.1). In Old French, its expression was optional and its forms consequently retained semantic connotations of a deictic or accentuating nature, in addition to their initial value. The article became obligatory in Modern French (with some exceptions, e.g. *avoir faim* “to be hungry”, literally, “to have hunger”), where it partially compensates the loss (in speech) of the inflectional nominal markers of gender and number, which had disappeared as a result of phonetic change. In the other Romance languages, the article is subject to fewer constraints and can be omitted under certain conditions²³.

Romanian exhibits the particularity of a postposed article, which in the past was generally interpreted as ‘enclitic’. However, there is considerable evidence that it is to be considered ‘suffixal’ and thus inflectional in the proper sense (*cf.* Ledgeway, *The Romanian definite article in a comparative Romance perspective*, 2017).

In French, the absence of the ‘zero article’ is partially correlated with the use of partitive articles (*du, de l’, de la*), which express the neutralisation of number distinction for non-count nouns (or mass nouns): *du vin* may represent *un vin* “a (specific) wine” or *des vins* “(different) wines”.

→ Bossong, *Classification*, OxfGuide 6.4.2 [‘The partitive’]
 Carlier/Lamiroy, *Partitive articles in the Romance languages*, OxfEnc
cf. also 8.2.4 below for the function of the article in reduced NPs.

²³ Aside from proper nouns (*cf.* n. 22 above), the omission of the article is possible in phrases involving abstract nouns, with nouns following prepositions as well as with noun phrases in postverbal position, such as the indefinite plural in the position of an object complement (e.g. Sp. *también invitó extranjeros*, It. *ha invitato anche stranieri* vs. Fr. *il a aussi invité des étrangers* “he also invited (some) strangers”). It can also be omitted before nouns in subject function with unaccusative verbs (e.g. Sp. *Han muerto niños* “(some) children have died” vs. **Niños han muerto*).

8.2.4 Modification

In addition to determiners, noun phrases may also include (optional) elements of modification. A noun functioning as the head of an NP can be modified in various ways: by qualifying adjectives (It. *una strada stretta* “a narrow road”), by relative constructions (Sp. *el hombre que está hablando* “the man who is speaking”) or by other nominal elements, which, however, require a preposition (les *photos de nos amis* “the photographs of our friends”).

The position occupied by the various elements within the NP is fixed, to a greater or lesser extent. In the Romance languages, modifying elements usually follow the modified element:

Fr. *C'est un garçon sympathique* “He’s a nice boy” [but also: *Ces trois belles maisons* “These three lovely houses”]

Sp. *Ha preparado una cena estupenda* “He prepared a wonderful supper” (cf. 8.4.2 for constituent order).

Diverging configurations are the exception. In Romanian, for example, the presence of the postposed definite article causes the modifying elements to be displaced to the right:

<i>bun-ul</i>	<i>prieten</i>	→	<i>prieten-ul</i>	<i>bun</i>	“the good friend”
good DEF.ART.M.SG	friend M.SG		friend DEF.ART.M.SG	good M.SG	
<i>cel mai bun prieten al meu</i>	→	<i>prietenul meu cel mai bun</i>	“my best friend”		
the most good friend of mine		the friend mine the most good			

In medieval Romance languages, some qualifying adjectives or nominal determiners frequently precede the determined noun; cf. the examples for Old French:

<i>sa blanche barbe tîret</i>	“he pulls at his white beard”
<i>le rei cort</i>	“the king’s court”
<i>Dieu merci</i>	“by the grace of God”

(cf. Oesterreicher, *LRL* 2/1, art. 107, p. 289).

Whereas modifiers typically follow the modified element, the position of qualifying adjectives additionally depends on their semantic content. If they express a predominantly identifying value, they are more likely to be placed to the left (this is also the case for judgements relating to value or size); if a descriptive or characterising element prevails, they are placed to the right. The following examples illustrate this phenomenon:

Sp. *nueva botella* vs. *botella nueva*. The first example means “another bottle”, whereas in the second, the object is characterised as being “(brand) new”;

Sp. *pobre mujer* vs. *mujer pobre*. The first example illustrates a value judgement, as in the exclamation “the poor woman!”; the second characterises the head noun as “lacking financial means”.

Cf. also *les trois belles maisons blanches*, Sp. *las tres hermosas casas blancas*, It. *le tre belle case bianche* “the three pretty white houses”, O.Fr. *sa blanche barbe tiret* “he pulls at his white beard”.

Divergence from these rules has the effect of generating specific meanings, usually involving an increase in emphasis or expressivity:

Sp. <i>ese hombre</i>	vs. <i>el hombre ese</i> “that man”
<i>no tiene ningún motivo</i>	vs. <i>no tiene motivo ninguno</i> “there is (absolutely) no reason”

At the level of the phrase, qualifying adjectives may be considered as analogous to subordinate relative clauses at the level of the sentence (compare *les enfants heureux* “the happy children” and *les enfants qui sont heureux* “the children who are happy”)²⁴.

The NP may also be modified by other nominal elements:

- The position of the qualifying adjective may be occupied not only by adjectives, but also by nouns, which immediately follow the determined element (e.g. *tarte maison* “home-made tart”, *conflit Iran-Iraq* “Iran-Iraq conflict”, *opposition verbe-nom* “verb-noun distinction”). In these cases, the modifying noun(s) add(s) an element of characterisation relating to the identity of the modified noun.

Similarly, N-N appositions are also frequent, often consisting of a common noun combined with a proper noun: e.g. *le président Mitterand* “President Mitterand”, *la Cathédrale Notre-Dame* “Notre-Dame Cathedral”, It. *la Piazza San Marco* “Saint Mark’s Square”, *il volcano Etna* “Mount Etna”, Sp. *el río Ebro* “the river Ebro”, *el planeta Venus* “the planet Venus”.

Formally, such constructions resemble compound nouns of the type *chou-fleur* “cauliflower” (which is not actually a ‘cabbage / Fr. *chou*’, but a different species of Brassica).

- As we have seen, prepositional constructions may also function as noun complements (e.g. *la maison de mon père* “my father’s house”); such complements constitute separate noun phrases introduced by prepositions. NPs of the type $N_1 + de +$

²⁴ It is necessary to distinguish between restrictive (or ‘defining’) and non-restrictive (‘non-defining’) relative clauses: *mes cousines qui habitent la campagne sont venues* “my cousins who live in the countryside have come” (restrictive) vs. *mes cousines, qui habitent la campagne, sont venues* “my cousins, who live in the countryside, have come” (non-restrictive). In spoken language, these two types of relative clauses may be marked by a different prosodic contour.

N_2 are extremely frequent in the Romance languages (e.g. Fr. *maison de campagne* “country house”, *tremblement de terre* “earthquake”, Sp. *el mes de mayo* “the month of May”, *sala de espera* “waiting room”). Besides the common structure expressing possession, they can also express various other kinds of relations (e.g. Fr. *la ville de Rome* “the city of Rome”, *ce fripon de valet* “that rascal of a valet”)

- O.Fr. and O.Occ. display a specific form for this construction (the ‘absolute oblique case’): O.Fr. *la maison mon pere* (= *de mon pere*), “my father’s house”. This type survived into Modern French in the form of relics such as *hôtel-Dieu* (= *hôtel de Dieu*) “clerical hospital” or the town name *Château-Thierry* (= *château de Thierry*); cf. also the example of *le rei cort* (above), and it is still present in various dialects of southern Italy where the possessor is animate and highly referential (e.g. equivalents of *la casa il sindaco* “the mayor’s house”).

Modified NPs can be reduced under certain conditions: the nominal element may be omitted, provided that it has been previously identified in the discourse:

- in constructions that involve a specifying nominal determiner: Fr. *les deux pays* “the two countries” → *les deux*, Sp. *los dos países* → *los dos*
- in the presence of a modifying (characterising) adjective: Fr. *le chien noir* “the black dog” → *le noir*, Sp. *el perro negro* → *el negro*. The article in these reduced constructions does not correspond to an actual pronominal form (such as Sp. *lo negro* “that which is black, blackness”) but assumes the syntactic role of a nominal element
- in Sp. this also applies to NPs modified by prepositional phrases: *la casa del vecino* “the neighbour’s house” → *la del vecino*. This is not possible in Fr., which employs a pronominal form in such cases: *la maison du voisin* → *celle du voisin*.

Finally, three further observations may be made:

1. Adjective phrases (AP) follow the same rules as noun phrases: Romance adjectives can be determined by elements placed before the adjectival head (e.g. *trop rapide*) or following it (e.g. *le plus beau de mes rêves* “the best dream I’ve (ever) had”); they may be modified by elements placed after the phrase head (*plein d’eau* “full of water”, *content de rester à la maison* “happy to stay at home”), which function as complements of the adjective.
2. An NP may also be replaced by a pronoun (*Son frère aîné est parti* “His older brother has left” → *Il est parti* “he has left”). The pronominalisation thus does not only reference the head noun (*frère*), but also the components that determine (*son*) and modify (*aîné*) the nominal.
3. Furthermore, Romance pronouns may take their own modifiers, which are then generally placed to the right: Fr. *moi seul* “only me”, Sp. *yo mismo*, It. *io stesso* “myself”/ Fr. *rien de plus étrange* “nothing stranger”, Sp. *nada de lo que tú dices* (*me conviene*), literally “nothing of what you say (suits me)”, It. *niente di nuovo* “nothing

new”. Placement to the left, by contrast, is marked: It. *solo lui* “only he/him” vs. *lui solo* “he/him alone”, *tutti noi* “all of us” vs. *noi tutti* “us all”.

8.3 The verb phrase (VP) in the Romance languages

8.3.1 Components and structure of the VP

A verb phrase (VP) may be simple (Sp. *escribe* “he/she writes”) or complex (Sp. *no ha estado escribiendo mucho tiempo* “he/she hasn’t been writing for very long”). Although it may function as a sentence in its own right, it generally functions as the central element within a sentence. The verb phrase consists of the head verb (or the ‘predicate’) and its modifiers; generative grammar also includes all its complements. Even when considered according to the narrow definition (i.e. the head and determiners only), the verbal event generally conveys a considerable amount of information, at least in the Indo-European languages. The VP may encode the basic verbal categories (person, number, cf. 7.3.1 and 7.3.2) and the morphosyntactic features expressed by the predicate (tense, mood and aspect as well as voice, cf. 8.3.2), also encompassing its immediate environment (pronouns, cf. 7.5) and its optional determiners: auxiliary and other functional verbs creating periphrases (Pg. *devo cantar*, cf. 8.3.2 no. 1 below) – or expressing voice, diathesis (no. 5) and ‘action moods’ (no. 6) – as well as elements of negation (Sp. *nunca canta*, no. 7) and adverbs (Fr. *vendre cher*, *ibid.*). The position of these elements is fixed to a greater or lesser extent.

The VP thus specifies the persons involved in the action, as well as the content of the utterance act and that of the denoted event or situation, while at the same time placing the latter elements in a spatial-temporal context. By means of its inflectional elements and valency frame, the predicate exerts a great deal of influence on the construction of the sentence as a whole (cf. 8.3.2 no. 6). Moreover, information linked to the lexical meaning of the verb also comes into play (no. 6.4). Grammatical and lexical information thus interact to a large extent within the verb phrase.

The different Romance languages display a great deal of variation in their deployment of verbal morphosyntactic features, particularly those relating to mood, tense and (passive) voice. These phenomena of variation pertain to a universal level, which also determines the relevance and proximity of the different features in relation to the verb. The relevance hierarchy as described by Bybee recognises the following order of decreasing relevance:

- direct agreement relationships (person and number)
- mood, which concerns the entire context of utterance
- tense and aspect

- diathesis and verb valency, which, in the Romance languages, concern auxiliaries in particular
- negation

In some cases, the different categories fulfil complementary or parallel functions, making detailed classification difficult. Due to the large amount of grammatical information attached to verbs, an empirical description quickly encounters problems as a result of phenomena of interference and ellipsis, especially in diachronic and variational syntax (cf. e.g. Loporcaro, *Grammatica storica del dialetto di Altamura*, 1988, or Ledgeway, *Grammatica diacronica del napoletano*, 2009).

→ Sheenan, *Complex predicates*, OxfGuide 61

Dragomirescu/Nicolae/Pană Dindelegan, *Complex predicates*, CambrHandb 19

8.3.2 Grammatical categories of the predicate: TAM

The system of tense, mood and aspect is responsible for the anchoring of a verbal event or situation within the utterance context. The three dimensions are interdependent and are capable of fulfilling similar functions; nevertheless, they remain distinct in the different languages of the world.

→ Gosselin, *Sémantique de la temporalité*, 1996; *Temporalité et modalité*, 2005

1 Tense

The linguistic parameter of tense is correlated with the extralinguistic dimension of time; the latter, however, remains a relative factor and constitutes a series of relations. As stated by the GMF, “The moment in which a person speaks marks the origin of the process; it establishes the concept of the ‘present’.” (2009: 514 *sqq.*, transl. from French). The moment of utterance thus provides a first point of reference, which is then placed in relation to the verbal event. In this way, the three temporal dimensions of present, past and future, and their relationships to the moment of utterance – of simultaneity, anteriority and posteriority, respectively – are established. These three basic dimensions thus constitute the primary framework of temporality.

The categories of this primary framework in the Romance languages do not fully coincide with those present in (Classical) Latin; they have undergone several significant changes, which include:

- the replacement of the simple future forms by periphrastic forms, already present in Late Latin (*cantabo* → *cantare habeo* > *chanterai*; cf. 7.6.1 no. 4 and 8.3.3 no. 2);
- the development and grammaticalisation of a compound past perfect (in French ‘*passé composé*’: *j’ai chanté*) in addition to the Latin synthetic perfect (*cantavi* > Fr.

'*passé simple*' *je chantai*); the latter has been weakened in usage in several Romance languages, but preserves a fully functional and productive paradigm to this day (e.g. all Spanish varieties, Brazilian and European Portuguese, Galician, Leonese, dialects of southern Italy, southern Romanian varieties, most southern Occitan varieties);

- the development of the conditional (which has a modal use and also expresses a temporal value of 'future in the past', cf. 8.3.3 no. 3);
- the development of 'double compound' past forms (e.g. the Fr. '*passé surcomposé*': *j'ai eu chanté*).

In the Romance languages, the three dimensions of the primary framework have become established in a form that also specifies the internal progression of an action: we distinguish the representation of perfective action (compound past ~ simple past – future) from that of non-perfective (or 'non-actual') action (past imperfect – conditional). This distinction represents a secondary framework corresponding to the value of aspect, which will be treated in the following section.

→ Maiden, *Inflectional morphology*, OxfGuide 27.4

Bertinetto/Squartini, *Tense and aspect*, OxfGuide 58

2 Aspect

The aspectual values expressed by Romance tenses, which relate to the secondary framework of temporality, form a complementary system to that of tense. Aspect highlights the internal progression of an action as well as the point of view of the speaker.

With regard to the internal progression of an action, a perfective aspect is distinguished from an imperfective (or unaccomplished) aspect. In the Romance languages, this distinction becomes apparent above all in the past tenses (where it is often falsely interpreted as a category of tense).

- the past imperfect paradigm describes an action in progress at some point in the past, without consideration of its temporal limits (*je lisais* / Sp. *leía* / It. *leggevo*);
- the compound past and the simple perfect describe a completed action (*j'ai lu* – *je lus* / Sp. *he leído* – *léí* / It. *ho letto* – *lessi*); in a certain sense, the accomplished aspect thus coincides with the post-terminal phase of an action (cf. no. 5.1 below).

A further contrast may be observed between simple and compound past tense forms, which concerns the point of view of the speaker and his or her relation to the tense expressed in the sentence. Simple forms in standard Peninsular Spanish, Italian or French do not show any relationship with the present and are part of the primary temporal framework (Sp. *leí*, It. *lessi* "I read [past simple] (at a given moment)" [referring to a completed episode in one's life]). Compound forms, in contrast, are connected to the present: they belong to the secondary framework and convey the post-terminal phase

of an action (Sp. *he leído*, It. *ho letto* “I have read (the book, and now I know what it is about)”).

This distinction is weak, however, and tends to disappear or to undergo transformation: it may, as in Italian, evolve into a contrast between different regional varieties (the simple form of the ‘*passato remoto*’ is more typical of the dialects and regional varieties of southern Italy than of the northern varieties), or, as in Spanish, between national varieties (the simple form of the ‘*indefinido*’ is much more frequent in the Hispanic countries of the Americas than in the Iberian Peninsula). In present-day French, the distinction has become one of register (the *passé simple* is characteristic of high register and literary language, and is almost never used in normal speech).

3 Mood and modality

Like aspect, mood primarily concerns the point of view of the speaker (the source of an utterance) and determines his or her involvement in the utterance act. Moreover, in the Romance languages, modality in grammaticalised form plays a significant role in complex sentence constructions, where the subjunctive mood stands in contrast to the indicative mood, functioning as a marker for subordinate clauses (cf. 8.4.5).

The subjunctive constitutes the most striking grammatical representation of a modality that conveys multiple values; however, the conditional may also assume a modal role, as can some indicative tenses such as the future and the past imperfect. Other processes of modalisation exploit lexical means, such as modal verbs or modal particles.

A large number of definitions (both semantic and formal) have been suggested for the subjunctive mood, though no holistic definition has become established to date. The use of the subjunctive is correlated with a diverse array of pragmatic-semantic and syntactic elements. Sometimes it stands in true semantic or syntactic opposition to the indicative, while it is a redundant marker in cases where sentence modality is already present in the meaning of the main verb (cf. for example Bally 1944) or where its use is induced by grammaticalisation, especially when it appears in the role of a marker of subordination. In cases where it stands in true contrast to the indicative, the subjunctive mainly expresses imagined or potential action. It thus represents a ‘suspension of the category of truth within the proposition’, according to Confais (*Temps, mode et aspect*, 1990).

More recently, Gévaudan has suggested interpreting different moods as grammaticalised markers of modalisation; modalisation would thus be a phenomenon of ‘linguistic polyphony’, which describes the ability of an utterance to refer to multiple points of view (= ‘pv’, cf. *Les rapports entre la modalité et la polyphonie linguistique*, 2013). In the sentence *Marie pense [que Pierre est venu]* (“Mary thinks that Peter has arrived”), one may attribute a first assertion, ‘*Pierre est venu*’ (= pv1) to ‘*Marie*’, and a second, ‘*Marie pense [pv1]*’ to the speaker (who utters the sentence). The subordinate point of view (pv1) is represented syntactically by a subordinate clause. The indicative is used in

assertive sentences, whereas non-assertive sentences require the subjunctive (consider *Marie souhaite* [or: *ne pense pas* / *se réjouit* / *demande*] *que Pierre vienne* “Mary wishes that Pierre would come / M. does not think [or: is pleased] that P. will come / M. asks / is asking P. to come”). Finally, in both cases, the infinitive marks coreference between the grammatical subject in the main and subordinate clauses (*Marie croit* [or: *espère*] *venir demain* “Mary believe 3F.SG [hope 3F.SG] come INF tomorrow”).

→ Maiden, *Inflectional morphology*, OxfGuide 27.5; Quer, *Mood*, OxfGuide 59

4 Verbal periphrases

While the Romance languages maintain many characteristics of the Latin inflectional system, they also systematically exhibit a large number of innovative periphrastic constructions. These periphrases are used to convey various morphosyntactic categories:

- voice (in passive constructions such as: *La porte a été ouverte* “the door has been opened” vs. active: *il a ouvert la porte* “he has opened the door”)
- tense (e.g. compound past *elle est arrivée* “she has arrived / she arrived”, pluperfect *elle était arrivée* (“she had arrived”)
- aspect (e.g. present progressive: Sp. *estoy durmiendo* “I am sleeping” vs. simple present *duermo* “I sleep”)
- mood (by means of modal verbs, e.g. *il veut partir* “he wants to leave”, *elle peut chanter* “she can sing”)

The complementarity of inflection and periphrasis has become an essential characteristic of the Romance verbal system. Latin was a predominantly inflectional language, and although periphrases did exist, they were relatively infrequent (examples are the perfect passive *laudatus est* “he has been praised” and the future active participle *amaturus sum* “I am about to love”).

Verbal periphrases include all cases in which the main verb is auxiliarised; typically, the verb expressing the denoted situation or the event (the ‘genuine’ lexical predicate) appears in a specific non-finite form, accompanied by a freely inflected verb which semantically ‘modifies’ the VP. A few general rules should be mentioned (cf. also *Transitive and intransitive verbs*, no. 6.3 below):

- the main verb usually takes a non-inflected form (i.e. participial, infinitive or gerundive);
- the auxiliary verb may vary: HABERE, ESSE(RE) – STARE, TENERE; *VOLERE (for VELLE); VENIRE; FACERE;

- in modern Romance languages, the auxiliary is generally placed before the main verb (except in future and conditional forms, where it is an original Latin construction; cf. 7.6.4, as well as 8.3.3 no. 3 below);
- medieval languages also show a V-Aux construction (sometimes called stylistic fronting), e.g. in O.Sp. up to *La Celestina* (15th c.), or in O.Fr. *grant joie menee avoit* “he had shown great joy” (where, however, not only has the non-finite verb been fronted, but the entire VP, including its argument *grant joie*);
- sometimes, a preposition is used to link the auxiliary with the main verb (Sp. *estoy por hacer* “I’m going to do”, *voy a jugar* “I’m going to play”);
- exceptionally, a periphrastic construction can be formed with two finite verbs, in a kind of pseudo-coordination (e.g. Sicilian *vaiu (a) mmanciu* “I’m gonna eat”, cf. Ledgeway, *From coordination to subordination*, 2016).

When periphrastic formations first appeared, the auxiliary could easily be separated from the main verb (O.Fr. *jo ai paiens veüz* “I have seen pagans”, *Song of Roland*). Over time, and with the grammaticalisation of these constructions, a process of merging and partial fixation took place. Essentially, in present-day Romance languages only adverbs and negation particles can appear between auxiliaries and main verbs, and even then only with restrictions (e.g. in Ibero-Romance or Romanian, the auxiliary verb and the participle of analytic perfective forms can not be separated: Sp. *Siempre había cantado*, but not *había *siempre cantado*; in French and Italian, however, insertion is possible: It. *Aveva sempre cantato*).

Periphrases may semantically reinforce temporal *deixis* (cf. 4.5.3 no. 2), as in the case of the periphrastic future. They also specify modal elements or reinforce the progressive aspect (cf. Sp. *Juan está yendo a misa los domingos* “John usually goes to Mass on Sundays”, Pg. (Brazil): *O brasileiro está comendo menos melão espanhol e argentino* “Brazilians have become less used to eating Spanish and Argentinian melons” / Pg. (Portugal) *O brasileiro está a comer ...*).

The development of verbal periphrases in the Romance languages initially reflected a tendency towards greater transparency of grammatical information, as well as towards expressivity. As with compound demonstrative pronouns, the initial semantic connotations of the elements in question were later lost, and they were desemanticised during the grammaticalisation process.

→ Laca, *Non-passive verbal periphrases in the Romance languages*, OxfEnc
Ledgeway/Smith/Vincent (eds), *Periphrasis and Inflexion in Diachrony: A View from Romance*, 2022

5 Voice and diathesis

5.1 Passive voice and diathesis

The expression of voice (Lat. *genus verbī*), which can be active or passive, underwent extensive changes between Latin and Romance. Latin deponent verbs (corresponding to a ‘middle voice’ / *vox media*) disappeared and the passive was morphologically weakened, its functions being partially assimilated by reflexive periphrastic forms (cf. passive reflexive constructions such as Fr. *il s’est fait mal voir* ‘he put himself in a bad light’). In the Romance languages, these functions can also be adopted by simple periphrastic passive forms:

[active] *La jeune femme lui a volé son portefeuille* ‘The young woman stole his wallet.’

[passive reflexive] *Il s’est fait voler son portefeuille* ‘He had his wallet stolen.’

[passive periphrastic] *Son portefeuille lui a été volé* ‘His wallet was stolen.’

The difference between active and passive is linked to the status of the semantic agent: in the active voice, the expression of the agent is expected and generalised, whereas in the passive voice, it is optional:

active voice:	<i>La jeune femme</i> _{AGENT}	<i>lui</i> _{EXP}	<i>a volé</i>	<i>son portefeuille</i> _{PATIENT}
passive voice:	<i>Son portefeuille</i> _{PATIENT}	<i>lui</i> _{EXP}	<i>a été volé</i>	<i>(par la jeune femme</i> _{AGENT} <i>)</i>

When a sentence is transformed from active to passive voice, the expression of the agent is no longer obligatory. This equates to a reduction of verbal valency; cf. the following examples of partial and total active-to-passive transformation in Romance:

Sp. *el agente llama a Pedro* ‘The policeman is calling Peter’ → *Pedro es llamado por el agente* ‘Peter is being called by the policeman’ (partial transformation) → *Pedro es llamado* ‘Peter is being called’ (total transformation, as the number of arguments, or the verbal valency, is reduced)

It. *mia sorella canta la canzone* ‘my sister is singing the song’ → *la canzone è cantata da mia sorella* ‘the song is being sung by my sister’ → *la canzone è cantata* ‘the song is being sung’

The grammatical transformation of verb valency is generally labelled ‘diathesis’, a concept that refers to the number of arguments and their constellation (cf. the definition established by Tesnière: ‘sens dans lequel l’action exprimée par un verbe s’exerce d’un actant vers un autre’ / ‘the way in which the action expressed by a verb defines the relationship between its actants’, *Éléments de syntaxe structurale*, 1959: 667). The formation of the Romance passive (be it reflexive or periphrastic) thus represents a valency reduction. According to Tesnière’s terminology, this phenomenon is known as ‘recessive’ diathesis. In contrast, an increase of valency is found in ‘causative’ diathesis,

mostly due to the introduction of auxiliaries, such as in Fr. *laisser jouer qqn* “let sb. play” or *faire faire qqch. à qqn* “have sb. do sth.”.

Recessive diathesis cannot always be distinguished from cases of contextual ellipsis or the deletion of an optional argument. Moreover, there are some cases in which diathetic alternation is not possible, and these cannot be easily explained (consider Fr. *Jacques a une maison* → **Une maison est eue*; likewise, in Engl. *Jack has a house* is acceptable but **A house is had* is not). Nevertheless, the prototypical (and unmarked) types of diathesis are easy to identify.

→ Cennamo, *Voice*, OxfGuide 60

Ledgeway, *Passive periphrases in the Romance languages*, OxfEnc

5.2 Recessive diathesis

Recessive diathesis is expressed mainly through grammatical pseudo-reflexivity, which deletes the semantic role of the agent. Thus, in the sentence *les journaux se vendent partout* “the newspapers are being sold everywhere”, an agent is implied, but not mentioned explicitly (it may be reintroduced if it is atypical or if its identity is not self-evident, as in Sp. *la proposición se rechazó por todo el mundo* “the idea was rejected by all”). In contrast, no agent is implied in French *le repas s’achève* “the meal is coming to an end” or Spanish *la cuerda se rompe* “the rope is breaking”. Finally, in impersonal constructions such as Italian *si vendono fragole* “strawberries are being sold”, an agent is implied, but not named. In the last example, not only the semantic role of the agent, but also the syntactic role of the subject is suppressed²⁵.

The different types of diathetic passive (periphrastic or reflexive) are used in variable combinations in the Romance languages (Sp. *las luces se apagaron* vs. *las luces fueron apagadas* for “the lights were switched off”).

Although passive constructions display the syntactic characteristics of a true reflexive construction, semantically, they do not employ true pronouns. In the Spanish example *las luces se apagaron* the reflexive function word *se* marks the suppression of an agent with regard to the action of putting out the lights, but does not refer to a specific (internal or external) argument. Thus it does not correspond to the pronoun in examples such as Fr. *Colette se coiffe chez Jean* “Colette is doing her hair at John’s place”, where *se* is a true pronoun and refers to the agentive subject *Colette*.

²⁵ Impersonal constructions are also used in Mod.Fr. (*il se vend pas mal de journaux ici* “quite a few newspapers are being sold here”); however, since the expression of the subject is mandatory, the (morpho-)syntactic role of the ‘grammatical’ subject is filled in such constructions (cf. 2.3.4 no. 3.3).

5.3 Causative diathesis

Causative diathesis involves an increase in verb valency; it may make use of causative auxiliaries (Fr. *faire*, *laisser*, *rendre*, Sp. *hacer*, *poner*, It. *fare*, *rendere*) or verbs of perception (*voir*, *entendre*, *sentir*, Sp. *ver*, *oír*, It. *vedere*, *sentire*). The agent is thus replaced by new semantic roles such as that of the instigator:

Alfred est heureux “Alfred is happy” → *Marie rend Alfred heureux* “Marie makes Alfred happy”

Note the change from three arguments to four in the following example:

*Paul*_{A1} *a écrit* *une lettre d’excuses*_{A2} *à Marie*_{A3} “Paul wrote a letter of apology to Marie.”

→ *J’ai fait écrire à Paul*_{A3} *une lettre d’excuses*_{A2} *pour / à Marie*_{A4} “I made Paul write a letter of apology to Marie.”

This type of construction is widespread in the Romance languages (with the exception of modern Romanian and – partially – the dialects of the extreme south of Italy), although it was not usual in Latin. In standard Italian, it even allows some degree of passivity (*la macchina sarà fatta riparare domani* “We will have the car repaired tomorrow”), which is not possible in French or Spanish: **Juan fue hecho salir* “John was made to leave”.

The distribution of verbs involved in cases of the causative diathesis, as well as that of the pronouns and prepositions implied in their construction, varies greatly in different Romance languages (Fr. *laisser* / *rendre* / *faire*: *j’ai fait laver la vaisselle à Jean* / *par Jean* “I had John wash the dishes” vs. Sp. *hacer* / *volver* / *poner*: *le / lo hizo temblar* “it made him tremble”).

→ Vincent, *Causatives in Latin and Romance*, 2016

6 Lexical aspects of the VP: *Aktionsart*, verb valency, transitivity and semantic groups of verbs

6.1 *Aktionsart*

So-called ‘*Aktionsarten*’ (‘modes of action’, from German) describe the manner in which the action expressed by a verb is realised. Their role is to signal one of five universal ‘phases’ structuring the verbal event (pre-initial, initial, progressive, terminal and post-terminal):

- the initial phase marks the beginning of an action. This type is underdeveloped in the Romance languages: Sp. *tomo y me voy* “I make up my mind and set off on my way”, Fr. *je me mets en route* “I’m going to start” (cf. English *get moving* !, *Let’s get going*), Sp. *me pongo a escribir*, *je me mets à écrire* “I’m starting to write”

- the progressive phase describes an action in progress: Sp. *está cantando* (cf. the English progressive *I'm singing*), *sigue cantando*, *ando escribiendo*, Fr. *je suis en train d'écrire*
- the terminal phase marks the completion of an action. The grammaticalised expression of this phase is relatively common in Romance languages: Sp. *acabo de cantar* “I’ve just sung”, Fr. *je viens de chanter* (whereas it is not grammaticalised in English)
- the pre-initial and post-terminal phases correspond to the periphrastic future (*voy a cantar*, *je vais chanter*, cf. English *I will sing*) and perfect forms (*he cantado*, *j’ai chanté*, cf. English *I have sung*).

In the Romance languages, these ‘Aktionsarten’ are chiefly expressed by means of the grammaticalised periphrastic constructions illustrated above (cf. also no. 4). They may, however, also involve lexical expressions (consider *rêvasser* instead of *rêver* “to dream” or Lat. *cantare* instead of *canere* “to sing”; these two cases represent iterative forms, which can be likened to the progressive phase).

Aktionsarten in Romance only sporadically come into play in the expression of TAM (tense – aspect – mood, cf. no. 2 above); their role is nevertheless integral to the verbal event. Grammaticalisation of the five phases is particularly advanced in Ibero-Romance.

From a terminological point of view, modes of action are sometimes also referred to as ‘aspects’ in a lexicological sense; they should, however, not be confused with the phenomenon of grammatical aspect, which concerns perfective vs. imperfective (cf. *ibid.*).

6.2 The valency frames of verbs

Verb valency refers to the number of argument slots of a given verb (or, according to Tesnière’s terminology, the number of ‘actants’), and is directly linked to its lexical meaning. As the semantic and syntactic aspects of arguments will be discussed below (8.4.3), only the main configurations will be mentioned here:

- presence of either an agent or a patient in the case of monovalent verbs (i.e. verbs requiring only one argument):

*Pierre*_{AGT} *dort* “Peter sleeps/is sleeping”;

*La porte*_{PAT} *s’ouvre* “The door opens/is opening”

- presence of an agent and a patient (or in some cases the theme of an action, its content or a beneficiary) in the case of bivalent verbs (i.e. verbs requiring two arguments):

*Paul*_{AGT} *mange une pomme*_{PAT} “Paul eats/is eating an apple”;

*Marie*_{AGT} *pense à sa soeur*_{GOAL} “Mary thinks about/is thinking of her sister”

- presence of an agent, a patient and a theme/content/beneficiary in the case of trivalent verbs (i.e. verbs requiring three arguments):

*Pierre*_{AGT} *prête le livre*_{PAT} à Marie “Peter lends/is lending the book to Mary/Mary the book”

Argument marking and the presence or absence of circumstantial adjuncts differ considerably from one verb to another, and individual verbs can display multiple valency frames (consider Fr. *il joue de la flûte* “he plays the flute” vs. *il joue au foot* “he plays football”). Even lexical combinations within a given valency frame are subject to significant restrictions (which nouns appear frequently in combination with a given verb? in which order? cf. 9.5.2). These usage restrictions are shown in valency dictionaries (e.g. Busse/Dubost, *Französisches Verblexikon*, 1977; Busse, *Dicionário sintático de verbos portugueses*, 1994; Blumenthal/Rovere, *PONS-Wörterbuch der italienischen Verben: Konstruktionen etc.*, 1998) and in studies on verb and noun collocations (e.g. Stein, *Semantik und Syntax italienischer Verben*, 2005b or Gross, *La sémantique de la cause*, 2009).

Indeed, the various valency frames of a single verb are often linked to different lexical meanings (*il pense déjà aux vacances* = “he is already thinking about the holidays” vs. *il pense que tu as raison* = “he thinks/believes you’re right”); in this respect, then, valency is subordinate to lexical semantics. In other cases, however, different frames simply equate to free variants, with no semantic distinction (*fournir qqch. à qqn* vs. *fournir qqn en qqch.*, both “provide sb. with sth.”).

Optional circumstantial adjuncts display maximum variation. These are determined by the meanings of different arguments, as well as by the usage of a particular historical period (*il travaille la nuit / pendant les vacances / pour gagner sa vie / sans avoir vérifié ses outils* “he works/is working at night / during the holidays / to earn a living / without having checked his tools”).

Changes affecting valency frames and the prepositions they require are very common in diachrony. Thus, the valency frame of a verb in Old Spanish or Old French frequently differs from that of its modern equivalent even though its lexical meaning remains stable (e.g. verbs such as ASK, OBEY or PARDON across old and modern Romance). The addition of prefixes can also alter the argument structure of verbs. The stability of verb radicals and their lexical meaning does not preclude significant change, the impact of which is difficult to evaluate today, as historical research on verb collocations is still in its infancy (the only historical dictionary that systematically considers the verbal context is Cuervo’s *Diccionario de construcción y régimen*, cf. 9.9.2 no. 4).

6.3 Transitive and intransitive verbs – unaccusative and unergative verbs

The concept of verb valency is related to that of transitivity or, in other words, the ability of a verb to take object complements. A primary distinction is made between transitive verbs, which take object complements, and intransitive verbs, which do not. In many cases, the object of a transitive verb can be omitted (e.g. transitive: *Paul mange une pomme* “Paul eats/is eating an apple” vs. absolute use: *Paul mange* “Paul eats/is eating”).

More specifically, transitive verbs which establish a syntagmatic relationship between a subject and a single object (O1) are known as ‘monotransitive’ verbs. The latter is in most cases a direct object, which is unmarked (e.g. It. *vede la sua fidanzata* “he sees/is seeing his girlfriend”), although it can also be an indirect object, which is introduced by a preposition in the Romance languages (e.g. Fr. *il pense à son amie*, Rom.n *se gândește la prietena lui* “he thinks/is thinking about/of his (girl)friend”). Verbs that require two object complements (which thus fulfil the syntactic roles of O1 and O2) are called ‘ditransitive’ (e.g. *Marie prête le livre*_{O1} *à Pierre*_{O2} “Mary lends/is lending the book to Peter”, Sp. *Le ha dado a mi hermana*_{O2} *una manzana*_{O1} “He/she has given/gave my sister an apple”).

Intransitive verbs, on the other hand, never take direct objects and only take non-direct object complements (C). However the latter are frequent and varied, as they can be accompanied by indirect objects (e.g. Sp. *le*_C *murió la abuela* “his grand-mother died”), locatives (e.g. *elle dort à la maison*_C “she sleeps/is sleeping at home”; cf. verbs as ARRIVE, LEAVE, GO, COME), temporal arguments (verbs as DIE, BE BORN, EMERGE) or other optional circumstantial adjuncts (e.g. It. *quest'uomo ride senza motivo*_C “this man laughs/is laughing for no reason”).

In addition to transitive and intransitive verbs, one finds attributive (or ‘copular’) verbs. Like intransitive verbs, they lack an object; instead, they establish an immediate relationship between a subject and an attribute (e.g. the French verb *être* in the construction *être pasteur* “to be a priest”).

Finally, in periphrastic verb forms, the (in)transitivity of a verb involves a specific perfective auxiliary: both transitive and the majority of intransitive verbs employ an auxiliary derived from Lat. HABERE (e.g. Fr. *avoir*). In Modern Italian and French, many verbs still employ a descendant of Lat. *ESSERE (It. *essere* and Fr. *être*, the paradigm of the latter involving suppletion by means of the inclusion of descendants of Lat. STARE “to stand”). These include pronominal verbs such as Fr. *se laver* “to wash”, as well as a number of verbs used in perfective clauses, such as Fr. *il est allé*, Occ. *es anat*, It. *è andato* (vs. Sp. *ha ido*, Cat. *ha marxat*, Rom.n *a plecat* “he has gone/left”).

A second important distinction involves only intransitive verbs, among which one can distinguish between ‘unaccusative’ and ‘unergative’ verbs (cf. La Fauci, *op. cit.*, pp. 88–93). This distinction, introduced by David Perlmutter (*Impersonal passives and the unaccusative hypothesis*, 1978) specifically involves the properties of the subjects of these verbs. In French, Occitan or Italian, unaccusative verbs conjugated with perfective auxiliaries take *ESSERE and require agreement with their grammatical subject on the past participle (*Marie est tombée* “Mary fell/has fallen”); the grammatical subject thus assumes some properties of a 1st object in transitive constructions since it does not correspond to the semantic role of an agent (e.g. *Marie*_{PATIENT} *tombe* “Mary falls/is falling”); like direct objects, it triggers participle agreement and it also occurs in the typical (= postverbal) object position (e.g. (unmarked/thetic) *è arrivato Gianni*; cf. Fr. *Il est arrivé un enfant*). According to La Fauci, the subject is ‘in a commutative relationship

with the function of the direct object' (*Compendio*, 2009: 90). The subjects of intransitive unaccusative propositions and the subjects that appear in passive constructions display similar semantic and syntactic properties (e.g. It. *La porta*_{PATIENT} *è stata aperta* "The door has been opened"; cf. La Fauci 2009: 90).

Unaccusative verbs stand in opposition to so-called 'unergative' verbs. Complex forms of the latter take an auxiliary based on HABERE even in French and Italian, and have 'canonical' subjects that do not display the properties of a 1st object (e.g. Fr. *Les enfants ont rigolé* "The children had fun"). While the subjects of unaccusative verbs relate to the semantic patient, the subjects of unergative verbs typically fulfil the semantic function of agents (unerg. *Les enfants*_{AGENT} *rigolent* vs. unacc. *Marie*_{PATIENT} *tombe*).

It should be noted that these two classes of verbs are also to be found in languages which do not display auxiliary alternation, such as modern Ibero-Romance or English (cf. Sp. unacc. *Muerto el abuelo, Juan se marchó* vs. unerg. **Llorado el abuelo, Juan se marchó*).

6.4 Semantic groups of verbs

Unsurprisingly, the lexical meaning of verbs primarily concerns lexis. A classification of verbs founded on broader concepts nevertheless involves the domain of grammatical semantics. A fundamental distinction can be made between dynamic verbs (which describe actions or events, e.g. *faire* "do", *écrire* "write") and stative verbs (which express states of being, e.g. *être* "be", *connaître* "know", *aimer* "like, love", cf. Zeno Vendler, *Linguistics in Philosophy*, 1967). At the level of the sentence, modal or temporal meaning can be expressed syntactically by means of modal verbs (*vouloir* "want", *devoir* "must", *pouvoir* "can"), epistemic verbs (*croire* "believe", *savoir* "know") or verbs of movement (*aller* "go").

Other fundamental categories include:

- existential verbs (Sp. *faltan posibilidades para los niños* "there are not enough opportunities for children")
- verbs that describe a natural event (Sp. *salió el sol* "the sun rose")
- verbs expressing sensory phenomena (Sp. *me duele la cabeza* "I have a headache (= I feel pain in my head)", *me gusta el vino* "I like the wine (= the wine tastes good to me)")
- verbs expressing the availability of something (Fr. *prêter* "lend", *vendre* "sell")

Verbs expressing illocutionary acts (cf. 4.5.3 no. 1) simultaneously describe the speech acts in which they are involved (*constater* "notice", *prétendre* "claim" / *demander* "ask", *ordonner* "order", *souhaiter* "desire" / *saluer* "greet", *mettre en garde* "warn", *promettre* "promise", *reprocher* "reproach" / *demander pardon* "beg forgiveness" / *nommer* "name", *déshériter* "disinherit").

The grouping of verbs according to concept clusters implies a psycho-linguistic and cognitive perspective, which is in line with more general research on lexical concepts, and allows a more accurate understanding of syntactic meaning.

- Heidinger, *Valency in the Romance languages*, OxfEnc
- Loporcaro, *Auxiliary selection and participial agreement*, OxfGuide 49
- Bentley, *Split intransitivity*, OxfGuide 50
- Bentley/Ciconte, *Copular and existential constructions*, OxfGuide 52
- Pountain, *Copulas in the Romance languages*, OxfEnc
- Bentley, *Existential and locative constructions in the Romance languages*, OxfEnc
- Acedo-Matellán/Mateu/Pineda, *Argument structure and argument realization*, CambrHandb 16
- Cabredo Hofherr, *Verbal plurality in the Romance languages*, OxfEnc

7 Optional determiners: negation and adverbs

Simple sentential negation is normally placed before the verb in the Romance languages (like the forms *non* and *ne* in Latin): Sp. *no canta(r)* “he/she does not sing (do not sing)”, Pg. *não canta(r)*, It. *non canta(re)*, Rom.n *nu cânta (nu cântă)*. French is an exception here, as it exhibits a discontinuous negation of the type *ne ... pas* (the cultivated French type *je ne sais* is subject to lexical restrictions; the loss of the preverbal adverb *ne* in familiar registers of French, however, is generalised).

Formally, the French construction is similar to another type, the double negation, in which a second reinforcing negative particle (e.g. Fr. *jamais*, *plus*, *personne*, *nulle part* “never, no more, nobody, nowhere”) is placed after the verb (Fr. *il ne chante jamais* or Sp. *no canta nunca* “he/she never sings”, It. *non lo credo mica* “I don’t believe it one little bit”). In Spanish or Italian, this reinforcement can also be expressed by a preverbal negator (Sp. *nunca canta*, It. *mica lo credo*).

In both types of negation, the auxiliary verbs in modern Romance languages are placed before the main verb (*cf.* the description of periphrases in no. 3 above), whereas adverbs normally follow the verb. The adverb may be inserted between the elements of a periphrastic construction, (e.g. in French: *elle ne l’a certainement pas encore rencontré* “she definitely hasn’t met him yet”; this configuration is impossible in Ibero-Romance and exceptional in Romanian).

The position of adverbs can also imply a certain difference between an adverb modifying the VP (*il parle très bien français* “he speaks very good French”; *il reprit ses esprits curieusement* “he regained his spirits in an odd manner”) and adverbial elements modifying the entire sentence, subject included (*il parle le français très bien* “he speaks French very well”; *curieusement, il reprit ses esprits* “strangely, he regained his spirits”).

Note that adverbs can function outside the VP as elements modifying adjective phrases (*elle est* [[très]_{ADV} [jolie]_{ADJ}]_{ADVP} / *she is* [[very]_{ADV} [pretty]_{ADJ}]_{ADVP}) or even stand alone in the place of an entire sentence (*Oui!*, *Assurément!* / “Yes! Certainly!”). Furthermore, adverb phrases can be expanded by means of further adverbs (*Il travaille vrai-*

ment bien “He really works well”), often marking degrees of comparison (*Il saute plus/moins/aussi haut* “He jumps higher/less high/as high”); in some cases, this results in more complex constructions that make use of prepositional complements (*J’ai agi* [[*conformément*]_{PREP} [à la loi]_{PP}]_{ADVP} “I have acted in accordance with the law”).

→ Poletto, *Negation*, OxfGuide 51

Moscatti, *Negation and polarity in the Romance languages*, OxfEnc

8.3.3 The transformation of the verbal system from Latin to Romance

A comparison between the Latin and Romance verbal systems is provided below. The following cases are distinguished: continuity of both form and function (no. 1), continuity of function with replacement of formal elements (no. 2), categories which constitute innovations in the Romance languages (no. 3) and finally, the abandonment of both Latin form and function (no. 4).

1 Continuity of form and function

As far as the main categories are concerned, the Romance languages preserve the present indicative, subjunctive and imperative as well as the past imperfect indicative and subjunctive, which do not display significant variation. The past perfect indicative is reduced, but is preserved in specific varieties, as shown in the summary below:

- present indicative: (cf. the paradigms provided in 7.6.1)
- present subjunctive: CANTEM > It. *canti*, Sp. *cante*, (O.)Fr. *(je) chant(e)* etc.
- present imperative: CANTA > It. Sp. *canta*, Fr. *chante* etc.
- past imperfect indicative: (cf. the paradigms provided in 7.6.3); the functions are maintained for the most part
- past perfect indicative: CANTA(V)I > It. *cantai* (generalised in the dialects of the extreme south of Italy for the past perfect, but defunct in the dialects of Northern Italy today), Sp. *canté* (especially in the Americas), Fr. *je chantai* (high register).

→ Schaden, *Perfects in the Romance languages*, OxfEnc

2 Replacement of formal elements with continuity of function

For the Latin categories whose forms are replaced by new constructions, which are usually periphrastic, the preservation of functions is also frequent. The replacement of the simple future (future I) forms by periphrastic future forms has already been described (cf. 7.6.4). The emergence of new ‘synthetic’ future forms in Late Latin was followed by the development of further periphrastic forms, as in French *je vais chanter* and (American) Spanish *voy a cantar* “I’m going to sing” (cf. 8.3.2 no. 5.1).

The periodic appearance of new future forms has been a topic of interest for many linguists, who have observed regularities in what may be termed a ‘future cycle’ (cf. Fleischman, *The future in thought and language*, 1982 and the fundamental work by Koch and Oesterreicher (1996), the latter including essential references to previous researchers, Helmut Lüdtke in particular). The stages of a linguistic cycle are as follows: (i) a given form undergoes phonetic reduction; (ii) it is replaced by a new form, into which new lexical material is incorporated (particularly elements of a modal nature); (iii) the new form then undergoes phonetic reduction.

These observations allow the identification of evolutionary patterns such as the following: (i) phonetic reduction: *cantans* + Indo-Eur. **b^{hw}o* (subjunctive of “to be”) > Lat. *cantabo*; → (ii) replacement by a new form: Pop. Lat. *cantare habeo* “I have to sing” > (iii) reduction Fr. *chanterai* “I will sing” → (ii) new form: Fr. Mod. *je vais chanter* “I’m going to sing”.

The synthetic forms of other Latin tenses were replaced by Romance periphrases combining an inflected form of HABERE with the past participle of the main verb:

- pluperfect indicative: forms were generally replaced by past imperfect forms of HABERE / *ESSERE + past part: CANTAVERAM → It. *avevo cantato*, Sp. *había cantado*, Fr. *j’avais chanté*; Rom.n shows a very marked change of its own, using forms of the original Latin pluperfect subjunctive: *cântasem* (cf. below).

However, the synthetic Latin forms have survived with a new function in Modern Spanish, where they have assumed the role of the conditional (*quisiera* ...) and the past subjunctive (e.g. *quería que viniera*); in many dialects of the upper south of Italy, the Latin pluperfect also functions as a conditional (e.g. Cal. (Cosenza) *cci jerra subbitu* “I’d go there at once”. In O.Fr. the pluperfect forms *auret* 3SG (< *aveir* INF) and *roveret* 3SG (< *rover* INF “ask, request”) appear in the *Canticle* (or *Sequence*) of *Saint Eulalia* (one of the earliest French texts, dating from the end of the 9th century) with past perfect function (cf. 10.3.5 no. 1 and 11.2.4).

In Ibero-Romance, the occasional maintenance of both the form and the function of the Lat. pluperfect can even be observed, such as in Modern Spanish journalistic writing and in certain Latin-American varieties; cf. also Pt. *falara* “I had spoken”, belonging to a higher register, alongside the more popular *tinha falado*;

- past perfect subjunctive: the forms were partly replaced by those of the present subjunctive of HABERE/*ESSERE + past part.: CANTAVERIM → It. *abbia cantato*, Sp. *haya cantado*, Fr. *j’aie chanté*, and partly by those of the pluperfect subjunctive (cf. the following remarks);
- past imperfect subjunctive: the forms were generally replaced by those of the pluperfect subjunctive and the function has been extended to a ‘past subjunctive’ (which involves both imperfective and perfective aspects and which thus also assumes the functions of the past perfect subjunctive): CANTAREM (+ CANTAVERIM)

- CANTAVISSEM > CANTASSEM > It. *cantassi*, Sp. *cantase*, Fr. *je chantasse* (rare in Mod.Fr.); hence, It. *Non sapevo che Maria venisse a casa nostra* conveys both the perfective meaning “I didn’t know that Maria came to our house” and the imperfective meaning “... was coming to our house”. In Sardinian, however, the forms and functions of the past imperfect subjunctive have been preserved (parallel to the periphrastic past perfect and pluperfect subjunctive): CANTAREM > Sard. *cantare*;
- pluperfect subjunctive: the Latin forms (which survive in the function of the past subjunctive, cf. above) were replaced by different types of periphrases: It. *avessi cantato*, Sp. *hubiese ~ hubiera cantado*, Fr. *j’eusse chanté*;
 - future perfect: forms were replaced by those of the future of HABERE / *ESSERE + past part.: CANTAVERO → It. *avrò cantato*, Sp. *habré cantado*, Fr. *j’aurai chanté*. The forms of the future perfect survive in part in the Ibero-Romance languages (but with the function of future subjunctive): Sp. *cantare*, Pg. *cantar*.

3 Romance innovations

Cases of true innovation are comparatively rare, and concern only the conditional, the periphrastic perfect and past anterior, as well as the double compound past.

In almost all Romance languages, the formation of the conditional is parallel to that of the periphrastic future. The infinitive of the main verb is followed by the past imperfect of HABERE; phonetic contraction then leads to the fusing of the two elements to produce a synthetic form:

CANTARE (HAB)EBAM > Sp. *cantaría*, Pg. *cantaria*, Fr. (*je*) *chanterais*; as in the case of the future, the grammaticalised morpheme can be separated from the verb base in medieval Ibero-Romance varieties in particular (cf. 7.6.4).

Modern Italian conditional forms are based on the perfect of HABERE:

CANTARE HABUI > *canterei*; O.It. also shows the common Romance type *cantaria*, *cantaria*.

In modern Romance languages (but not in Italian), the conditional assumes temporal functions (‘future in the past’: *je crois qu’il viendra* “I think he’ll come”. → *je croyais qu’il viendrait* “I thought he would come”) as well as modal functions (‘hypothetical’ conditional sentences [= ‘potentialis’] or ‘counterfactual’ [= ‘irrealis’]: *s’il faisait beau, nous passerions la journée dans la montagne* (“if it were fine weather, we would spend the day in the mountains”).

The quality of the future in the past can also be expressed by other verbal paradigms (e.g. the Fr. past imperfect in *j’aimais à l’époque une fille qui devait mourir peu après* “at the time I loved a girl who would die soon afterwards” could be expressed in

It. either by the past imperfect ... *che doveva morire poco dopo*, or by a conditional form: [...] *che sarebbe morta poco dopo*).

The conditional perfect is a compound tense (*avrei cantato, habría cantado, j'aurais chanté*) presenting no further peculiarities in most Romance languages; in Italian, however, it can assume the function of the future in the past, instead of the conditional (e.g. Fr. *Il a dit qu'il payerait* vs. It. *Ha detto che avrebbe pagato* [**pagherebbe*] “He said he would pay”).

Depending on the language and the verb in question, the periphrastic perfect (= compound past) is formed by the present form of HABERE, ESSE(RE) or TENERE followed by the past participle of the main verb:

Sp. *he cantado*, Rom.n *am cântat*, It. *ho cantato* ~ *sono stato* (< ESSE), Fr. *j'ai chanté* ~ *je suis allé*, Pg. *tenho cantado* (< TENERE)

We have seen that this category reinforces the parameter of aspect in the verbal system (the past is seen from the angle of the present).

The past anterior emerged from the simple past of HABERE + past part.:

ebbi cantato, hube cantado, j'eus chanté

Double compound past paradigms are very common in the Romance languages (e.g. Fr. *quand elle a eu terminé ses études* – instead of *quand elle avait ...* –, literally “when she has had finished her studies”); they may occasionally replace periphrastic forms that have lost their transparency, but they have also developed their own specific aspectual and modal values (cf. Schaden, *La sémantique du parfait*, 2007; Ledgeway, ‘Aspectual and irrealis marking: the distribution of the Wallon *temps surcomposés*’, 2023).

The inventory of periphrastic forms in Romance verb inflection is extremely rich; cf. the French examples *j'avais chanté* (pluperf.), *j'aurai chanté* (fut. perf.), *j'aie chanté* (perf. subj.), *j'eusse chanté* (pluperf. subj.), *j'aurais chanté* (past cond.), *j'eus chanté* (past ant.), *en ayant chanté* (past gerund), *être chanté* (present passive infinitive).

The remaining innovations in the verbal system are more sporadic and mostly concern precise functions relating to the different categories. The following functional changes also involve formal modifications:

- the inflected infinitive in Pg. and Galician (as well as in Sardinian, O. Leonese and O. Neapolitan); cf. 8.3.3 no. 5 for example;
- the negation of the imperative by means of a particle of negation followed by an infinitive in the singular or by a subjunctive (It. *non cantare*, Sp. *no cantar, no cantes*); only Fr. displays a homogeneous paradigm for both positive and negative forms of the imperative (*crie* – *ne crie pas* “shout – don’t shout”);
- ‘subject conjugation’ (*je joue* “I play”) and ‘object conjugation’ (*je le vois* “I see him/it”), which will be examined in greater detail below (cf. 8.4.3 nos. 3 and 4).

4 Abandonment of Latin form and function

The innovations described above were countered by the disappearance of some Latin structural categories or elements during the transition to the Romance languages:

- the ‘synthetic’ passive voice (cf. 8.3.2 no. 5.1 above)
- the future imperative, participle and infinitive (CANTATO, AMATURUS, AMATURUS ESSE)
- deponent verbs [= verbs which are active in meaning but which take the forms of the passive voice] (MORI → MORIRE > *mourir*, *morir* “die”)
- the supine [= a nominal form of the verb] (DICTU), with the exception of Romanian, where the supine continues to this day (e.g. *frumos de auzit*, literally “lovely to hear”, *am terminat de cântat* “I have finished singing”)
- the formal distinction between a present and a perfect verb stem

5 Non-conjugated verb forms

Non-conjugated verb forms underwent changes similar to those described for finite forms during the transition from Latin to Romance:

- form and function were maintained in the case of the active present infinitive: CANTARE (> *chanter*); since the Latin passive present infinitive has been lost in Romance, this form can have active or passive meaning (e.g. active: *voglio amare quella ragazza* vs. passive: *una ragazza da amare*);
- the Latin passive present infinitive can be replaced by periphrastic forms: CANTARI → *être chanté*
- this also applies to the active and passive perfect infinitive: CANTAVISSE → *avoir chanté*, CANTATUS ESSE → *avoir été chanté*;
- the forms of the Latin gerund were reduced to a single form (probably the original ablative form of the Latin gerund): CANTANDO > It. Sp. *cantando*, Fr. (en) *chantant* (homophonous with the present participle); note that the use of the present participle is restricted, as is that of participial and gerundive constructions.

It should also be noted that in French, verbal present participle forms (e.g. *ses mains tremblant de peur* “her hands trembling with fear”) have been distinct from verbal adjective forms (*ses mains tremblantes* “her trembling hands”) since the 18th century.

Despite the profound restructuring of the Latin system of verbal categories, numerous formal and functional elements were maintained, to a greater or lesser extent. This continuity is all the more remarkable when considered in the light of the major processes of restructuration which took place during late Antiquity and the early medieval period. Thus, from their beginnings through to the modern varieties, the Romance languages lend themselves particularly well to comparative studies. Consequently, it is

useful to analyse phenomena pertaining to the early stages of the Romance languages from a ‘differential’ perspective (i.e. by identifying the ways in which earlier stages differ from the language of today).

8.4 The sentence in the Romance languages

8.4.1 Denotative content and principles of linear marking

The study of the sentence involves questions relating to the statement, i.e. the decontextualised semantic content of a sentence, to information structure (which anchors the sentence to its immediate syntactic and communicative environment), and to the wider textual environment, which is principally defined by pragmatic factors.

These textual and pragmatic aspects include strategies of cohesion and coherence (cf. 4.5.3, no. 2 in particular), as well as temporal regulation and structuring markers and their use in different Romance textual genres over time (cf. 4.5.4). Textual and discourse traditions also determine the degree of complexity of sentence structure: while simple sentences are preferred in spoken discourse, written language favours the development of subordination. From this point of view, linguistic elaboration in Romance may be regarded as a history of the complex sentence from the Middle Ages onwards.

The following section, however, will focus on the denotative content. The notion of ‘assertion’ or ‘statement’ shows that syntax distinguishes two levels of structure: one level relating to the concrete realisation of syntactic structures in a given language, and a second, more abstract level that draws on general, anthropological concepts. The second level implies a hierarchical organisation of syntactic information. Linguistic expression (i.e. the first level), in contrast, is rigorously linear; in addition, it is characterised by an extreme diversity of patterns of realisation, both among different languages and within each individual language.

The same assertive content can thus give rise to different linear constructions: the denotative content in the sentence *toutes ses robes*_{A1} *avaient en commun*_p *un je ne sais quoi*_{A2} *qui lui*_{A3} *échappait* ‘all of her dresses had in common a *je ne sais quoi* that eluded him’ is identical to that in *ce je ne sais quoi*_{A2} *que toutes ses robes*_{A1} *avaient en commun*_p *lui*_{A3} *échappait* ‘that *je ne sais quoi* that all of her dresses had in common eluded him’. Rather, it is the information structure of the sentence that is affected by such differences (i.e. the theme-rheme structure, cf. no. 4 below). In the same way, while the TAM (tense – aspect – mood) system is universal, its features are not realised in the same way in all languages. To give one example, in Romance languages, the verbal system is based on tense, whereas in Semitic languages, aspect is central²⁶.

²⁶ It is conceivable that, in the act of utterance, thoughts are first reorganised into hierarchical linguistic units, which are then restructured as smaller entities and given a linear form by means of the neces-

Examples of functional equivalence are numerous in syntax: in synchrony, a nominal argument can be replaced by an entire (subordinate) clause; in diachrony, inflectional markers can be replaced over time by agglutinating markers or vice versa; in synchrony and in diachrony, the functions carried out by word position can be replaced by grammatical markers.

The principles of syntactic marking allow the expression of complex and hierarchical relationships in a purely linear manner. The fundamental means by which this may be achieved are universal. They include:

1. the order, and thus the relative position of elements within syntactic constructions; here, linearity is directly reflected by the syntax (e.g. *Paul aime Pauline* “Paul loves Pauline” vs. *Pauline aime Paul* “Pauline loves Paul”). This is labelled ‘taxematic determination’ (from Gr. *taxis* = Lat. *ordo* “order”) or ‘configurationality’;
2. grammatical markers that are added as new elements to the basic elements. This is called ‘grammatical determination’; it makes use of inflection (cf. 7.2.3) as well as function words (e.g. Sp. *Pablo quiere a Juanita* vs. *a Juanita le quiere Pablo*); it may also use principles of ‘zero’ marking as in the subject case (*Pablo*), which is often unmarked (cf. 7.1.2 no. 3);
3. prosodic markers that intervene simultaneously and that mostly affect the global interpretation of the sentence (cf. 6.3) e.g. *il parle simplement* vs. *il parle, simplement* (where the pause acts as a prosodic marker); similarly, *je pensais qu’elle y serait* “I thought she’d be here” conveys either the presence or the absence of *elle* “she” (the subject of the subordinate clause), depending on whether the stress, which acts as a prosodic marker in this case, is placed on *pen’sais* or *se’rait*).

The position of grammatical markers in relation to lexemes is also important (cf. Bosson, LangTyp, art. 48); these may be placed before a lexeme (= prefixed, GRAMM – LEX: *a la vecina*), after it (= suffixed, LEX – GRAMM: *vicin-ae, venie-ba-tis*) or on either side of it (= circumfixed, e.g. *Jean ne vient pas*); conversely, infixes are not usual in the Romance languages (in contrast to Semitic languages, for example). The position of the markers may vary according to specific verbal forms (e.g. It. *me lo dai* vs. *dammelo*), the position occupied by a lexeme within the sentence (Pg. *quero-te* vs. *não te quero*) or prosodic conditions.

Though the multiplicity of means of concrete expression in different languages may seem disconcerting, these actually correspond to a very limited number of principles, mainly relating to position and grammatical marking.

sary marking. The brain performs a sort of grammatical ‘tagging’ in order to lend meaningful elements a form. In the decoding of information, this process is reversed: linear chains with their grammatical encoding are interpreted and transformed into a hierarchical structure, to which a meaning is immediately attributed (Raible, LangTyp, art. 1, p. 13, compares this process to the computerised identification of parts of speech [morphological tagging] and of constituents [syntactic parsing]).

Configuration (or distribution) concerns the elements affected by grammatical marking. There are two possibilities: either two or more of the elements concerned may exhibit grammatical marking ('congruence'), or only one element may be marked. In Latin, the marking of the subject and the verb within the propositional framework is 'bilateral': *Socrate-s curr-it*. In the Romance languages, in contrast, 'unilateral' marking is predominant (and involves the predicate: Fr. *Socrate cour-t*, Sp. *Sócrates corr-e*, It. *Socrate corr-e*); bilateral marking may occur at the level of the proposition in the presence of periphrastic verbs (It. *è venut-a Giulia* vs. Sp. *ha venid-o Julia*)²⁷.

In conclusion, the transition from a hierarchical to a linear structure implies a number of constraints, since linguistic segments must necessarily follow one another in time (or space). This goes some way towards explaining the variability of syntactic realisations, as every language must find ways of encoding grammatical meaning at every stage of its development. However, since no linear form can adequately reflect a hierarchical configuration, structures remain unstable and are thus prone to variation. If phonetic, morphological and lexical changes, which have their own causes, are also taken into account, the diversity in the physiognomy of individual languages becomes easily comprehensible.

In the following sections, four aspects central to the (diachronic) study of the sentence in the Romance languages will be examined:

- the changes in constituent order which took place between Latin and the modern Romance languages (no. 2)
- semantic roles (no. 3)
- information structure, which transcends the limits of the sentence itself (no. 4)
- subordination, which constitutes the basic principle of complex sentences (no. 5)

Related topics which will not be considered here include:

- the different types of simple sentences (declarative, interrogative, exclamative, injunctive)
- certain types of complex sentences (the interaction between juxtaposition, coordination and subordination)
- atypical sentences (incomplete or elliptic).

- Cruschina/Ledgeway, *The structure of the clause*, OxfGuide 31
- Ledgeway/Schifano, *Parametric variation*, CambrHandb 21

²⁷ Bilateral marking between the subject and the verb is also maintained in the Romance varieties that exhibit a binary case system (especially Old Gallo-Romance and Romanian, cf. 7.3.4).

8.4.2 Constituent order and principles of position

Constituents are basic entities within the structure of sentences. In the modern Romance languages, they follow a more or less rigid subject-verb-object (SVO) order, whereas in Latin, they display a loose subject-object-verb (SOV) structure. The lexemes and grammemes involved in the formation of constituents also adhere to a specific internal order. In constituent order, as well as in the positioning of grammemes in relation to lexemes, so-called construction ‘heads’ and dependent elements are distinguished (head vs. dependent, following Greenberg 1963). The ‘head’ of the constituent is the component that determines the ‘plan’ for the whole construction, such as the type of complements it requires, if any, as well as the semantic properties of the complements.

According to dependency grammar, the verbal nucleus constitutes the head of the entire clause (e.g. *Jean donne le livre à son frère*), whereas most generative and functionalist theories consider the nominal group that functions as the subject of the clause to be situated on a higher level than the predicate, but nevertheless subordinate to the head of the clause, which is inflection (tense and person; 3SG.PRES.IND. in the example provided).

In the case of an NP, the head of the construction is a noun, according to most grammatical theories (e.g. *cette grande voiture blanche*); generativist approaches, however, consider that above the NP, there is an additional functional layer involving determiners, called the ‘determiner phrase’ (DP).

The syntactic changes which took place between Latin and the Romance languages represent a general process of reorganisation which may be described as an inversion of the direction of construction: whereas Latin tends to construct from right to left, the Romance languages construct from left to right; in Latin, the sentence head is most commonly situated further to the right, since the verb (i.e. the element on which, from the theoretical perspective of dependency grammar, the entire construction depends) is most often in sentence-final position:

Lat. *hominis domum video* [GEN-NOM-V] “I see the man’s house”

→ Fr. *Je vois la maison de l’homme*, Sp. *veo la casa del hombre*, Pg. *vejo a casa do homem*, It. *vedo la casa dell’uomo*, Rom.n *văd casa bărbatului* [V-NOM-COMPL]

In Classical Latin, SOV word order was dominant; in (spoken) Late Latin, it was replaced by VSO, which also characterises the syntax of the early Romance languages. This reorientation from a ‘centripetal’ to a ‘centrifugal’ type of syntax (in which the construction head is displaced from right to left) appears to have been linked to orality and thus to specific forms of oral expressivity in Late Latin (which after the fall of the Western Roman Empire in 476 was beginning to develop regional differentiation, thereby losing its former function as a vehicle for large-scale communication); however, this order also exhibits the tendency to place the verb before the object, which is characteristic of

creole languages. At a later stage, the subject – which typically assumes the semantic role of the agent – moved into first position. This change may be linked to the fact that agents occur more frequently in discourse than other elements, and perhaps even more so to the fact that subjects, as they are most frequently agents, are prototypically topics.

The VO order (which allows the subject to appear either in first or second position, i.e. VSO or SVO) was grammaticalised in all Romance languages from around 1200 onwards. Although an inverted OV structure remained possible, it was only used to express emphasis in order to focalise or topicalise the object (*cf.* Ledgeway, *V2 beyond borders: The Histoire ancienne jusqu'à César*, 2021).

The SOV word order of Classical Latin was still apparent in texts such as the O.Fr. *Strasbourg Oaths* (842, conserved in a manuscript dating from the 10th century), which was written in a linguistic form close to 'rustic' Latin (*cf.* 10.3.5 no. 3); it also survives to some extent in the O.Sp. *Poema de mio Cid* (ca. 1200, ms. 1235), but is more frequent in subordinate than in main clauses. However, in most Old Romance texts, the SOV word order had already disappeared:

- In the *Song of Roland* (ca. 1100, ms. 1225/50), 79% of sentences display (S/X)VO word order (e.g. *li reis Marsilie_s m'ad transmis_v ses messages_{oi}* "King Marsilie sent me his messengers" l. 181); the remaining 21% of sentences with an (X)OV structure may thus be conditioned by the syntactic context (e.g. emphasis of the O2: *a voz Fran-ceis_{oz} un counseill_{oi} en presistes_v* "you took advice from your French people" l. 205);
- In the *Queste del Saint Graal* (ca. 1235), OV(S) constructions are reduced to a mere 1% (*ceste parole_{oi} dist_v Salomons_s* "Salomon said the following words")²⁸.

The SV(O) order – with an agent subject (as opposed to a patient subject) – began to be generalised from the end of the 16th century onwards, but to varying extents from language to language: it displays few constraints in Spanish, is more common in Italian and Portuguese and is entirely grammaticalised in French:

- In the O.Sp. *Primera Crónica General* (compiled in the latter part of the 13th century), VS order still represents 60% of sentences, while 35% display SV order (*salieron_v a ellos_{oz} los moros_s et mataron_v y_c a todos_{oi}* "the Moors went out towards them and killed them all there").
- In the chronicles of the 15th century, this relationship has changed to 30–45% of VS structures compared to 55–77 % of SV structures. In the *Life of Lazarillo de Tormes* (16th century), the two orders are present in equal measure.
- In modern literary language, SV is generally dominant, accounting for up to 70% of all sentences (vs. 30% VS) according to certain studies and with variation among

²⁸ This section draws from the examples and percentages provided in an unpublished text by Georg Bossong (personal communication, 2000); *cf.* also *id.*, *Ausdrucksmöglichkeiten für grammatische Relationen*, LangTyp, art. 48, p. 662 *sqq.* and *La sintáxis de las Glosas Emilianenses*, 2006).

the individual Romance languages; VS order thus presupposes a patient subject or carries specific semantic connotations, e.g. as an indicator of narration (*cf.* Neruda: *trasladóse_v la familia_s a aquella vivienda_c* “the whole family moved to that residence”).

Processes of change are slow, and do not occur at the same pace in the different languages: in Spanish or in Romanian, the medieval VSO type still appears today, while in Portuguese it had already become rare in the Middle Ages; in Italian, VSO disappeared in the 19th century (though it was still present in the writings of Manzoni: *vide_v Lodovico_s spuntar da lontano un signor tale_{oi}* “Lodovico observed a certain gentleman appearing from a distance”).

The issue is further complicated by the frequency of constructions displaying dislocation, undoubtedly catalysed by the type of information structure particular to spoken language. In popular French, the expression of the ‘semantic’ subject can be delayed until after the VP, providing or ‘recalling’ the reference of a preverbal pronominal subject (*ils ~ on*):

Ils_(s) me prenaient_v pour un satyre tous ces cons_s (Queneau) “Those fools mistook me for a satyr”

On_(s) avait remarqué_v ça nous autres_s (Céline) “The rest of us had noticed that”

Ils_(s) sont_v fous ces Romains_s (Goscinnny) “These Romans are crazy”

The Romance languages provide an excellent illustration of the fundamental instability of basic constituent orders in the languages of the world. The different constellations – the most frequent being SOV, SVO and VSO – each carry specific advantages and disadvantages. The VSO order facilitates decoding, since it places the verb at the beginning of the sentence; however, the object does not directly follow the verb, potentially leading to problems of comprehension. The VOS model characteristic of Austronesian languages overcomes this difficulty, though it places the thematic subject, which should logically precede the rheme (i.e. what is being said about the theme), in final position. The transition from mental hierarchy to the temporal linearity of linguistic expression presents a particular challenge which constituent order must strive to overcome; none of the solutions is perfect and all remain unstable.

→ Salvi, *Word order*, OxfGuide 62

8.4.3 Arguments and semantic roles

1 Overview

As the syntactic features of the verb phrase have been dealt with above, this section will focus on the relationships between predicates and their arguments. Arguments display two main characteristics:

1. Arguments are finite in number. The links between a predicate and its arguments concern a finite number of elements. Normally, a predicate governs between zero and three arguments, depending on its individual (and thus lexical) valency. The most common configuration seems to be a bivalent structure (i.e. two arguments for one predicate), illustrated by the following examples from Sp. and Fr.:

- zero arguments (= absence of a semantic subject):
llueve / il pleut “it’s raining” (cf. 3.3 below)
- one argument (subject):
Pedro duerme / Pierre dort “Peter is sleeping”
- two arguments (subject + 1st object):
Pablo lee la carta / Paul lit la lettre “Paul is reading the letter”
- three arguments (subject + 1st object + 2nd object):
Carlota (le) da el libro a Pedro / Charlotte donne le livre à Pierre “Charlotte gives Peter the book”

A fourth argument can be introduced by means of causative diathesis (cf. 8.3.2 no. 5.3); however, in sentences with four arguments difficulties in parsing often appear. Consider the following:

Carlota (le) hace dar el libro a Pedro por Amador / Charlotte fait donner le livre à Pierre par sa soeur “Charlotte is having her sister give Peter the book”

In contrast, complement clauses fulfilling the syntactic role of a second 1st object are very common: *Je te parie 100 € que* + [object complement clause] “I bet you a hundred euro that ...”.

Moreover, the group formed by the fundamental relationship between a predicate and its arguments may be expanded by optional supplementary information in the form of circumstantial adjuncts. On the surface, these adjuncts may resemble arguments (compare *Paul mange une tarte au citron* “Paul is eating a lemon tart” (= argument) vs. *Paul mange toute la journée* (= circumstantial adjunct) “Paul eats all day long”.

2. Arguments form a hierarchy. The arguments governed by a predicate form a hierarchy corresponding to the syntactic roles of subject, 1st object and 2nd object. The subject is nearly always present in sentences and is only rarely grammatically marked; the 1st object is also often present and sometimes displays grammatical marking; the 2nd object appears less frequently but displays a greater degree of grammatical marking.

Note that in Tesnière's terminology, the subject is referred to as 'prime actant', the 1st object as 'second actant' and the 2nd object as 'third actant'. This terminology also envisages a 4th and 5th actant (i.e. argument) for object complements that are often introduced by prepositions (more commonly termed 'indirect complements' or 'circumstantial complements').

2 Semantic roles

The concept of semantic roles has been incorporated into various models of syntactic theory and they thus have a number of more or less synonymous designations: 'actantial roles' (Tesnière 1959) or 'actant functions' (*Aktantenfunktionen*, Heger 1976), 'participant roles' (Halliday 1967–68), 'case roles' or 'deep cases' (Fillmore, 1968) and 'theta' roles (Chomsky 1981). Theoretical approaches vary greatly regarding the number of different roles to be identified and their denomination: as a minimum, ten roles have been identified as occurring in the great majority of languages. For the most part, these adhere to a hierarchy based on their ability to function as a subject or object in utterances:

Agent → *Goal* → *Recipient* → *Beneficiary* [→ *Experiencer*] → *Instrument* → *Location* → *Time* (adapted from Dik, *Functional Grammar*, 1978: 70, and Fillmore, *The case for case*, 1968).

The first four semantic roles are by far the most frequent and are easily recognisable:

1. The agent (German *Agens* [= AGT]) is the acting or planning instance within a sentence (the 'instigator of the event' after Fillmore); it represents an agentive or causative role, and is typically animate.
2. The patient or, alternatively, goal (German *Patiens* [= PAT]) is the object affected by the event or action expressed by the predicate; it is typically inanimate.
3. The recipient (German *Dativ* [= REC]) is the receiver or addressee of a transaction denoted by the predicate (e.g. *dire/parler à* "say/speak to"; it is typically animate. In some languages, it can be placed on the same hierarchical level as the beneficiary (4).
4. The beneficiary (German *Benefaktiv* [= BEN]) may assume the position of the recipient, introducing different semantic implications with respect to the latter. A beneficiary is typically a human who benefits from the accomplishment of the action described.

These first four roles appear in a prototypical configuration in sentences such as the following:

Fr.	<u>Jean</u> _{AGT}	<u>donne</u>	<u>le livre</u> _{PAT}	<u>à Marie</u> _{REC}	<u>pour Pierre</u> _{BEN}
cf. English	<u>John</u> _{AGT}	<u>gives</u>	<u>Mary</u> _{REC}	<u>the book</u> _{PAT}	<u>for Peter</u> _{BEN}
Pg.	<u>O João</u> _{AGT}	<u>representa</u>	<u>os computadores</u> _{PAT}	<u>ao público</u> _{REC}	
cf. English	<u>John</u> _{AGT}	<u>is presenting</u>	<u>the computers</u> _{PAT}	<u>to the public</u> _{REC}	

The following four roles are less prototypical, but are nevertheless frequent and easily recognisable:

5. The experiencer [= EXP] is the (usually animate) locus of a (psychological or concrete) action: *Jean souffre* ‘John is suffering’; however, cf. also *le tonneau fuit* ‘the barrel is leaking’ (after Filmore, it is ‘the entity which receives or accepts or experiences or undergoes the effect of an action’).
6. The instrument [= INST] is an inanimate entity by means of which an action is carried out: *la clé ouvre la porte* ‘the key opens the door’.
7. The location [= LOC] identifies the place of an action or event: *les clés sont dans le tiroir* ‘the keys are in the drawer’.
8. Time [= TEMP] specifies the temporal framework within which an action or event takes place: *elle est partie hier* ‘She left yesterday’.

A series of less frequent and less specific roles also exist, such as manner and result.

The first four roles, which are the dominant ones, and the four complementary roles are characterised by clusters of semantic features that are fairly easy to specify: the agent is responsible for an action, the patient (or goal) is the target of an action, etc. Their concrete realisation depends as much on the lexical meaning of the noun involved as it does on that of the verb (cf. ch. 6, particularly 6.2 and 6.3). Although the two categories of syntactic and semantic roles do not directly depend on one another, they display certain affinities: the syntactic subject often corresponds to the semantic agent, which is the dominant semantic role, etc.

The agent is the most salient semantic role; its semantic content relies on the realisation of the distinctive features of ‘human’ or ‘causative’. Each of these features, in turn, takes its place within a hierarchical continuum based on degrees of saliency. The most important of these is the ‘animacy hierarchy’ (the concept of which originated in phenomenology or *Gestalt* psychology). The degree of salience of entities which can occur as agents is measured by their respective degree of resemblance to the ‘speaking subject’. The most prototypical agent is thus a ‘speech act person’ (that is, one of the persons directly involved in the utterance act), followed – in order of decreasing prototypicality – by ‘humans in general’, ‘discontinuous and concrete entities’ (such

as animals, plants and countable objects), ‘mass nouns’ (i.e. uncountable entities), and finally ‘abstract entities’.

Thus, a hierarchy emerges which is easy to interpret in terms of distinctive and salient features and which plays a major role in lexical semantic change (cf. 9.3.4):

speech act person → humans in general → animals → plants → countable objects
(natural / artificial) → uncountable objects → abstract entities

This hierarchy is fundamental to both grammatical and lexical semantics, and may therefore be equated with the basic relationship between the predicate and its arguments central to the proposition (cf. 8.1). It is anthropocentric and is based directly on the nature of human perception. The animacy hierarchy is also of importance for the remaining semantic roles, which it actualises in various ways: the most prototypical patient, for instance, is a countable object.

Other semantic hierarchies parallel to the animacy hierarchy are less specific. Proximity between a given entity and a prototypical agent is measured by the degree of causativity (immediate, indirect, absent), intentionality (strong, weak, absent) and control over the action (strong, weak, absent). According to these parameters, a prototypical agent is human, is an immediate instigator (of the action), acts with unequivocal intentionality, maintains complete control of the action and can be immediately identified as its instigator (salience).

These basic rules are considerably more complicated in concrete sentences since a wide variety of combinations between syntactic roles and semantic roles are possible: the subject of a sentence may play the semantic role of an agent, but it may also appear in the role of a patient or recipient. The following examples illustrate how the same subject or object can fulfil different semantic roles:

<i>le couteau lui a fait peur</i>	“the knife frightened him”	(subject - agent/cause)
<i>le couteau est propre</i>	“the knife is clean”	(subject - experiencer)
<i>le couteau coupe bien</i>	“the knife cuts well”	(subject - instrument)
<i>Luc nettoie le couteau</i>	“Luke is cleaning the knife”	(object - patient/goal)
<i>Luc a fabriqué ce couteau</i>	“Luke produced this knife”	(result) (cf. GME, p. 238)

Nevertheless, semantic roles are a useful concept for the understanding of certain grammatical phenomena, allowing, for example, a better interpretation of changes affecting valency and diathesis. An argument may be omitted or displaced depending on its degree of prototypicality:

Paul fume une cigarette “Paul is smoking a cigarette”
→ *Paul fume beaucoup* “Paul smokes a lot”

In the second example, *une cigarette* can be omitted due to its high degree of prototypicality in combination with the verb *fumer*.

3 Argument marking in Romance – the subject

In the Romance languages, arguments are marked by a combination of positional and inflectional markers, including agreement as well as ‘zero marking’ (i.e. the absence of explicit marking, *cf.* 8.4.1). The subject, in particular, which appears in the large majority of sentences, usually bears no grammatical marking in most of the world’s languages (the marked masc. sg. subject case in Old French and Old Occitan constitutes an exception, *cf.* 7.3.4). The grammatical identification of the subject thus relies on the person/number agreement with the verb and on its position within the sentence: in modern Romance languages an agent subject is usually placed before the verb; in the case of a patient subject with unaccusatives, by contrast, the unmarked position is postverbal. However, in French – and other Gallo-Romance varieties – the preverbal position is generally grammaticalised (SV, *cf.* no. 2 above).

The non-expression of the subject is not unusual in most Romance languages, and even represents the unmarked case in Italian:

- no subject: It. *piove* “it’s raining”, Sp. *llueve* “id.”; Pg. *chove* “id.”, *é quinta feira* “it’s Friday”; things are more complicated in French: in the equivalent *il pleut*, *il* functions as a ‘grammatical’ subject (or ‘dummy’ subject, *cf.* below)
- covert subject: It. *vengo* / *viene* “I’m / he is coming” vs. (less common and marked) *io vengo*, *lui viene*

In Latin, moreover, passive sentences involve a reduction of valency that results in the omission of the subject (‘recessive diathesis’: *bene editur* “it is well published”, *cf.* 8.3.2 no. 5.2). In the Romance languages, however, the subject normally occupies the highest position on the argument hierarchy; it alone determines the agreement in person and number of the conjugated verb at the clausal level.

While the theoretical status of full subjects is relatively clear, their replacement by clitic subject pronouns (Fr. *tu chantes*) raises problems of interpretation and tends to differ among the Romance varieties. According to some approaches, personal subject pronouns are analysed as inflectional affixes bound to verb forms; others consider these pronominalised forms to be arguments in their own right with the value of a full subject. To harmonise the two positions, it can be argued that in French the subject pronouns are phonological clitics but full syntactic subjects, whereas in many dialects of northern Italy the subject pronouns are considered to be both phonological and syntactic clitics.

Some Romance languages tended to develop a series of unstressed clitic forms (*je*, *tu*, ...) alongside the existing stressed pronouns (*moi*, *toi*, ...). French represents an extreme case in this respect, as the expression of subject clitics has become obligatory. According to the long-standing theory that assumes the existence of a ‘subject conjugation’ in French, the Latin personal pronouns were transformed into grammemes or argument markers that were fully integrated into the verbal paradigm.

It should be noted that the expression of subject clitic pronouns in French became obligatory during the 16th to 17th centuries. In Old and Middle French, by contrast, the expression of subject pronouns gave rise to pragmatic differences (involving focus and topic shift), as is still the case in Spanish and Italian today. In older language stages, these pronouns were thus employed for the purpose of semantic intensification (as in *toi, tu crois* in Mod.Fr.).

In French, moreover, subject pronouns are omitted in the presence of a nominal subject (*mon père vient* “my father is coming”); this argues in favour of an interpretation as arguments in their own right. Informal French is more complex in this respect, as it displays double subject constructions (*mon père, il vient*, literally “my father, he is coming”). In order to uphold the hypothesis that the subject clitic retains a status of syntactic autonomy in constructions of this type, it is analysed as an instance of ‘dislocation’, in which the pronoun *il* moves to the left, so as to occupy the syntactic position of the subject (cf. the overview of this issue provided by Auger, *Les clitiques pronominaux en français parlé informel*, 1995).

The transitional status of subject clitics becomes even more apparent when one compares them to object clitics (cf. no. 3.4 below): the latter clearly function as arguments, despite the fact that they are not autonomous (*il me le donne* = *il donne cela à moi*). Such forms belong both to the NP (since they are reduced nominal constituents) and to the VP (since clitics are formally integrated into the verb phrase).

Like the formation of verbal periphrases, the integration of fixed subject clitics into the VP is a prototypical example of secondary or re-grammaticalisation (cf. 5.1.3 and Bossong, *Vers une typologie des indices actantiels*, 1998).

→ D'Alessandro, *Agreement*, CambrHandb 17

4 Romance argument markers – 1st object (O1) and 2nd object (O2)

While in all Romance languages the subject is linked to the semantic role of the agent, the 1st object only partially corresponds to the semantic role of the patient. Rather, this ‘second argument’ (or ‘second actant’ according to Tesnière) corresponds to a direct object. The 1st object does not necessarily carry grammatical markers, and in some cases it can be distinguished from the subject by its position alone, following the verb (VO). The latter constellation is found in French and standard Italian.

Fr. *le garçon voit le film* / *il prend le train* / *il admire Michel*

It. *il ragazzo vede un film* / *prende il treno* / *ammira Michele*

Pg. *o rapaz vê um filme* / *toma o comboio* / *admira Miguel*

[Engl. “the boy is watching the film” / “he is catching the train” / “he admires Michael”]

In the other Romance languages, the 1st object may display grammatical marking. In these cases, such marking is determined by semantic factors and relates to the animacy

hierarchy. Referred to as ‘differential object marking’ (DOM), it concerns, above all, terms designating human beings:

Sp. *el niño ve a su hermano* vs. *el niño ve el árbol*

Rom.n *băiatul îl vede pe fratele său* vs. *băiatul vede copacul*

[Engl. “the child sees his brother” vs. “the child sees the tree”]

Less frequently, DOM is used for other animate beings or even objects or abstract entities:

Sp. *el entusiasmo venció (a) la dificultad* “the enthusiasm overcame the difficulty”

In standard Italian and in some Romance varieties such as Walloon or the Romansh variety of Lower Engadin, these markers only appear when the argument in question is a ‘speech act person’; DOM is thus used to highlight 1st objects whose semantic role corresponds to that of a prototypical agent (human, speech act person).

In Spanish, differential object marking began to replace the unmarked 1st object from the 12th century onwards; positional marking was thus the original starting point in all Romance languages. In French – which does not exhibit DOM – partitive articles were introduced to mark a second argument corresponding to a non-countable object (*ils préfèrent boire du vin* “They prefer drinking wine”; *cet homme a du courage* “this man has courage”; cf. Lazard, *Le marquage différentiel de l’objet*, LangTyp, art. 65).

In most Romance languages, the 2nd object is marked by the function word *a*, which is derived from the Latin preposition *AD* (e.g. *donner de l’argent à Etienne* “give [INF] money to Stephen”, *ne rien dire à personne* “not tell [INF] anything to anyone”, Sp. *darle dinero a Esteban*). Romanian, which has systematically retained a higher degree of inflection, can show positional marking with no prepositional device (e.g. *a-i da bani lui Ștefan*); however, it can also display prepositional marking, especially with non-human referents (e.g. *a da mânca la porci* “feeding pigs”).

This results in at least partial object conjugation for the 2nd object in Romanian. This is particularly characteristic of Spanish, parallel to differential marking of the 1st object. In Spanish, a marked 2nd object in initial position must be doubled by an argument marker consisting of a clitic pronoun (*a Juan le duele la cabeza* “John has a headache (= John feels pain in his head”); in postverbal position, the clitic is frequent but not obligatory (*(le) he dicho la verdad a mi padre* “I have told my father the truth”). The grammaticalised nature of this phenomenon is highlighted by the absence of agreement: fam. Sp. *le he dicho la verdad a mis padres* (rather than the expected *les*).

In addition to this redundant form of object conjugation for the 2nd object, there is a complementary form for the 1st object, which is common in the Romance languages (Fr. *je le vois*, It. *lo vedo* vs. **lo vedo l’uomo*).

As far as possible, object conjugation should be distinguished from dislocated sentences that display pronominal redundancy: a construction such as *le parfum, je l’aime*

reflects a specific communicative intention, but not a form of object conjugation (the same applies for oft-cited sentences such as *moi, de l'argent, à ce filou, je ne lui en donnerai jamais* (literally “me, money, to this rogue, I’ll never give him any”), Bossong, *Vers une typologie des indices actanciels*, 1998: 36).

Like subject clitics, object clitics reflect a grammaticalisation process that has reached different stages in different Romance languages:

- the pronoun first of all becomes fixed within the verbal group, but occupies a variable position depending on prosodic factors: in sentence-initial positions, the clitic follows the verb in Portuguese (*conheço-te*); in specific non-initial position, it precedes the verb (*não te conheço* “I don’t know you”).
- a second stage has been reached when the position of clitics no longer depends on prosody, but purely on grammatical parameters such as the status of the verb (= full or auxiliary) or verbal modality; in French, they are proclitic to the infinitive (i.e. in preverbal position) (*sans le dire*), whereas Sp. and It. display enclisis (i.e. the use of clitics in postverbal position), e.g. *sin decirlo*, *senza dirlo* “without saying [INF] it”); the three languages display proclisis in indicative forms and enclisis in positive imperative forms (*tu me le donnes* → *donne-le-moi*, *me lo dai* → *dammelo* “give it to me”).
- the fixation of the position of clitics is complicated by the presence of auxiliaries within the verbal framework (e.g. It. *voglio farlo* vs. *lo voglio fare*); the concrete combination of different clitics, modal verbs or auxiliaries and predicates has been subject to intense variation throughout the historical stages of the Romance languages.

In French, for example, a clitic functioning as the complement of an infinitive could still be placed in front of the auxiliary during the 17th century (as it still is today in causative constructions with *faire/laisser*, e.g. *elle m’a fait chanter*, not **elle a fait me chanter*); in these constructions, the direct object was placed before the indirect object, as was the case in earlier stages of the language: *je le vous ai voulu dire* (Malherbe III, 254) → *j’ai voulu vous le dire* “I wanted to tell you (it)”; cf. also the word order in Mod.Fr. in sentences such as *elle n’a pas encore pu le lui expliquer* “She hasn’t been able to explain it to him yet”.

Furthermore, in Italian, Spanish and Portuguese, clitics can still be attached to the inflected verb (= auxiliary) rather than the main (lexical) verb. Finally, in Romanian, clitic climbing is not merely an option but obligatory (e.g. *îl pot face*, not **pot îl face* “I can do it”).

5 Arguments and circumstantial adjuncts

In summary, the subject is unmarked, the 1st object is generally marked by position (or by a preposition in the case of DOM), while the 2nd object is always marked by a function word (also a preposition), with the exception of Romanian.

The construction of concrete sentences depends on verbal valency frames, which are highly variable. These frames introduce lexical elements which impose constraints and which are responsible for the large degree of formal variance characterising the resulting syntactic structures, thus further complicating the analysis of argument markers.

It is especially important to distinguish objects introduced by a preposition (e.g. *se réjouir de son succès* “rejoice [INF] in his/her own success”, *se diriger vers le lac* “head [INF] towards the lake”, Sp. *pensar en Juan* “think [INF] about John”) from circumstantial adjuncts that may have identical grammatical markers (e.g. *il est parti en vacances*₀₁ “he has gone on holiday” vs. *il est parti en avion*_c “he has gone on a plane”).

Depending on the verb or the particular meaning, the markers employed for the prepositional 1st object (an argument, which is an obligatory element), as well as those accompanying circumstantial adjuncts (which are optional) consist of a wide range of prepositions, which are all of Latin origin.

6 Latin, Romance and creole systems

We have seen that the Latin system, which largely relied on inflectional marking, was replaced by a system of combined marking relying on positional and grammatical markers, in which the agglutinating type is dominant and the inflectional type is negligible, at least in the nominal system.

Grammatically speaking, however, the Romance system is no less complex than that of Classical Latin. The abandonment of the latter system forms part of a development which took place on a macroscopic scale, from Proto-Indo-European (which possessed a more flexional system than that of Latin) to the modern Indo-European languages. Phonetic change may have played a role, as the loss of final consonants and vowels beginning in the 1st century AD could potentially have contributed to the weakening of the Latin inflectional system. The structure of the diasystem, characterised by intermittent phases of elaboration and the predominance of oral communication, may have had a certain impact. To date, however, there is no coherent and generally accepted explanation for the grammatical reorganisation which took place between Latin and the Romance languages.

The Romance languages are nevertheless far removed from the systems which may be observed in creole languages with a Romance lexical basis, which make exclusive use of positional marking for arguments. Creole languages adhere rigidly to a defined constituent order:

SUBJ – TAM markers – NEG – V – OBJ2 – OBJ – COMPL ADV/PREP

This prototypical order is the only means by which syntactic roles are marked in these languages. Roles are therefore not encoded by means of grammatical or prosodic marking; cf. the following example of Haitian Creole:

<i>li</i>	<i>rakonte</i>	<i>papa-li</i>	<i>istwa</i>	<i>sa-a</i>
S-3SG	V	O2-3SG	O1	PronDem
“he” (<i>il</i>)	“told” (<i>raconta</i>)	“his father” (<i>à son père</i>)	“story” (<i>histoire</i>)	“this” (<i>cette</i>)

Source: Muysken, *Creolization*, LangTyp, art. 117, p. 1660.

Consider also the example of Caribbean Papiamentu:

<i>bo</i>	<i>a</i>	<i>duna</i>	<i>mi ruman</i>	<i>e</i>
S-2SG	TAM-past	V	O2-3SG	O1-3SG
“you” (<i>vous</i>)	“have” (<i>avez</i>)	“given” (<i>donné</i>)	“my brother” (<i>à mon frère</i>)	“it” (<i>le</i>)

Source: *ibid.*

In sign languages, constituent order is similar but less rigid; here, the SVO order is concurrent with the SOV order.

In conclusion, it is clear that the system exhibited by the Romance languages is more widely distributed among the languages of the world than the Latin or creole systems, both of which display only one type of marking.

8.4.4 Information structure (topic-comment)

Information structure does not relate exclusively to the domain of the proposition *per se*. Rather, it creates a link between sentence and context, and assumes specific pragmatic functions. The distinction of a theme and a rheme presupposes a binary clause that starts out from given information (i.e. the theme), about which new information (i.e. the rheme) is provided; individual sentences have only one theme. This configuration is dominant in modern Romance languages; it is correlated with the SVO order, which attributes a salient position to the subject. Specifically, a prototypical subject is an agent, which, at the same time, assumes the function of a theme.

This configuration is not infrequent among the languages of the world, since the theme generally tends to be placed in first position and thus to correspond to the subject. Even in Latin, which displays an extremely variable word order, the theme–rheme sequence is used in unmarked constructions. At the same time, this positional freedom easily allows the rheme to be displaced for the purpose of emphasis. This type of displacement, which serves informational or pragmatic purposes, is more difficult to achieve in Modern French, where positions are fixed and tied to specific semantic roles.

Consequently, French uses other means to create emphasis, in particular, dislocation (*l'abeille a piqué mon doigt* “the bee stung my finger” → *elle l'a piqué, mon doigt, l'abeille / mon doigt, l'abeille l'a piqué*, literally “it stung it, my finger, the bee / my finger, the bee stung it”; cf. also constructions that lack verbal elements: *admirable, ce tableau* “admirable, this painting”); such constructions result in the pronominal repetition of certain elements (cf. the example cited in no. 3.4 above: *moi, de l'argent, à ce filou ...* “me, money, to this rogue ...”; or other forms without dislocated theme-rheme constructions: *l'amour, elle appelle ça*; or *ses mains seulement qu'on voyait*, literally “love, she calls that”; “his/her hands only that could be seen”). Spanish occupies an intermediate position between Latin and French in this respect (*Jorge da el libro a su hermano* “George gives the book to his brother” → *a su hermano le da el libro Jorge*, literally “to his brother gives the book George”).

Moreover, differences of position between Latin and the Romance languages concern the verb in particular, which is in final position in Latin, in 2nd position in the Romance languages (cf. 8.4.2), and sometimes even in 1st position, in cases of mono- or bivalent verbs (existential or presentative verbs such as *il existe ...*, etc.) or in cases of the ‘rhematisation’ of the subject (*bailaba la gente en la calle* “people were dancing in the streets”).

The information structure of the sentence thus makes use of different utterance techniques with variable semantic values. True rhematisation by means of the displacement of the object cannot be achieved without some degree of positional freedom. If this is not possible, dislocation (‘left dislocation’ or ‘right dislocation’) or ‘clefting’ necessarily occurs (to be further distinguished from pseudo-cleft constructions such as *j'ai mon vélo qui est cassé*, literally “I’ve my bike which is broken”; Sp. *él que canta bien es Pedro*, literally “He (the one) who sings well is Pedro”). In addition, there is some interaction between information structure and recessive diathesis: the latter causes the agent to disappear and the patient to be thematised (instead of *mon doigt, l'abeille l'a piqué*, literally “my finger, the bee stung it”, a diathetic passive construction results in: *mon doigt a été piqué (= par l'abeille)* “My finger has been stung [= by the bee]”). In spoken language, prosody also plays an important role in informational emphasis.

- Cruschina, *Information and discourse structure*, OxfGuide 34
- Giurgea/Remberger, *Illocutionary force*, OxfGuide 53
- Cruschina/Giurgea/Remberger, *Information structure*, CambrHandb 26
- Cruschina, *Focus and focus structures in the Romance languages*, OxfEnc
- Id.*, *Topicalization in the Romance languages*, OxfEnc
- Manoliu, *Pragmatic and discourse changes*, CambrHist 1, 9

8.4.5 Subordination

Subordination (or ‘hypotaxis’) is the most complex kind of syntactic relationship that can be governed by the fundamental relation between the predicate and its arguments (cf. 8.1 and Bosson, LangTyp, art. 48: 658). It establishes an explicit relationship between a main clause and a subordinate clause by means of subordinating conjunctions or relative pronouns. As such, it stands in contrast to ‘parataxis’, which refers to the juxtaposition of clauses with relations that are often implicit (cf. the extreme case of juxtaposition in the absence of coordinating conjunctions, e.g. Fr. *Les chiens aboient, la caravane passe* “dogs bark (but) the caravan goes on” or *Vienne la nuit, sonne l’heure / Les jours s’en vont, je demeure* “Comes the night, sounds the hour / The days go by, I endure”, Apollinaire (transl. A. S. Kilne); Lat. *veni, vidi, vici* “I came, I saw, I conquered”).

From a syntactic perspective, subordination introduces a subordinate clause in place of an argument or a circumstantial adjunct within the main clause (consider Sp. *me gusta que vengas a verme*, literally “I like that you come to see me” which corresponds to *me gusta el pescado*, “I like fish”; the 1st object slot is filled by a complement clause). By means of hypotaxis, in other words, clauses are transformed into arguments and subordinated under a further predicate.

A subordinate clause may fulfil the role of the subject or an object within the main clause (we speak of ‘argument clauses’ or ‘complement clauses’) or it may assume the position of a circumstantial adjunct (‘adjunct clauses’ or ‘adverbial clauses’). Adjunct clauses are introduced by specific conjunctions, at least in the Romance languages (e.g. in Sp. *como, según, a pesar de que*, etc.: *lo haré puesto que lo dices* “I will do it since you say so”, *lo digo a fin de que todos lo sepan* “I’m saying it so that everyone knows”).

Subordinate clauses may also form noun complements, which typically occur as relative clauses (e.g. Sp. *vengo del pueblo en que nací* “I come from the town where I was born”, which refers back to *pueblo*; *cogimos la fruta que estaba madura* “we picked the fruit that was ripe”, which refers back to *fruta*; Fr. *Le petit chat [que nous avons adopté] amuse beaucoup les enfants* “The little cat [that we’ve adopted] really amuses the children”, cf. 8.2.4).

From a universal perspective, subordinate clauses are marked either by word order alone (minimal marking) or by complementary grammatical markers such as conjunctions or modal forms (maximal marking). The Romance languages use maximal marking, which most often takes the form of a specific conjunction (Fr. *que*, Sp. *que*, It. *che*), doubled in many cases by the subjunctive mood (e.g. in French, when the subordinate subject is placed before the main clause: *Qu’il vienne, ne me surprendrait pas*)²⁹.

²⁹ The modal semantic value inherent in the subjunctive (cf. 8.4.5), leads to semantic ‘conflict’ in some cases: the meaning of epistemic verbs (*savoir, constater, dire*), for example, stands in opposition to this modal value; consequently, the use of the subjunctive in a subordinate clause that is governed by such

For some epistemic forms, use of the indicative is commonplace in subordinate subject clauses:

il est vrai / évident / certain que les gens souffrent beaucoup, It. *è vero / evidente / chiaro / certo che la gente soffre molto* [Engl. “it’s true / obvious / clear that people are suffering greatly”]

cf. also object clauses: Fr. *tout le monde sait que la terre est ronde* “everyone knows the Earth is round”, Sp. *todo el mundo sabe que la tierra es redonda* “id.”, It. *vediamo che il caffè è finito* “we (can) see that the coffee is finished”, Pg. *o ministro declarou que as eleições terão lugar* “the minister declared that the elections are going to take place”

In such cases, the indicative may become established as the normal and natural mood used in subordinate clauses, with the subjunctive becoming the deviating mood.

In ‘Eastern’ Romance (encompassing Romanian as well as southern and central-southern varieties of Italian), patterns of subordination using two complementary conjunctions developed as a result of the influence of Greek. Successors of Lat. *QUIA* (Rom.n *că*, south. It. *ca*) appear in conjunction with epistemic verbs and the indicative mood; a different conjunction (Rom.n *să*, north-eastern Sic. *mi*, Cal. *mu/mi/ma*, Salentino *cu*) may be required by subjunctive contexts (although expressed by the same forms as the indicative in these varieties). In this situation, infinitives within the sentence are replaced by person-inflected subordinate constructions: Cal. *vogghiu cu dormu* “I want to sleep” (here with the indicative) vs. It. *voglio dormire* “id.”.

Subordinate clauses without conjunctions are less common in the Romance languages: we have already mentioned the Portuguese *infinitivo pessoal* (‘personal infinitive’) (cf. 8.3.3 no. 5: *falar, falar-es, falar-mos, falar-des, falar-em*), which produces subordination without a specific modal value (*é conveniente ter-es cuidado* “it’s necessary for you to pay attention”). In Spanish or Italian, there are constructions that do not use conjunctions (= ‘complementiser deletion’) but require the subjunctive (or another irrealis form, such as the conditional or future) after specific verbs (*pedir* “to ask, request”, *querer* “to desire”, *temer* “to fear”): *los usuarios piden Ø se amplíe el horario del bar* “the regular customers asked if the bar’s opening hours could be extended”.

In Spanish these constructions can also be licensed by more complex subordinate structures, as in *La mujer que Juan supone fue responsable del crimen* “The woman John assumed was responsible for the crime”. According to a recent diachronic study (Laca and Bertolotti, *Orden de constituyentes*, forthcoming), such constructions, also common

verbs leads to modal conflict. This explains some variations in the use of the subjunctive, such as *je crois qu’il est malade* “I believe he is ill” vs. *je ne crois pas qu’il soit malade* “I don’t believe he is ill”.

after verbs of request or demand, were a feature of 16th-century Spanish which subsequently disappeared.

To summarise, subordinate argument (or complement) clauses may be introduced by a subordinating conjunction (e.g. Fr. *je préfère que tu ne viennes pas* “I’d prefer you not to come/I’d prefer (it) if you didn’t come”, conjunctive complement clause), by interrogatives (*J’aimerais savoir pourquoi il est si triste* “I’d like to know why he is so sad”, interrogative complement clause) or they may take the form of infinitive clauses, with or without an introductory complementiser (Fr. *J’ai peur de chanter devant tous / J’entends Ø chanter les oiseaux* “I (can) hear the birds singing”, infinitive complement clause).

The position of adjunct clauses introduced by conjunctions may also be occupied by participial clauses, which are not introduced by a subordinating function word (*Les vacances terminées, les cours ont recommencé* “As soon as the holidays were over, classes began again”).

Aside from its functions as an argument (or complement) and adjunct (or adverbial), the subordinate clause may also fulfil the role of an attribute (*mon avis est que tu as raison* “in my opinion, you’re right”) or an appositive element (*Il est grand, et qui plus est, très fort* “He’s tall, and what’s more, he’s very strong”).

To conclude the discussion of subordination, a few diachronic observations should be made:

1. Romance variation in subordinate constructions only rarely concerns the position of conjunctions. In contrast, the use of the subjunctive or the choice of conjunctions is often subject to change. All in all, the diversity of constructions is (almost) infinite and lends itself to numerous research questions, particularly those relating to textual genres.
2. In written texts from the medieval period, we observe an increase in the use of subordinate clauses, correlated with a decrease in the use of participial constructions. The latter were nevertheless reintroduced through linguistic standardisation and elaboration (based on the imitation of Latin) during the 16th century. After a period of drastic reduction in Late Latin, the whole system of subordinating conjunctions underwent gradual modification and innovation in the Romance languages. Still today, sentence constructions containing a large number of subordinate clauses generally belong to the written register and are developed through linguistic elaboration.
3. The simplification of relative subordination is a frequent phenomenon, not only in the Romance languages but in modern languages in general. This essentially occurs in spoken language and leads to the use of an all-purpose subordination marker (*que*, It. *che*) or even to the omission of the conjunctive marker (e.g. *tu sais, la fille [qu’]il connaissait le père !*, literally “you know, the girl [that] he knew the father of !”). Examples can nevertheless be found in Romance varieties from earlier historical periods (cf. D’Achille, *Sintassi del parlato e tradizione scritta*, 1990).

- Abeillé, *Coordination in syntax in the Romance languages*, OxfEnc
 Ledgeway, *Clausal complementation*, OxfGuide 63
 Stark, *Relative clauses*, OxfGuide 64
 Mensching, *Infinitival clauses in the Romance languages*, OxfEnc

8.5 Perspectives: the grammatical description of Romance and grammatical change

8.5.1 Grammatical description of the Romance languages: observations

Grammatical research on the Romance languages was very productive throughout the 20th-century, as was the publication of grammars, and both are still in full swing today. Studies focusing on a broad range of individual topics such as the article, the relative clause, past tenses, etc., could fill entire libraries. Most of these studies focus on a single Romance language, or even on a particular variety of a Romance language. More often than not, it is a standard, or modern literary language, though it may also be a non-standard present-day variety or the language (literary, in most cases) of a given historical period (cf. 11.1.1).

The oldest tradition of modern grammatical studies on the Romance languages is undoubtedly of French origin, beginning with the ‘Port-Royal grammar’ (*Grammaire générale et raisonnée*, 1660; cf. 10.5.3 no. 3 for Romance precursors in the 16th century). The publication of a grammar in seven volumes, *Des mots à la pensée. Essai de grammaire de la langue française*, by Damourette and Pichon (1911–1945), marked the beginning of the 20th century, appearing shortly after Nyrop’s historical grammar of French in six volumes (*Grammaire historique de la langue française*, 1903 [1899]–1930). The following decades witnessed the publication of Grevisse’s *Bon Usage*, further editions of which have appeared at regular intervals, each time with additions (¹1936, ¹³1993 – the last edition was undertaken by Grévisse’s stepson Goosse). The last decade of the 20th century saw the appearance of the *Grammaire méthodique du français* (GMF, Riegel / Pellat / Rioul ¹1994, ⁸2021), inspired both by the French and the Anglo-American traditions of grammatical research, the latter only partially generative. Recent contributions to grammatical description include the *Grande grammaire du français* (Abeillé / Godard, 2 vols., 2021), with 59 contributors, as well as – for diachrony – the *Grande grammaire historique du français* (Marchello-Nizia *et al.*, 2 vols., 2020), with 34 contributors, and, finally, Buridant’s monumental personal achievement, the *Grammaire du français médiéval* (¹2019, ²2022, which replaces a preliminary version published in 2000).

Several important grammars of Italian appeared in the 1980s and 1990s: Serianni, *Grammatica italiana*, 1988; Renzi, *Grande grammatica di consultazione* (3 vols., 1988–1995, ²2001); Schwarze, *Grammatik der italienischen Sprache* (1988), and, for Old Italian, Salvi/Renzi (eds), *Grammatica dell’italiano antico* (2 vols., 2010).

As for Spanish, the first grammar to incorporate linguistic variation to a large extent was the *Gramática descriptiva de la lengua española* edited by Bosque and Demonte, which did not appear until the end of the 20th century (1999, 3 vols.); cf. also the generativist work by Hernanz and Brucart, *La sintaxis: principios teóricos* (1987), as well as Ramírez's *Gramática española* (volume 1 appearing in 1951, the remaining volumes posthumously in 1985–86); the latter is a revised version of the *Gramática de la lengua española* published by the Real Academia española in 1928 (*Nueva Gramática de la lengua española*; vol. 1: *Morfología y sintaxis*; vol. 2: *Sintaxis II*).

Many works of grammatical description written from a synchronic perspective also describe elements of phonetics and orthography, or even semantics and word formation. Older 'historical grammars' tended to concentrate on phonetic and morphological elements to the almost complete exclusion of syntax, the study of which was less advanced in the first half of the 20th century. Explanatory models adopted by these works rely heavily on theories of grammar elaborated during Antiquity for Greek and Latin. In contrast, recent grammars such as the fundamental works of Renzi and Bosque/Demonte are often based on a generative approach. Of a more idiosyncratic nature is Buridant's *Grammaire du français médiéval* (2022), as he assumes a Guillaumian approach.

A general and comparative overview of the Romance languages has not been attempted since Meyer-Lübke's work, published in 1890/1902, now largely surpassed by more recent works (in contrast to the etymological dictionary by the same author: *Romanisches etymologisches Wörterbuch* (REW), which is still the only work of this type based on all Romance languages, cf. 9.9). Nevertheless, there are numerous works devoted to specific aspects of Romance grammar, and the major Romance encyclopedias (cf. 2.4.1) such as the OxfEnc, which employs a balanced methodological approach, combined with the OxfGuide, the CambrHandb, the CambrHist and the LRL, come close to providing a general overview.

Grammar represents an immense field of study as the ways in which morphological and syntactic elements can be combined are extremely varied. Despite the fact that the domain is considered to be 'systemic', the system of grammar displays strong resistance to simple interpretations. The ontological and empirical difficulties which stand in the way of adequate grammatical description arise from the fact that language is deeply rooted in the organisational structures of the human brain. The best results obtained by (semi-)automatic translation are achieved on the basis of large quantities of texts, which allows the identification of habitually related words in certain contexts. This method thus eliminates any necessity for the true comprehension of grammar.

Due to the immense complexity of grammar, every work of grammatical description is destined to remain fragmentary. Even the most extensive grammatical descriptions of a present-day language require prior knowledge of the language in question; while such works facilitate the comprehension of a certain number of features and improve one's command of the language, they do not, in themselves, suffice to enable the production of a text. This is all the more evident when we consider the older stages of modern lan-

guages, as well as extinct languages, where the relationship between linguistic description and usage no longer exists. The differential method – whereby a previous language stage is compared with a current stage – is required wherever possible, although this may lead to a certain amount of inconsistency. The description of grammatical phenomena from a diachronic and dialectological-variational perspective is thus no less challenging than the description of the vocabulary of the different Romance languages and dialects the numerous reference dictionaries attempt to provide (*cf.* 9.9).

What purpose is served by grammatical description, aside from its undeniable intellectual interest or the (no doubt vain) hope of achieving artificial (re)production of language? Above all, grammatical description, in conjunction with lexicography, has a real impact on language evolution, language elaboration and linguistic culture, even though natural language use is possible without grammatical description. Often, speakers of dialects, marked sociolects or languages that have not been put into writing lament the absence of a ‘grammar’ for their particular variety, claiming that it ‘lacks rules’. This does not, however, reflect linguistic reality; at most, it is possible that the speakers of such varieties do not possess a high level of linguistic ‘awareness’ and that their language lacks widely diffused works of grammatical description. Even if human beings are not consciously aware of the grammar of the language they acquire naturally, this language nevertheless constitutes a complete, coherent system (just as it possesses a phonological system and a network of lexical meaning(s)). Its coherence is entirely independent of the existence of grammatical works tasked with its description, or with the prescription of rules.

The existence and wide diffusion of grammars and dictionaries contribute to language standardisation, and thereby to the process of artificial normalisation, as well as to a more intense and efficient use of the expressive potential of language. In the case of morphology, for example, standardisation leads to the alignment and regularisation of paradigms, as well as to the reduction of exceptions: it is no coincidence that the verbal morphology of French, a language with strictly defined rules, is more regular than that of Catalan, which only recently underwent standardisation. Conscious intervention in syntax favours an increase in the complexity of syntactic constructions and thus the development of complex sentences. In contrast, the intense variation in medieval Romance texts can be explained in part by the absence of influential grammatical description at the time. Standard French and Italian are examples of partially artificial languages, whereas familiar and popular varieties of the same languages are entirely natural.

Depending on a more or less overt intention to standardise which pervades published grammars, linguists distinguish between ‘normative’ or ‘prescriptive’ grammars and ‘descriptive’ grammars; while the former intend to regulate language usage, the latter merely attempt to describe it. The choice between the two perspectives naturally depends on the targeted linguistic variety: a grammar of an elite variety (a high literary language, for instance) may intend to be prescriptive, since its model is destined to be

diffused throughout society; in contrast, a grammar of a variety of low social prestige (such as a regional dialect or popular sociolect) cannot aspire to more than a mere description; as a model, such a grammar would be redundant, since those who already speak the language would be highly unlikely to read it. Diffusion is thus of major importance with regard to the prescriptive role of a grammar. At the same time, however, a grammar which is too restrictive would not succeed in achieving the same level of influence as a prescriptive grammar that includes non-exemplary varieties of the language in question to a certain extent, such as everyday spoken language.

Finally, historical – or dialectological – grammatical description may appear to be of minimal immediate interest; indirectly, however, it often proves invaluable: on the one hand, it contributes to improving and refining the grammars of present-day languages, allowing a better categorisation of individual phenomena; on the other, it is an important source for linguistic history, knowledge of which increases the opportunity for a society and its members to develop a balanced linguistic identity.

8.5.2 General conclusions regarding the transformations in morphology and syntax between Latin and Romance

Any attempt to explain the linguistic changes which took place between Latin and the Romance languages must necessarily be founded both on detailed analysis and on observations of a more general nature. To facilitate an understanding of the complexities inherent in the subject matter, general observations relating to grammatical phenomena have played a central role in this chapter.

Despite this broader focus, a large number of changes specific to the Romance languages have been identified and compared with Latin. These include:

- constituent order and the position of grammatical markers
- syntactic role marking and hierarchy (the expression of the subject is reinforced in the Romance languages)
- phenomena of grammaticalisation such as periphrasis and diathesis
- the reinforcement of temporality
- more sporadic phenomena such as the introduction of polite plural forms (which serve as a means to modulate intrusion into the sphere of the speech partner)

Numerous cases of divergence among the Romance languages have also been observed: in several instances, the southern (and sometimes central) Italian varieties, Sardinian and Romanian stand in opposition to northern Italian, as well as to the Gallo- and Ibero-Romance languages. Sometimes, Romanian and Ibero-Romance form a contrast with the ‘central’ languages. Often, one of the linguistic groups displays a solution that differs from the others: for example, in the Ibero-Romance languages, verbal aspect has evolved to a greater extent, while French displays a rigid word order with the

resulting consequences. Moreover, it has been demonstrated that medieval Romance languages often differed less from one another than do their modern counterparts, which supports the hypothesis of increasing diversification over time within the family of Romance languages.

Not all aspects of the exemplary domain offered by the Romance languages for historical and variational observation are of relevance for general linguistics; nevertheless, they provide significant contributions to a variety of fields (e.g. constituent order, noun determination, verbal periphrasis or information structure). Beyond these concrete contributions, the Romance languages allow the detailed study of the evolutionary paths of grammatical forms and functions; this also applies to grammaticalisation, as well as to the links that are established on a more general scale between syntax and morphology on the one hand, and syntax and lexis on the other.

Moreover, the Romance languages offer a particularly promising basis for the study of a variety of further aspects, including the impact of textual genres and discourse traditions on grammatical choices, and the importance of speakers' reflections on their own language in relation to internal changes (*cf.* 10.1.3 no. 4). The exceptionally abundant historical documentation available for these languages has by no means been exhausted, and remains available for research focusing on new topics, following new objectives and using newly-developed instruments.

9 Lexis

9.1 The lexeme: general properties

Lexicology studies the stock of lexemes of a language (= its lexis, lexicon or vocabulary). It delimits, classifies and studies these elements and their structure from the point of view of their semantic and formal characteristics and identifies the relationships between different lexemes and categories of lexemes, determining the conditions and restrictions which govern their usage.

Like all subdisciplines of linguistics, lexicological research always requires the compilation of empirical data as a basis before analysis can take place. In phonetics and phonology as well as in grammar, forms and marking principles are limited in number, and their functions and hierarchical structuring are of primary importance. In lexicology, in contrast, the immense number of lexemes complicates the creation of data inventories, which are indispensable for any interpretative approach. Such lexical inventories may be limited to the vocabulary of an individual text (glossography) or a specific number of defined texts, or they may lead to the compilation of dictionaries (lexicography), which are a privileged source for studies on lexis.

→ RSG 3, sect. XV, *Histoire interne: lexique* (Rumänisch, Ernst, art. 224a; Bündnerromanisch, Liver, art. 229b; italiano, Schweickard, art. 232; français (Europe), Glessgen, art. 237; français (hors d'Europe), Queffélec, art. 238; catalán, Bruguera, art. 243; Spanisch, Geckeler, art. 247; Portugiesisch (Europa), Monjour, art. 253; Portugiesisch (außerhalb Europas), Schmidt-Radefeldt, art. 254; Konvergenz- und Divergenzphänomene in der Romania, Geckeler, art. 258)

OxfEnc, *History of the Lexicon* (Romanian, Iliescu; Raeto-Romance, Grünert; Sardinian, Putzu; Occitan and Gascon, Carles/Glessgen; Spanish, Dworkin; Portuguese, Pöll)

9.1.1 Basic properties of the lexeme

A definition has already been provided for the 'lexeme', which refers to a concrete or abstract concept, in contrast to the function word, which expresses a grammatical relationship (cf. 7.1.2 no. 1). This definition is reintroduced and further specified below:

1 Form and meaning

A lexeme consists of a form (which is phonic and, in written languages, graphic) and a lexical meaning.

From a formal point of view, it always comprises a lexical base (which includes its variant forms), to which derivational affixes (which may include word class markers) and inflectional affixes (which may imply stem alternation) can be added.

Lexical meaning corresponds to an assemblage of information that belongs to the shared knowledge of a community of speakers. In the utterance act, it is evoked by the lexical form, which thus calls forth identical or very similar concepts or knowledge in different persons. The essential element of the lexeme is its meaning, its *raison d'être* in a language and the semantic foundation of all communication.

A lexeme consists of the combination of a given form and a specific lexical meaning; a lexical form may nevertheless be associated with multiple lexical meanings (in such cases, we speak of 'polysemy', cf. 9.2.2 no. 7(8)) and it is necessary to suppose the existence of as many lexemes as there are lexical meanings.

2 Complementary parameters

In addition to the properties mentioned above, a lexeme includes at least five complementary parameters that are linked to its basic structure (one form – one meaning; cf. Glessgen, *Le statut épistémologique du lexème*, 2011: 422):

- inherent morphosyntactic features (part of speech and, for nouns, gender, cf. 7.2.1);
- derivational morphological structure in the case of complex words (a derivative can be associated with other lexemes that have an identical base or an identical affix);
- syntagmatic contexts that are inherent to a lexeme when it is inserted into a sentence, including, in particular, verb valency (cf. 8.3.2 no. 6.2; 9.5.1); a polysemic verb can govern several valency frames, depending on its meaning;
- diasystematic markers that are linked to a lexeme (its 'variational connotation', e.g. as a familiar or regional term, cf. 9.1.4); note that a polysemic form may have different diasystematic markers depending on its meaning.
- finally, the semantic relationships that emerge between different lexical meanings that are close to one another (such as synonymy, polysemy, (co-)hyponymy and antonymy, cf. 9.3.1); in reality, the limits of lexical meaning are defined by the semantic limits of contiguous lexemes: meaning does not exist autonomously but is integrated into semantic networks.

3 Usage and identification of a lexeme

As we have seen, the use of a lexeme is tied to more or less restrictive syntagmatic and diasystematic contexts (in French, one says *prononcer un discours* rather than *dire un discours*, just as in English, one says *to give a talk* and not *to make a talk* (this is referred to as 'collocation', cf. 9.5.2); the synonymous lexemes *voiture*, *automobile*, *bagnole* or *caisse* cannot be used in the same context; the same applies to the English words *car*, *auto*, *automobile*, *wheels*, *jalopy* or *old banger* [diasystematic restrictions]; etc.).

Moreover, a lexeme may be very widely used within a linguistic community (e.g. Fr. *matin*, *arbre*, It. *mattina*, *albero*, Sp. *mañana*, *árbol*) or be very rare: think of special-

ised terms such as *oronym* (cf. 9.1.2 below), *phylogeny* (cf. 1.2.5) or *variationist*; however, the frequency of use does not change the status of the lexeme in any way. Even words formed spontaneously by means of common semantic and derivational mechanisms are lexemes in their own right (we speak of ‘occasional words’, in contrast to ‘lexicalised words’ that have become established in usage). When a new lexeme begins to appear more frequently, it is called a ‘neologism’.

Finally, a lexeme cannot be identified based on formal and morphological criteria alone (cf. 7.1.2); intuitive recognition, which relies on speaker awareness, is also inconclusive. The best results are obtained by means of memorisation tests which aim to identify the semantico-formal lexical entities that are recognised and used by speakers (cf. Glessgen, *ibid.*: 416–421).

9.1.2 Lexemes and proper nouns

Within the category of nouns, it is important to distinguish common nouns from proper nouns. The two types both have an extralinguistic referent, but common nouns have the quality of generalising (*la maison blanche* “the white house” may refer to any one of the millions of white houses in existence), whereas proper nouns have an inherently individualising function (*la Maison Blanche* “the White House” refers to one specific place in the world; cf. 9.7 for onomastics).

Proper nouns include names of persons (‘anthroponyms’) and place names (‘toponyms’), but also the names of mountains (‘oronyms’), animals, mythological figures or works of art (*La Gioconda* “The Mona Lisa”) etc.

Proper nouns always originate from lexemes, though this origin is sometimes no longer obvious (cf. 9.7):

- the origin of the family name *Boulangier* as a name of a profession remains transparent and identifiable for a speaker of Modern French; the family name *Lefèvre*, in contrast, has become opaque (it is based on Lat. *FABER* “craftsman” → “blacksmith”, including an agglutinated definite article);
- the base of the toponym *Villeneuve* can easily be discerned (even if one does not know that this type of toponym usually refers to places that were founded in the 13th century), in contrast to *Antibes* (< Gr. *anti-polis* “the city in front [of Nikaia/Nice]”).

Note that when a proper noun becomes established in usage (that is, when it becomes fixed), the semantic connotations of the lexeme on which it is based are lost (e.g. the toponym *Les Pommiers* refers to a specific geographical locality and not to fruit trees (*pommiers* = “apple trees”). Thus, from this point of view, proper nouns are desemantised.

Proper nouns can, in turn, contribute to the creation of new lexemes, thanks to different mechanisms of ‘deonymisation’:

- a proper noun can be converted to a common noun without transformation (‘conversion’):
e.g. in Fr. *poubelle* “rubbish bin” or *zeppelin*, by metonymy after the names of their inventors; see also cases of ‘antonomasia’: *un Tartuffe* “a hypocrite” (after the main character in Molière’s play *Tartuffe*), by metaphorisation (cf. 9.7.1);
- more frequent is derivation, which may imply a change of word class:
Gargantua (a giant in Rabelais’s novel *Gargantua* → *gargantuesque* “gargantuan, enormous”);
- less known, but also very frequent is the integration of proper names in phraseologisms (cf. 9.5 and 9.7.1) which implies specific semantic knowledge:
Pandore “Pandora” → *une boîte de Pandore* “a Pandora’s box”, *Trafalgar* (referring to the Battle of Trafalgar, in which the French under Napoleon were defeated) → *un coup de Trafalgar* “a disastrous blow”; It. *quarantotto* (referring to the revolution attempts in 1848) → *è successo un quarantotto* “a dramatic event happened”.

Owing to the semantic characteristics of proper nouns, they should be treated separately (within the domain of onomastics), even though their linguistic behaviour is for the most part analogous to that of common nouns. Note that from this point of view, onomastics is a field that belongs to linguistics.

→ 9.7

9.1.3 Number of words and frequency of use

It is very difficult to determine the number of words that exist in a language at any given time; this makes comparisons of the type ‘this language has more words than that one’ hazardous. Today’s major dictionaries and newspapers, however, allow an approximate estimation:

- the PtRob (*Petit Robert*) contains approximately 60,000 entries (‘lemmas’), the TLF (*Trésor de la Langue Française*), about 70,000 to 90,000 entries;
- in both dictionaries, the lemmas group various meanings under a given lexical form, making accurate quantification difficult (the PtRob claims to provide some 200,000 meanings, and thus an average of slightly more than three meanings per form, a figure which seems realistic);
- the German *Duden* in eight volumes contains 200,000 lemmas, while the *Oxford English Dictionary* encompasses 300,000 and provides definitions for at least 500,000 lexical forms (the lemmas include further derivatives – the higher number

is due in particular to the higher quality of English lexicography, but also to a more tolerant concept of language varieties and a favourable attitude towards the integration of loanwords);

- this last figure is comparable to the number of lexical forms contained in a number of American newspapers in any given year: of a total of 227,500 lexical units (a figure which does not account for cases of polysemy), 45,450 are base words, 43,000 are derivatives whose meaning is far removed from the basic meaning and 139,000 are derivatives with an immediately recognisable meaning (*cf.* Steven Pinker's invaluable work *The language instinct*, 1999, ch. 5).

Technical words and terminologies belonging to the different sciences (without even including names of medicines or commercial brands) far surpass these already impressive figures.

The frequency of use of the majority of words is extremely low:

- the 907 most frequently used words contained in the TLF account for 90% of all occurrences in an average French text;
- the following 5,800 words cover 8% of occurrences;
- the remaining 65,000 or 85,000 words account for only 2% of occurrences (Brunet, *Le Vocabulaire français de 1789 à nos jours*, 1981; *cf.* Picoche / Marchello-Nizia, 1994, ch. X, II.2).

Of the most frequently used forms, a large proportion are function words (which are few in number, amounting to approximately one hundred); lexemes of low frequency all carry very specific meanings and allow the establishment of associations with sophisticated cultural knowledge. The knowledge of frequently used words alone does not suffice in order to understand a language or an utterance: as an example, without an understanding of the linguistic terminology used in the present *Companion*, which does not account for even 1% of the forms used, the whole of the text is incomprehensible.

Individual knowledge of words is quite variable: an American BA-level student may have up to 45,000 different lexemes and more than 15,000 proper nouns in his or her passive vocabulary (i.e. words they understand but do not use spontaneously), as well as occasional borrowings, acronyms and idiomatic expressions (Pinker, 1999, *ibid.*); a seasoned writer will have triple this number.

During the process of learning a foreign language, the number of lexical forms acquired by the individual increases considerably (a thorough knowledge of four or five languages does not cause difficulties for the brain's processing capacities); forms from different languages nevertheless draw upon a single repertoire of concepts, to which each language adds only a few entries (*cf.* 9.2.2).

Literary language is highly restrictive in the number of words it uses: the complete writings of the most prolific authors contain up to 15,000 words (Shakespeare)

or 20,000 words (La Fontaine or Molière; Rabelais, whose vocabulary consists of over 30,000 words, is an exception here, which can be explained by his encyclopedic style, *cf.* the database compiled by Brunet, *Rabelais sur Internet*).

The vocabulary of French authors of the Classical period, who used a highly stylised and normalised language, did not exceed 4,000 words (e.g. Corneille, or Racine, who uses 3,263 words; *cf.* H.-J. Wolf, *Französische Sprachgeschichte*, 1991: 125).

Literary languages are diaphasic varieties that played a particular role in the development of current standard languages. The role of secular literature is particularly apparent in Italian, where the vocabulary of Dante and Boccaccio is still very much present in the modern language (*cf.* 10.5.3 no. 3). This fact, combined with the lexically restrictive character of literary genres, means that the knowledge of approximately one hundred words in addition to the vocabulary of a modern Romance language suffices to enable a reader to understand a literary text written in a medieval Romance language.

Although it is difficult to discern the number of words that exist at a given historical moment, the vocabulary of a language at any specific time is in principle finite. In contrast, a language's inventory of idiosyncratic usages – not to mention the sum of the potential words that only exist virtually, in our minds – is theoretically unlimited in number: words can be created *ad infinitum*. The number of words resulting from occasional creativity depends on the language and the historical period.

9.1.4 Variational connotation and communicative context

A word that is rooted in a community of speakers is not necessarily used by all its members, nor in every possible context. A historical language encompasses numerous subvarieties, defined by geographical space, social prestige, context of use, as well as change over time (*cf.* ch. 4). Every word has an individual distribution within such a cluster of varieties.

As an example (taken from Müller, *Le Français d'aujourd'hui*, 1985: 227), the terms *voiture*, *auto*, *automobile*, *bagnole*, *caisse* and *chiotte* “car” are identical with regard to their denotation (i.e. they have the same meaning); however, they do not belong to the same language registers (they carry different diastratic and diaphasic connotations). Similarly, *cornet*, *sachet* and *poche* have the same meaning (“bag”), but are used in different regions of France (they thus have different regional connotations).

A word may belong to a single variety or to more than one variety; in the first case, it is more strongly marked from a variational point of view: it has a ‘connotation’. The semantic connotations are particularly apparent when such a word is used outside its typical context. It is precisely for this reason that we can speak of ‘popular connotation’ with regard to French, for example, when a word such as *caisse* appears in a novel.

Different types of variational connotation are part of the complex of varieties of a historical language (according to Coseriu's use of the phrase, *cf.* 4.3.1). In this sense, words may reflect:

- regional or diatopic connotation, which is apparent when words associated with a specific region or country are used in another: e.g. *septante* “seventy” carries specific connotations in France, where *soixante-dix* is the habitual form, but not in Belgium, Switzerland or Canada; *mis papás* (American Sp.) would seem surprising in Spain, where *mis padres* is expected;
- diastratic and/or diaphasic connotation, in words belonging to different registers (e.g. formal, informal, popular or slang) and lexemes linked to a specific context of use: terms linked to age groups (e.g. youth language), sex (men/women) or specific text types (e.g. literary language or technical language);
- chronological connotation in the case of neologisms and archaisms: words that are in the process of entering usage and those that are disappearing from use easily provoke emotional reactions (words are perceived as being ‘nice’/‘ugly’, ‘good’/‘bad’); this shows, moreover, that speakers are aware of language that is in the process of changing. To take the example of borrowings in French, a long-standing Anglicism like *film* (1889) or the abbreviation *radar* (1943) go unnoticed, whereas terms such as *blush* (1969) or the adjective *short* (meaning “unsatisfactory”) are marked both as borrowings and as neologisms.

The variational connotation of a lexeme adds to its semantic connotations and is thus linked to its lexical meaning (the denotation of the word). In other words, a given lexical meaning can be linked to one or several specific communicative contexts. Hence, variational connotation belongs to the domain of lexical semantics (*cf.* 9.1 above and 9.2.2).

Variational connotation should not be confused with another aspect that also has semantic implications: the associations evoked by a word in the mind of speakers. This is a mental and extralinguistic category with indistinct limits. Associations evoked by a lexeme may be individual, but they may also take on an (anthropological or cultural) intersubjective character. The exact nature of the relationship between lexemes and associations will be specified in the following section.

A text containing a large number of lexemes that carry a connotation (such as regionalisms, familiar forms, technical terms or neologisms) acquires strong connotations as a whole. It is not easy to discern connotations in texts from older historical periods, since our knowledge is based on texts displaying a high level of elaboration, be they literary works, scientific and technical texts or legal and administrative documents. Nonetheless, it is often possible to distinguish a number of popular or regional words in medieval texts, or to recognise legal, medical, mathematical or architectural terms in treatises on these subjects. Latinisms that were introduced into texts as words of high prestige constitute an easily recognisable category of diastratically marked lexemes.

Even though (semantic, derivational or borrowed) archaisms and neologisms can be identified to some extent in medieval texts (*cf.* Carles, *L'innovation lexicale chez Chrétien de Troyes*, 2013), they become easier to detect in sources from the 16th century onwards, owing to an increased number of texts and the appearance of numerous dictionaries.

9.2 Semantic and semiotic theory

9.2.1 Lexical meaning, concept and referent

Semantics deals with a broad range of topics concerned with meaning – from the meaning of lexemes to the meaning of texts or speech acts, as well as the meaning expressed by grammatical relationships. Lexical semantics is concerned more specifically with lexemes, but it also allows a better explanation of the general characteristics of semantics than that provided by the other subdomains of this field.

The most fundamental distinction in lexical semantics concerns that between *lexical meanings*, which are linguistic in nature and tied to a specific language, and *concepts*, which are cognitive in nature and lie outside or at least close to the boundaries of language. Theoretically, the distinction between the two is relatively simple:

- the concepts 'TREE' and 'BEAUTY' do not change according to our mother tongue; at most, they may be conditioned culturally: a French-speaking individual from the Lorraine region and a German-speaking individual from neighbouring Saarland will have a rather similar idea of a tree (e.g. a beech tree), since they are surrounded by the same type of vegetation; in contrast, French-speaking inhabitants of Nice or Martinique will likely have a rather different idea (a maritime pine or a palm, for example), despite the fact that they use the same word as their fellow speakers from Lorraine.
- the lexical meaning (or the 'signified') is necessarily associated with a given lexical form and is thereby placed within a semantic network, where it is at the same time associated with synonyms, antonyms, hypernyms and (co-)hyponyms (*arbre*, *arbuste* "bush, shrub", *plante*, *chêne* "oak tree", *bois* "wood" etc.). Moreover, unlike the concept, the lexeme as a combination of lexical meaning and form is further bound to diastematic markers (e. g. familiar, standard, scientific) and to a specific word class (e. g. noun, verb or adjective), resulting in the addition of semantic and syntactic information to the purely 'denotative' meaning.

There are two ways of identifying and defining a lexical meaning. The first is introspective; one can reflect on the possible meaning of a lexical form by drawing on one's intuitive knowledge and thinking of the possible contexts in which that form may occur. This introspective method is fast, but it remains subjective. The second is based on the use of the large corpora that exist for many languages today, for the Romance languages as

well as for English, including their medieval stages; these allow the meaning attributed to a form in a given context to be identified; the comparison of a large number of such contextual meanings, in turn, allows a general definition to be established. This method is more accurate, but much more challenging.

It is more difficult to identify concepts than to define lexical meanings, as the latter can be approached by direct observation. Concepts, on the contrary, can only be distinguished by the observation of lexical meanings in individual languages, followed by the comparison of the results obtained for each language. This process can to a certain extent be accompanied by psycholinguistic tests.

The distinction between concepts and lexical meanings is well established in studies on literary and cultural history, and it forms the basis of questions such as the following:

- which words (= lexical forms and meanings) can be associated with the concept of 'LANGUAGE' in French? (Answer: *langue, idiome, langage, variété*, etc.)
- which terms are used to evoke 'FEELINGS' in French songs of the 12th century or in Italian literature of the 14th century?

The functions of lexical meaning and concepts are nonetheless very similar: they assemble information concerning a defined entity (an object such as a tree or an idea such as beauty), making it possible to refer to these entities efficiently in utterances. At the level of the brain, the signified is not actually distinct from the concept; rather, it represents a slightly more precise and sharply delineated form of the concept that only arises at the moment at which it is combined with the neural connections linked to a particular lexical form (at the same time, the previously mentioned semantic network of which it forms part is evoked, as well as the diasystematic and morphological information linked to the connection between form and meaning).

Interpreting the triad formed by signified, concept and (concrete or abstract) referent is not easy. In the following, the most important stages of the attempts at such an interpretation by 20th-century linguistics are presented, concluding with a synthesis reflecting the current state of research.

9.2.2 Sign theory and the structure of the lexeme

All semantic theories are chiefly concerned with the nature of the linguistic sign from a 'semiotic' point of view (cf. 1.2.2). The sign is primarily a sequence of sounds or letters that are segmented and interpreted as phonemes and graphemes, words and sentences, in order to refer ultimately to 'extralinguistic reality'. The majority of the proponents of semantic theories have represented their philosophy in the form of diagrams, beginning with Saussure's dichotomy of the sign (no. 1 below), which is followed by a semiotic triangle (no. 2), a semiotic trapezium (no. 3) and a semiotic pentagon (nos. 4 and 6), with

a version of the pentagon having been adapted – albeit unconvincingly – in the shape of a semiotic square (no. 5).

1 The dichotomy of the sign

The fundamental ‘dichotomy of the sign’ (Saussure 1916), which is still valid today, postulates the existence of a signifier (form) and a signified (meaning), which, together, form a unity at the moment of utterance (*cf.* 1.2.2). The relationship between signifier and signified is specific to each individual language, and is conventional and arbitrary in synchrony (*cf. ibid.*). This explains why the same concept may have different names in different languages. While Saussure’s dichotomy does not explicitly account for the relationship between the lexical meaning (the signified) and the concept, it paves the way for its description.

2 The semiotic triangle

The semiotic triangle (Ogden / Richards 1923) was inspired by the philosophy of Aristotle. It elaborates on the dichotomy established by Saussure by introducing a third element: the referent in extralinguistic reality to which the linguistic sign refers:

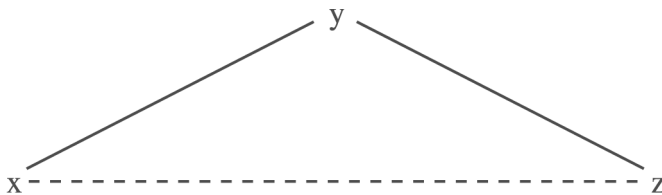


Fig. 27: The semiotic triangle (source: Ogden / Richards, *The Meaning of Meaning*, 1923).

- x: phonetic sequence (replaces Saussure’s ‘signifier’, corresponds to Hjelmslev’s ‘expression’ [1943])
- y: mental representation (replaces Saussure’s ‘signified’, corresponds to Hjelmslev’s ‘content’)
- z: external reality (‘referent’)

Like Saussure’s model, the semiotic triangle groups the lexical meaning and the concept into a single entity, with the crucial difference that Saussure focused on lexical meaning, whereas Ogden and Richards emphasised the concept. In contrast to Saussure’s model, the semiotic triangle does not consider the phonological signifier (the mental representation of *langue*), but rather the phonetic sequence (which is an element of *parole*).

3 The semiotic trapezium

The semiotic trapezium (Heger 1976 [1971]) offers the first complex representation of the linguistic sign. It differentiates between phonetic realisation (the phonetic sequence, belonging to *parole*) and phonemic representation (the phonological signifier, at the level of *langue*) and introduces an explicit distinction between the linguistic signified (the lexical meaning) and the designated (the extralinguistic concept, which Heger refers to as the 'noeme'). Furthermore, Heger applies his model not only to isolated words, but also to larger linguistic units such as entire texts.

The semantic trapezium model will not be presented in detail owing to its complexity, as well as to the fact that its components are more precisely defined in the semiotic pentagon model. Similarly, Coseriu's semiotic theory will not be described here. Although his structuralist ideas were of great importance, they nevertheless failed to clarify the distinction between meaning and concept, which would have to wait for the subsequent contributions of cognitivism.

4 The semiotic pentagon

Current semiotic theory is based on the semiotic pentagon (Raible 1983), which, like Heger's trapezium, relies on the distinction between lexical meaning (the 'signified') and concept (which Raible refers to as 'designated'). This model accounts for the dichotomy of the linguistic sign ('signified' = meaning / 'signifier' = phonological image) as well as for the relationship of the sign to its actual phonetic sequence (referred to as 'nomen', a term that did not become widely established). The pentagon model thus provides a synthesis of all the constitutive elements present in previous models.

In addition, the pentagon model specifies the nature of the 'referent', which is no longer considered as a category relating to the external world, but rather as a category of utterance. Realities outside our perception and faculties of conceptualisation are not directly involved in the process of reference. The referent is thus considered to be 'actual' (i.e. what the utterance is talking about); this refinement avoids, among other things, the necessity of distinguishing between concrete reality ('TREE') and abstract reality ('BEAUTY'), by placing both on the same level. Although we may intuitively perceive a major difference between entities of concrete and abstract reality (and although their linguistic representations are handled differently by the brain), this difference disappears at the level of utterance.

The semiotic pentagon consequently incorporates the following elements:

'the virtual' – corresponds to the brain's representation (of language)

- 'signifier' = lexical form (phonological image, corresponding to the formal half of the sign)
- 'signified' = lexical meaning, attached to the signifier in a given language

- ‘designation’ or ‘concept’ = the mental representation of the entity in question (a category that is cognitive, universal or linked to a defined culture, and not to an individual language)

‘the actual’ – corresponds to the concrete utterance (or its context)

- ‘nomen’ or ‘phonic / graphic form’ = sequence of sounds by means of which the signifier is actualised (category of *parole*)
- ‘referent’ (or ‘nominandum / denotatum’) = actual referent of the sign within the utterance context

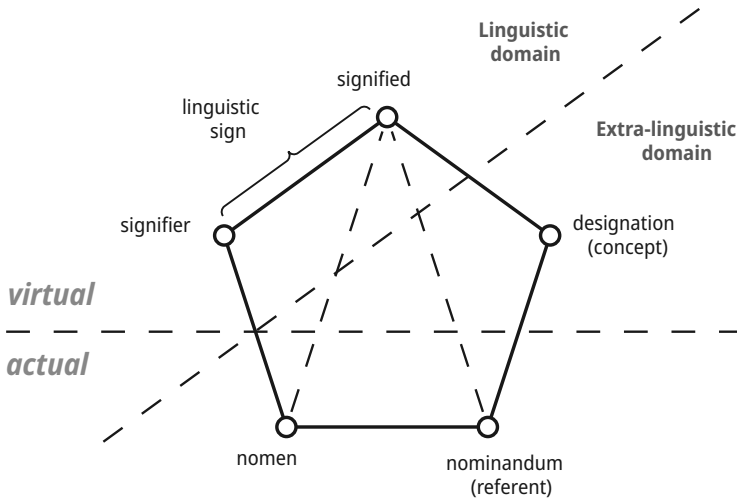


Fig. 28: The semiotic pentagon (source: Raible, *Zur Semantik des Französischen*, 1983)

5 The semiotic square

The semiotic square (Blank 1997) starts from Raible's pentagon, but restructures it and adds three essential complementary parameters (cf. 9.2.1 above), which are part of the linguistic sign beyond its phonological form and lexical meaning:

- semantic relationships: polysemy, semantic fields, frames (cf. 9.3)
- the syntagmatic context: valency, collocations and fixed expressions (cf. 9.5)
- the diasystematic context: variational connotations (cf. 9.1.4)

Using this as a basis, Blank developed his *Drei-Ebenen-Semantik* (= three-level theory of semantics, Blank 1999, ch. 7), in which the diasystematic parameters are considered to be as much part of the lexical sign as form and meaning.

Aside from these useful additions, the semiotic square has several weaknesses: it wrongly considers the concept and the extra-linguistic ‘encyclopedic knowledge’ as

synonymous. It also introduces too great a separation between the signified and the concept. Moreover, it mistakenly replaces the correctly conceived opposition between ‘virtual’ and ‘actual’ with that between ‘abstract’ and ‘concrete’, which is more banal and ultimately incorrect.

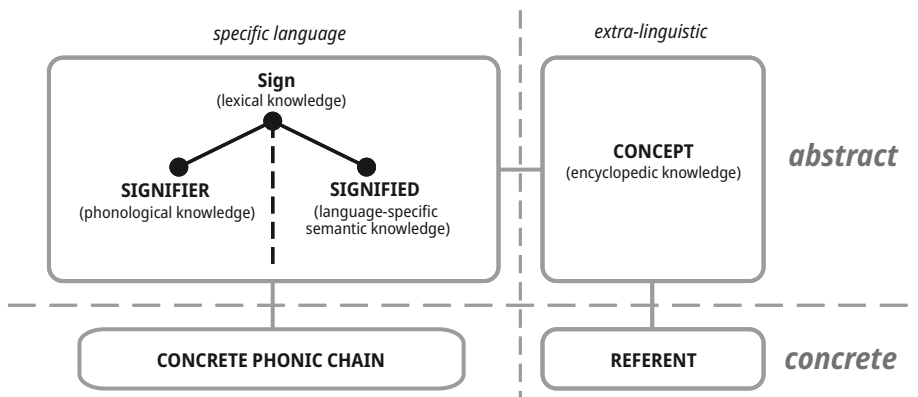


Fig. 29: The semiotic square (translated after Blank [*Prinzipien des lexikalischen Bedeutungswandels*, 1997: 148]; *Einführung in die lexikalische Semantik*, 2001: 9)

6 The enhanced semiotic pentagon

After a failed attempt to develop a modified version of the semiotic square (*cf.* Lebsanft / Glessgen, *Historische Semantik in den romanischen Sprachen*, 2004: 14–17), we returned to the more accurate pentagon model and elaborated the extended version below (first published in Glessgen, *Le statut épistémologique du lexème*, 2011: 453 and further developed with the aid of Peter Koch).

This model accounts for the organisation of lexical information in the brain. The brain possesses two distinct physiological systems, one for lexical forms and the other for concepts; together, these two systems constitute ‘semantic’ memory, which is part of ‘declarative’ memory (*cf.* 5.1.2). The manner in which lexical forms are stored in the brain is complex, in particular due to inflectional variation. We will limit ourselves here to the following remarks:

- memory concerns both phonological forms and orthographic forms;
- the storage of a given form (be it simple or compound) depends on its frequency of use;
- for persons who have knowledge of more than one language, the series of forms are necessarily organised according to each individual language.

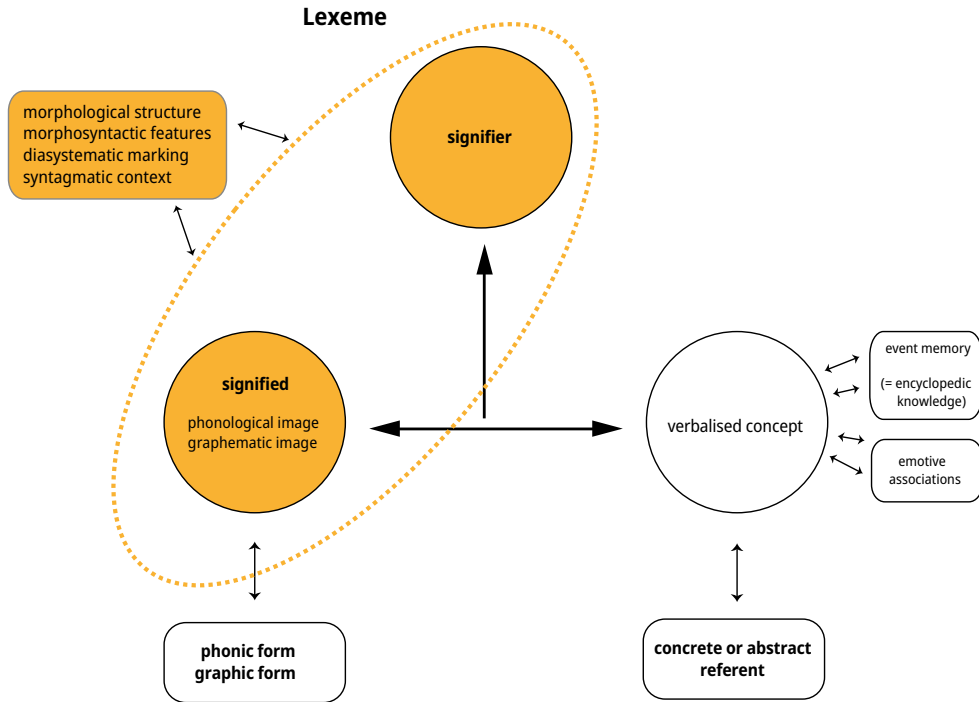


Fig. 30: The enhanced semiotic pentagon (Glossgen)

The memorisation of concepts is even more complex:

- it is linked to perception, vision in particular (e.g. the shape of objects and movements), and to the functionality of the referents in question;
- a single concept may be composed of different elements, which are interconnected by neural links (e.g. the concept of a dog evokes the fact of having four legs, barking or running in a certain way);
- the memorisation of a concept always occurs in interaction with that of other concepts; individual concepts are more easily retained due to their integration within conceptual networks;
- due to these various factors, concepts are by nature imprecise and vague;
- semantic memory is clearly detached from ‘episodic’ memory, which assembles recollections of concrete events from our past, such as that of a given place or experience (e.g. ‘my parents had two dogs; one of them once bit the postman’); concepts are thus abstract entities that are close to *langue*, and not truly ‘encyclopedic’ entities as is generally assumed);

- non-activated concepts end up disappearing (the brain functions according to the rule ‘use it or lose it’);
- a final and essential point is that multilingual persons have a single series of concepts, which is linked to different groups of lexical forms.

The great paradox inherent in this model is that the signified, which is the linchpin around which the processing of semantic information in language revolves, does not exist in itself at the level of the brain. To be more precise, it is an extract from the conceptual memory that relies on the links that are formed between different lexical forms and the conceptual network. In this way, the French forms *bois* and *forêt*, Italian *legno* and *bosco*, Spanish *madera* and *bosque* or English *wood* and *forest* are each linked to the same two conceptual groups (notwithstanding the fact that Fr. *bois* and English *wood* are polysemic and can also denote a type of “forest”).

The structure of the enhanced semiotic pentagon can be specified on this basis:

- it contains five basic entities (phonological form, lexical form, signified, concept and referent), two of which constitute the lexeme; it thus forms a pentagon like Raible’s model;
- there are neural links between lexical forms and corresponding (‘verbalised’) concepts; this relationship is represented by a two-way arrow;
- the signified is generated through these neural links; it is inseparable from the lexical form (the ‘signifier’) and draws from the concept; this is indicated by two arrows;
- complementary elements (concerning parts of speech, morphological structure, diasystematic markers and the syntagmatic context) are linked to the lexeme; encyclopedic elements (event memory) as well as emotional associations are linked to the concept.

The model explains, among other things, how a multilingual individual is able to control a single collection of concepts via different groups of lexical forms. It emphasises that the nature of the signified and that of the concept are fundamentally identical and that, contrary to what is generally supposed, concepts are not an ‘extra-linguistic’ entity. There are, of course, concepts for which there is no corresponding lexical form in a given language (‘non-verbalised’ concepts, cf. no. 7 (2) below), but the processing of the conceptual network largely and necessarily relies on lexical forms.

7 Signifieds, concepts and referents

The relationship between meaning (‘signifieds’), concepts and referents is at the very core of the functions of language since one of the main reasons we speak is to articulate

concepts. The relationship between these three entities thus requires further reflection. The following observations can be made:

1. The emergence of a sign starts with concrete and abstract referents. The two types of referents are different, but less so than might at first appear:

- Concrete referents are only defined as entities by humankind. They exist physically in reality before being defined, but it is only through our discernment that they are transformed into referents; e.g. a “volcano” only becomes a discrete entity because we consider it as such; before that, it is a part of the continuous surface of the Earth. Crucially, the existence of a concrete referent thus depends on human cognition.
- Abstract referents more immediately reflect ‘choices’ made by humankind. However, they also relate to concrete information: the concept of ‘BEAUTY’ is associated with physical forms; e.g. men, women, animals, landscapes or buildings may be considered to be beautiful.

In order for a given element of information to become an ‘actualised’ referent, it must be conceptualised. There are an infinite number of potential referents that have never undergone this process. For instance, the sensations one experiences when (1) letting oneself glide into a bathtub full of warm water or (2) diving into the sea or (3) plunging into a swimming pool are caused by three different types of contact with water that could theoretically give rise to conceptualisation.

2. The emergence of a concept presupposes a particular interest in a potential referent. The concept thus results from the sum of the knowledge one has of the referent (‘encyclopedia knowledge’). This also involves a significant amount of abstraction: we select the features that appear to us as most characteristic of the referent (its ‘salient features’) and we ‘categorise’ them.

The process may be interrupted at this stage without the concept having been associated with a given lexical form (= without having been ‘verbalised’). There are numerous concepts that do not have a corresponding fixed linguistic expression in a given language; this is easily proven by means of the large quantity of concepts for which there are specific terms in one language, but not in another (think of the French distinction between *espoir* and *espérance*, for which there is no equivalent in English, nor in many other languages, while *shallow* is frequently cited as an English word for which no single word can be found in other languages).

It can be presumed that there are a large number of other concepts for which no fixed form exists in a given language. In order to express such a concept, it is nevertheless possible to describe or paraphrase it. Put simply, if people have a strong interest in a concept, they will attempt to find a specific term for it, thereby verbalising it, allowing it to be referred to efficiently and unambiguously.

3. In order to verbalise a new concept, one inevitably draws from the previously existing inventory of lexical forms in a language: new meanings can be attributed to forms already in use (*cf.* 9.3.1); these forms can at the same time be altered (by means of derivation, compounding and phraseologisation, *cf.* 2.4.4 nos 2 and 3 as well as 2.4.5); furthermore, terms can be borrowed from foreign languages (*cf.* 9.6.3). In all these cases, the lexical form attributed to a new concept relies on the forms attributed to previously verbalised concepts; this, in turn, shows that while signs are not semantically motivated in synchrony, they must have been at some point during their evolution.

When a concept is verbalised, it is transformed and specified, and it becomes more widely used: the categorisation of a concept is completed through its verbalisation. The limits of a 'signified' are more distinct than those of the underlying concept, considering that a signified groups different conceptual elements, or even different concepts (e.g. the English lexeme *hair* groups together the conceptual elements that are expressed by separate lexemes in French: *cheveu* "head hair" and *poil* "body hair"). In reality, there are more concepts than signifieds (and, through polysemy, more signifieds than lexical forms, *cf.* (8) below).

In addition to the difference in nature, the ways in which concepts and signifieds can be observed differ: while a lexical meaning can be grasped through the analysis of the use of words in different linguistic contexts, the limits of a concept are very difficult to define. As mentioned above (*cf.* 9.2.1), it is necessary to start from the lexical meanings in different languages and to discern the concepts that can be identified by means of comparison.

4. It should be remembered that, by definition, the signified reflects the knowledge shared by a community of speakers. It belongs to a specific language and is part of the network of semantic, syntagmatic and diasystematic relationships in that language (as illustrated by the 'three levels' of Blank's theory of semantics).

5. During the process of communication, the hearer perceives a phonetic form and interprets it as its corresponding phonological form: the latter evokes a signified, the semantic content of which corresponds to a concept; the concept is associated with a referent and at the same time activates the event memory (encyclopedic knowledge) and any emotional associations that are linked to the concept.

In the act of utterance, the lexeme evokes a given referent, this time with a noticeable difference between concrete and abstract entities (contrary to the way in which concepts and forms are stored in the brain): a concrete referent may be unique and observable in the (external) world ('the tree [that I see outside in front of my window]'), whereas a mental referent always remains generic ('the 'BEAUTY' of the tree in front of my window').

6. The expressive force of language relies, in the end, *ex negativo*, on the choices that are made among innumerable concepts and potential referents that are not verbalised. A lexical form that has been forged for a concept in a given language bears witness to the particular interest that people have for this concept within a specific historical and cultural context.

7. The choice of an existing lexical form for the verbalisation of a new concept reflects recurrent and shared associations at the time of verbalisation: the new concept is associated with concepts that are linked to the form in question (*cf.* the example of the computer *mouse* in 1.2.2 and 9.3.4). Technically speaking, verbalisation thus corresponds to a semantic change, which, at the same time, adds a further meaning to the lexeme; it is therefore the origin of polysemy.

8. Polysemy is the synchronic equivalent of semantic change over time: the two make use of the same mechanisms. Polysemy may be described in a synchronic perspective, by the use of grids of salient and distinctive features, or, in a diachronic perspective, starting out from the semantic change at its origin. This second method, then, reflects the cognitive operations on which polysemy is based.

Polysemy is one of the basic elements of lexical organisation: the brain is more easily able to handle a limited number of forms that each have several meanings linked by cognitive mechanisms than a larger number of forms, each with a single meaning.

‘Natural’ languages are distinct from scientific terminologies in that the latter use unequivocal terms (one form = one meaning) and develop a large number of specialised terms.

Note that the distinction between polysemy (“a word with more than one meaning”) and homonymy (“more than one word that has the same form”) relies on genetic differences, according to the definition upheld here: polysemic forms reflect a conceptual association that was present when the signifieds in question emerged, whereas homonyms are identical words of different origins; they simply have the same form at a specific time, owing to the vagaries of phonetic change.

In practice, many linguists make a purely synchronic distinction between polysemy and homonymy. For polysemic words, they suppose the existence of a semantic link that can be identified in synchrony, whereas for homonymic words, they assume that there is no such identifiable link.

9.2.3 Onomasiology, structural semantics and cognitive semantics

The importance of semantic networks for language processing was understood as early as the end of the 19th century. The first attempts at systematically grouping lexemes according to their semantic proximity were made by Tappolet in 1895 (names of family

relations in the *România*) or Zauner in 1902 (names of body parts). The first half of the 20th century saw the development of the discipline of ‘onomasiology’, which preceded structural and cognitive semantics. Onomasiology studies the different forms used to name a given concept or those used to name several neighbouring concepts (cf. 4.2.1 and the bibliography by Quadri 1952).

Following these early studies, there were several attempts to create global onomasiological systems integrating all concepts verbalised in a given language (Casares, *Diccionario ideológico*, 1942) or in several languages. With regard to Romance studies, note in particular Hallig and Wartburg’s *Begriffssystem*, 1963 [1952], which was developed within the framework of the FEW (cf. 9.9.1. no. 3). It remains a valid alternative to the *Historical Thesaurus of the OED* (HTOED, 2020–), which is primarily based on the *Oxford English Dictionary* and which represents the most detailed onomasiological system to date. The internal and external contradictions displayed by these differing versions illustrate the difficulties inherent in structuring the semantic network of a language. In particular, the idea of a general framework applicable to more than one language inevitably encounters difficulties due to the close links between concepts and signifieds.

Despite these issues, the usefulness of partial semantic groupings for different linguistic ends, but also for literary and historiographical aims, is no less significant. The initial onomasiological current dissipated in the 1950s, but it was revived at the end of the 20th century within the context of cognitive studies.

Theories of structural semantics began to appear in the 1930s (cf. 2.2.4 no. 3) and developed significantly from the 1960s onwards. They are based on onomasiology in the sense that they also group words according to semantic fields; however, they focus on describing signifieds rather than concepts. They break down the meaning of words into smaller units (semantic features, also known as ‘semes’) and situate lexemes within a system of functional semantic relationships (the most important of which are synonymy and antonymy, hyponymy and hypernymy); together, these relationships constitute the semantic network of a language.

A typical example of a structural semantic analysis that breaks down word meaning into semantic features is the analysis of the semantic field *sièges* (“seats”) by Pottier (*Recherches sur l’analyse sémantique*, 1963), according to which the combination of different basic semantic units (semes) produces a superior unit of defined lexical meaning (the sememe). For example, the signifieds *chaise* (“chair”), *fauteuil* (“armchair”) and *canapé* (“sofa”) refer to items which have a backrest, whereas *tabouret* (“stool”) and *pouffe* (“pouffe, footstool”) do not. Every lexical or semantic field recognises hypernyms (also called ‘archilexemes’), which encompass different hyponyms; e.g. the hypernym *flower* includes the hyponyms *rose*, *tulip*, *lily*, etc., which, in turn, exist in a relationship of co-hyponymy.

Structural semantics encountered the same problems associated with the universal linguistic level that traditional onomasiology did: semes, lexical fields and archilexemes are not universal, but differ from language to language, even though languages often

share common features. Furthermore, the semantic network of languages is not systematic and does not cover the entire range of possible meanings in a complete and continuous manner.

Cognitive theories are also related to the long-standing tradition of onomasiology; they propose that lexical meanings immediately reflect the concepts with which they are associated. In particular, cognitive semantics was responsible for the introduction of the model of semantic prototypes (*cf.* Kleiber, *La sémantique du prototype*, 1990, as well as various studies by Koch). The model assumes that concepts are organised in groups within the brain according to semantic proximity (or ‘contiguity’); each semantic group is centred around a nucleus constituted by the ‘best example’ in our imagination. According to this logic, a hypernym such as *bird* is *not* the central element of the semantic field to which *swans* and *cormorants* belong, but rather the *sparrow* in Europe, the *robin* in North America (or, in large cities, the *pigeon*), each of which represents the prototype of this category (*cf.* 9.3.2).

Cognitive theory places more importance on lexemes that reflect a basic semantic level; hypernyms, in contrast, which represent a hierarchically superior level, are considered to be secondary from both a cognitive and historical point of view (the most salient individual birds are named first, and only afterwards do we name the generic category; *cf.* Geeraerts, *Theories of lexical semantics*, 2004). The basic semantic level more accurately reflects the world of everyday references that contribute to structuring our cognitive universe (*cf.* 9.3.7 no. 2).

The different semantic theories refer to relational categories such as synonymy (which does not actually exist, given that each word has its own particular usage) and antonymy, as well as to fundamental distinctive features of the animacy hierarchy: human / non-human and animate / inanimate [countable / uncountable], concrete / abstract, *cf.* 8.4.3 no. 2 and 9.3.3). Each of these theories has contributed important insights to the understanding of lexical and conceptual meaning and none of them can be ignored today.

- Detges, *Cognitive semantics in the Romance languages*, OxfEnc
- Casas Gómez/Hummel, *Structural semantics in the Romance languages*, OxfEnc
- Winter-Froemel, *Onomasiology in the Romance languages*, OxfEnc

9.3 Semantic change and semantic relationships in the Romance languages

9.3.1 Types of semantic change and semantic relationships

Of all types of language change, lexical change is the most dynamic. Given the vast quantity of forms and lexical meanings in a language, linguistic variation proper continually leads to transformations in everyday lexis. Semantic change takes place in parallel with word formation (*cf.* 2.4.4) and borrowing (*cf.* 2.4.6). These three mechanisms fulfil the same functions, namely (1) accounting for changes in the world of referents and the sphere of cultural concepts and (2) implementing the expressive capacity inherent in lexis.

A thorough understanding of the mechanisms of semantic change is essential as they are highly productive and are thus omnipresent in language use. Knowledge of semantic change allows one to better discern the semantic relationships that govern the brain's system of language processing with regard to different signifieds (or concepts). Semantic change relies on principles of 'taxonomy' (hypernyms vs. hyponyms) and 'metonymy' (the proximity of concepts to one another in our everyday life), which are precisely the same principles as those underlying the organisation of concepts in the brain. A semantic change occurs when a new concept is attributed to an already existing form (possibly accompanied by the addition of an affix to the latter); the semantic relationship that is established between the old and the new concept, linked to the same lexical form, implies a cognitive association between the two. Semantic change thus points to the existence of preferential links between certain concepts at the level of the brain.

When a semantic change takes place, the relationship between the initial meaning and the new meaning is always perceptible and transparent to speakers, even though they rarely pay attention to this phenomenon. This rule is consistent with what we have already observed in the case of phonetic changes (*cf.* 6.1.2; 6.2.4) and grammatical changes (*cf.* 5.1.3, 'grammaticalisation') and the same is true for derivational changes (*cf.* 9.4.6). When a change first occurs, speakers can normally establish a logical association between the initial form and the innovative form. It is only later that awareness of this relationship is lost.

The usefulness of the notion of semantic change for the understanding of conceptual networks, even in synchrony, explains the development of historical semantics. It has gained particular importance in Romance linguistics owing to the volume and the quality of the historical documentation available for these languages.

The first attempts at classification relied on rhetorical and psychological interpretation (*cf.* in particular Bréal, *Essai de sémantique*, 1897; Roudet, *Sur la classification psychologique des changements sémantiques*, 1921; Ullmann, *Précis de sémantique française*,

1952). These attempts were enhanced by structuralism and, more recently, by cognitive linguistics (cf. the syntheses by Blank, 1997 and 2001). The overview presented below in the form of a diagram is based on the whole of this tradition and develops it further (cf. once again Glessgen 2011: 430–437).

According to Roudet's classification, all semantic change is based either on contiguity (proximity) or similarity between the initial lexeme and the new lexeme. These two types of relationships can apply to lexical meanings as well as lexical forms, as shown in the following table:

	contiguity	similarity
<i>content / concept</i>	metonymy <i>nez</i> “nose (organ)” → “sense of smell”	metaphor <i>nef</i> “ship” → “nave (of a church)”
<i>form / expression</i>	ellipsis <i>voiture automobile</i> “car” → <i>voiture</i> (or: <i>automobile</i>) “id.”	popular (folk) etymology <i>forain</i> “stranger” → “relating to the fair” (influenced by <i>foire</i> “fair”)

Remarks:

- contiguity between two meanings results in metonymy (consider the further example *boire un verre* “to drink a glass” for “to drink the contents of a glass”;
- similarity between two meanings leads to metaphor (e.g. Fr. *trompe* “trumpet” → “trunk (of an elephant)”). Such assimilations are situated at the level of concepts and then manifest themselves at the level of signifieds;
- contiguity between two forms is at the origin of ellipsis (e.g. Fr. *diligence* is used to mean *carrosse de diligence* “stage-coach” and *cabinet* can refer to a *cabinet d’avocat* “lawyer’s office, legal firm”, *cabinet de toilette* “bathroom”, etc.;
- similarity between forms may give rise to semantic popular (or folk) etymology (e.g. the term *miniature* has taken on the meaning of “type of illustration of small dimensions” by association with the word family of MINUS, even though, etymologically speaking, it is derived from Lat. *minium* “lead tetroxide”, a red-coloured pigment used to illuminate manuscripts; the initial meaning of *miniature* was thus “illustration coloured with minium”; cf. 11.2.5).

To these four types of semantic change, taxonomic change (hypernym → hyponym) should be added: e.g. Late Lat. *auca* “bird” → Fr. *oie*, It. *oca* “goose”. This type of change takes place within a semantic family and reflects partial identity (a goose is a kind of bird), rather than a form of contiguity.

The following table summarises the fundamental characteristics of the three categories of conceptual change – taxonomy, metonymy and metaphor (cf. Steiner, *I centri di espansione nel cambio semantico*, 2016). They will be presented in detail in the follow-

ing sections, together with changes that have a formal motivation (ellipsis and popular etymology).

partial identity	contiguity	similarity (contrast)
<i>taxonomy</i> (= within the same semantic family)	<i>metonymy</i> (= within the same 'frame' / 'domain' / 'script')	<i>metaphor</i> (= 'domain mapping' = conceptual bridge between two frames or two remote taxonomies)
Vertical changes [generalisation] [specialisation]	[part-whole] [container-content]	[antiphrasis] [auto-antonymy]
Horizontal changes [co-hyponymy]		
close cognitive association		weak cognitive association

9.3.2 Taxonomic change

Taxonomic change is the least pronounced type of semantic change, since it relies on partial identity. It occurs within a semantic family (or semantic field, e.g. INSECTS) and affects the structural categories of hyponyms (*fly*, *bee*, etc.) and hypernyms (*insect*). There are three types of taxonomic semantic changes: the use of a hyponym to refer to a hypernym (generalisation), the opposite process (specialisation) and the assimilation of two hyponyms belonging to the same family (co-hyponymy).

1 Generalisation (*species* → *genus*)

Generalisation occurs when the lexical form of a hyponym is transferred and also applied to its corresponding hypernym: e.g. the Latin word PANARIUM “bread basket” evolved into Fr. *panier* “basket (in general)”; Lat. PASSER “sparrow” survives in Spanish as *pájaro* “bird”.

Generalisation reduces the number of specific conceptual features linked to a form while increasing its number of potential referents (i.e. the form *pájaro* now includes all types of birds, rather than sparrows alone). Generalisation may also be referred to as ‘broadening of meaning’; this term is, however, problematic because the broadening applies not to the meaning but to the referents in question.

Generalisation is not very frequent in the history of the Romance languages, but it is nonetheless a recurrent mechanism, which can be further illustrated by the example of the terms for “uncle” and “aunt” in Latin and the Romance languages:

- Latin has four distinct terms: PATRUS “paternal uncle”, AVUNCULUS “maternal uncle”, AMITA “paternal aunt” and MATER “maternal aunt”;

- in the Romance languages, only one term each for “aunt” and “uncle” survive; specifically, those used to designate the maternal uncle (Fr. Occ. *oncle*, Rom.n *unchi*) and the paternal aunt (O.Fr. *ante*, Fr. *tante*, Engad. *amda*, Rom.n *mătușă* [with the (probably borrowed) suffix -ușă]); hence, a generalisation has taken place in both cases;
- the motivations behind generalisation can be both semantic (if the contrast between the meanings of the terms in question is weak) and formal: the types PATRUU and MATERTA were disadvantaged by their marked phonetic proximity to PATRE and MATRE (in Fr., for example, PATRUU and PATRE would have led to the same result, following hereditary phonetic evolution).

2 Specialisation (*genus* → *species*)

In contrast to generalisation, specialisation (or ‘restriction of meaning’) increases the number of specific conceptual features and reduces the number of potential referents (as in the example of Late Latin *auca* → Fr. *oie*).

Like generalisation, specialisation reflects the organisation of semantic fields around prototypical referents: the meaning of Fr. *blé* (< BLATUM) evolved from “cereals (in general)” in Latin to “wheat” in French by a process of specialisation based on the principle of the ‘best example’ (the most frequent, the most characteristic or the more valued representative). In the case of *auca*, one possible interpretation is that in a rural context, geese were dominant amongst the poultry kept in the backyard. Further examples of specialisation are:

Lat. VIVENDA “that which is necessary for life” → Sp. *vivanda* “food”, O.Fr. *viande* “id.” → Mod.Fr. *viande* “meat” (meat was not the most common food, but the most valued).

Lat. COMPARARE “obtain” → various Romance languages “buy” (Engadinian *cumprar*, It. *comprare*, Cat. Occ. Pg. Sp. *comprar*, Logudorese [= Sard.] *kompurare*, Rom.n *a cumpăra*). Buying – as opposed to stealing, for example – was the most usual means of obtaining something.

Lat. HOMO “man (human being)” → Fr. *homme*, It. *uomo*, Sp. *hombre* “man (more specifically a male person)”. This change presupposes a patriarchal society, a supposition which is confirmed by the later grammaticalisation of O.Fr. *om* as the Mod. Fr. pronoun *on* “one”, which finds an equivalent in the German synonym *man* “one” < *Mann* “man”. It should be noted that Romanian preserves the original Latin meaning (*Maria este un om inteligent* “Maria is an intelligent person”), which may be indicative of semantic specialisation between the end of the 3rd century (after Roman colonisation in the Danube region) and the end of the 5th century.

3 Co-hyponymy

A third, similar type of semantic change concerns relationships of co-hyponymy: two concepts that represent hyponyms of a same hypernym are associated with one another. This change has not always been recognised as a type of taxonomic change parallel to the restriction and broadening of meaning. It is particularly frequent in the domains of flora and fauna and rarer for artefacts; *cf.* the following examples:

in the context of rural animals: Lat. TALPA “mole” → It. *topo* “mouse”

Sp. *tigre* “tiger” → Am. Sp. “jaguar”, Sp. *león* “lion” → Am. Sp. “puma”

Fr. *chapelle* → La Réunion creole *sapèl* “Indian temple”

at an abstract level of meaning: Lat. SUPERFLUUS “exorbitant, excessive” → Fr. *superflu* “superfluous”

Like other taxonomic changes, co-hyponymic relationships are based on ‘folk categories’ rather than scientific categories. They involve elements of perception (shape, colour), functionality or cultural knowledge (thus, in this context a ‘WHALE’ would be categorised as a ‘FISH’ rather than a ‘MAMMAL’); *cf.* the following changes in the *oïl* dialects (*cf.* FEW s.v. SERPENS):

Fr. *serpent* → “(non-venomous) snake” [absence of a verbalised hypernym; the association is non-scientific]

Fr. *serpent* → “eel” [id.]

Fr. *serpent* → “dragon” [id. – the mythical animal is considered to exist in popular belief and belongs to the same ‘family’ as the snake]

9.3.3 Metonymy

Metonymy is the most varied and frequent type of semantic change. It relies on the principle of ‘contiguity’ (or proximity) between two concepts in terms of space, time, functionality or causality. The two concepts may be connected by their co-presence or succession within the same stereotypical situation or ‘frame’ (also referred to as ‘domain’ or ‘script’).

While taxonomic and, above all, co-hyponymic relationships predominate in adult cognition, for children, metonymic relationships are clearly more salient. Consequently, children aged six establish a closer association between ‘CARROT’ and ‘RABBIT’ or ‘BANANA’ and ‘MONKEY’ – i.e. concepts belonging to the same frame – than between ‘CARROT’ and ‘BANANA’ and ‘RABBIT’ and ‘MONKEY’, which are related by co-hyponymy; older children more easily perceive the link between fruits and vegetables or different types of animals.

1 Co-presence

A first type of metonymy groups pairs of concepts that are spatially or temporally co-present within the same frame. The following examples are taken from Blank 1997, especially p. 250 *sqq.*):

- within the frame ‘mourning’: Lat. *PLANGERE* “beat one’s chest in mourning” → Lat. “lament” → Romance “cry, weep” (It. *piangere*, Rom.n *a plânge*, Sard. *prangere*)
- within the frame ‘restaurant’: Fr. *nous avons quatre assiettes de plus à la table 15* “we have four additional plates at table fifteen”, referring to “four additional guests” (corresponding to the English use of the word *cover* in the same context)
- temporal contiguity: It. *vendemmia* “grape harvest” → “grape harvesting season” (action → time when action takes place)
- spatial contiguity: Late Lat. *coxa* “hip” → Fr. *cuisse* “thigh”; Lat. *BUCCA* “cheek” → Fr. *bouche* “mouth”, It. *bocca*; *CAPITIA* “hood” → Sp. *cabeza*, Pg. *cabeça* “head”
Id.: Fr. *bureau* “table at which one works, desk” → “office in which one’s desk is situated” (object involved in an action → location of object); Fr. *cabinet* “assembly hall for ministers” → “group of ministers”
- causal contiguity: Fr. *sceptre* “sceptre” → “sovereign authority who has a sceptre as a symbol (characteristic attribute of a person)”; O.Fr. *message* “message” → “messenger” (object involved in action → actor); O.Sp. *pregón*, O.Pg. *pregão* “messenger” → “message” (actor → object involved in action) and Fr. *courrier* “messenger, postman” → “mail, letter”; Lat. *TESTIMONIUM* “testimony, evidence” → O.Fr. *témoign* “witness” (action → actor); *cf.* also the example of *NASUS* / *nez* “nose (olfactory organ)” → “nose (sense of smell)” provided in the table above.

Co-presence, then, takes various factors into consideration: the actors, the activities in which they are involved, the means these require, the objects and the places within a given frame, as well as the characteristic aspects of the frame.

2 Part-whole and whole-part relationships

A particular type of co-presence is the *part-whole* relationship, that is, the association between an individual element and its entire frame; e.g. Fr. *l’électeur* “the voter” can also mean “the whole group of voters”. Further examples are:

- part (a characteristic) → whole: Fr. *nez* “nose” → “face”; Fr. *âme* “soul” → “man”; It. *tetto* “roof” → “house”
- whole → part: Sp. *quebrada* “ravine” → Am. Sp. “creek”

3 Succession

Succession groups highly diverse relationships that are determined by space, time or logical relationships (cause – effect). Succession can involve an activity, the results or cause of an activity, the persons it implicates or the time and place of its occurrence:

It. *spina* “thorn” → “acute pain” (object → effect produced by the object)

Fr. *addition* “operation consisting of adding together several units into a whole, addition” → “piece of paper presenting the total of the expenses incurred at a restaurant, bill” (action → effect brought about by action)

Fr. *sourcil* “eyebrow” → *froncer les sourcils* “knit one’s brows, frown; show a state of discontent” (means used for the activity → cause)

Lat. *VIGILIA* “day before a religious festival” → Fr. *veille*, It. *vigilia* “day preceding another” (succession of two intervals of time)

4 Auto-converse

A final type of metonymy is ‘auto-converse’ change, sometimes called ‘internal metonymy’; it is particularly apparent in the case of certain verbs that express two opposing semantic roles (agent and patient) as in Fr. *louer*, which is used to refer to both the action of “renting something (from someone else)” and that of “renting something out (to someone else)” (the same applies to It. *affittare*, Sp. *alquilar* and Cat. *arrendar*). Other examples include:

Late Lat. *inodiare* “hate (somebody)” → “be considered hateful (by somebody)” (→ Fr. *ennuyer* “bore somebody, etc.”)

Late Lat. *abhorrescere* “feel repugnance (for something)” → Sp. Pg. *aborrecer* “*id.*; evoke a feeling of repugnance (in somebody)”

Sp. *gustar de algo / alguien* “like something / somebody” → *gustarle a alguien* “be liked by somebody”

In these cases, the relationship between the two semantic roles is inverted; in a certain sense, this can be seen as the lexicalisation of a grammatical semantic relationship. Cases of auto-converse change are less characteristic of nouns and adjectives; *cf.* however, examples such as Fr. *hôte* < Lat. *HOSPES* “person who gives hospitality, host” → “guest”.

Finally, it should be noted that each lexeme is simultaneously part of both taxonomic and metonymic networks (the colour red belongs to the semantic field of colours, is involved in the naming of red fruits and may evoke frames associated with “fire” (→ “fire truck” → “danger”).

9.3.4 Metaphor

Metaphor constructs a relationship of similarity between two concepts that belong to two different taxonomies and that are not associated within a common frame. In a process that is sometimes referred to as ‘domain mapping’, it produces a ‘projection’ of a source concept onto a distant target concept. Metaphors emphasise one or more pertinent or ‘salient’ aspects, which are associated through similarity, despite a marked distance between the concepts involved; *cf.* the classic example of metaphorical names for ‘head’ in the Romance languages. These start from a variety of initial concepts, all of which evoke the round shape of the head:

- Sardinian *konka* is based on Latin *CONCHA*, meaning “shell (of a scallop)”;
 - the most wide-spread type is based on Lat. *TESTA*, which signifies “pot, potsherd” (*cf.* Engad. It. Sard. *testa*, O.Sp. *tiesta*, Fr. *tête*, Occ. *testa*); in these examples, semantic change first led via metaphor to the meaning of “skull”, then via metonymy to that of “head”; *cf.* also the more recent popular French terms *terrine* “ceramic dish”, *calebasse* “gourd, calabash”, *carafe* “carafe, jug”, *cafetière* “coffee pot”;
- more recent denominations in French rely on the general meaning of “round or plump fruit” (Fr. *poire* (*se faire fendre la poire*); *cf.* German *Kürbis*, *Birne*); other languages (Greek, Persian, Sanskrit) use elements meaning “summit”, “upper part” for the head (*cf.* Sp. *azotea* “rooftop” → *estar mal de la azotea*); *cf.* also *UNGULA* and *GAMBA* below.

Examples of metaphorical change are less frequent than taxonomic and metonymic change. The last two mechanisms make use of associations that already exist in the brain, whereas metaphors link concepts between which there are no pre-established neural links. Consequently, metaphors are less common, more unpredictable and they create a more striking impression.

The similarity upon which metaphors are based relies, in turn, on the psychological phenomenon of ‘salience’: the concepts concerned share at least one salient feature, that is, a distinctive feature that is especially easily perceived or pertinent for the description of the concept.

The animacy hierarchy also plays a determining role in the process of metaphori-sation. As shown above in the chapter on argument hierarchy (8.4.3 no. 2), the animacy hierarchy consists of the following elements:

speech act person → humans in general → animals → plants → countable objects
(natural / artificial) → uncountable objects → abstract entities

Often, a salient feature of one of these categories is transposed onto another (i.e. from humans to animals or plants and vice versa, or from a concrete, often verbal concept, to an abstract concept, etc.). In this way, the distinctive feature ‘non-human’ can be

replaced, for instance, by the feature ‘human’, while the other salient elements of a concept are maintained.

As an example, consider the use of the verb *ruminate* with reference to a person thinking instead of an animal chewing (the same goes for Fr. *ruminer*, It. *ruminare*, Sp. *rumiar*) or the numerous examples of verbs denoting ‘shouting’ such as *bellow* or *roar* (cf. Fr. *beugler*, *mugir*, It. *muggire*, *ruggire*, Sp. *bramar*).

The following examples illustrate the different cases:

animal → human:	Sp. <i>zorro</i> , Fr. <i>renard</i> “fox” → “cunning person”
object → human:	Fr. <i>cafetière</i> “coffee pot” → Pop. Fr. ~ “head” (cf. the other examples given above for names of fruits used to designate the “head”)
human → plant:	Sp. <i>barba</i> “beard” → “tangle of fine roots or shoots”
human → object:	Lat. SATELLĒS “bodyguard” → Fr. etc. <i>satellite</i> “satellite”
animal → object:	Fr. <i>aile</i> , It. Pg. <i>ala</i> “wing” → “wing of a building”; Fr. <i>souris</i> , etc. “mouse” → “computer mouse” (English loanword, cf. 9.6.3 no. 2)
concrete → abstract:	Lat. SAPERE “to have a taste for” → “to be able to judge; understand, know” (i.e. change from “to be able to distinguish flavours” to “be able to distinguish concepts”) Lat. PENSARE “to weigh up” → Romance “to think (about), ponder” Late Lat. TROPARE (9 th century, [borrowed from Greek]) “to find” → “to speak in tropes, to invent melodies” (by metaphor) → “to compose (by metonymy)” (cf. Pfister/Lupis, <i>Introduzione all’etimologia romanza</i> , 2001: 119–123, recently corrected by Georgescu/Georgescu, <i>Fr. trouver, occ. trobar</i> , 2020)

The reverse, that is, the change from abstract to concrete, is highly unusual.

The substitution of distinctive features is less characteristic of metonymy, but nevertheless occurs (cf. the examples of terms for “messenger” and “message”, which reflect the substitution of the feature ‘human’ for ‘object’ and vice versa, 9.3.3 no. 1 above).

Two other metaphorical constellations appear in the phenomena of antiphrasis and auto-antonymy (cf. Blank 1997: 220 *sqq.*): antiphrasis expresses similarity that relies on extreme contrast, involving an axiological inversion (i.e. an inversion of a value judgement: positive evaluation → negative evaluation):

It. *famigerato* “famous” → “infamous”

It. *donnola* “damsel” and *bellola* “beautiful creature” → “weasel”; these examples involve the naming of a concept that is taboo, weasels being considered deceitful. The French form *belette* is probably derived from a Celtic basis **bel-* “white, gleaming”, even though it is associated with *BELLUS* by popular etymology; cf. also the synonymous Rom.n *nevăstuică*, literally “little bride”; cf. Pfister/Lupis, 2001: 174–181.

O.Fr. *oste* “guest” → “hostage”

Numerous antiphrastic denominations for “prostitute”: It. *benefattrice*, normally “benefactress”, *serva di Dio* “servant of God”; Fr. *respectueuse*, *religieuse*, *sœur* “nun, sister” (with parallel expansions within the same frame, cf. *couvent* “convent” → “brothel”, (*mère*) *abbesse* “abbess” → “madam, woman running a brothel”); Pg. *filha-de-santo* “saint’s daughter”.

Auto-antonymy reflects an even more marked opposition between two semantic extremes:

Lat. *DEVOTARE* (and O.Fr. *devoer*) “devote, dedicate” → “damn”

Fr. *sacré* “sacred” → “damned” (*sacré menteur* ! “damned liar !”)

cf. the following functional words (where the semantic change is a result of reanalysis and grammaticalisation): Fr. *rien* “thing” → “nothing”; *aucun* “somebody” → “nobody”; *personne* “id.”

These two types of change are very striking from a semantic point of view, though relatively infrequent in usage.

9.3.5 Changes by formal motivation

A semantic change may also be motivated by formal proximity between two words (‘paronomasia’), similarity or resemblance of forms (popular or ‘folk’ etymology) or formal contiguity (ellipsis). Folk etymology relies on false etymological interpretation by speakers: two lexemes whose forms resemble one another are interpreted as originating from an identical family. This causes semantic interference: e.g. O.Fr. *faissier* “beat with a switch” (< *faisse* “bundle of switches” < *FASCIA*) is associated with *fesse(s)* “buttocks” (< *FISSA* “id.” < *FISSUM* “split, cracked”) and its form and meaning are altered as a result: today, *fesser* means “spank” (cf. 9.4.5. no. 3, blending).

Folk etymology thus describes the incorporation of a lexeme within another genetic family by means of false etymological attribution, whereby the new association provokes a phono-semantic reorganisation of the words involved. This shows that people continually re-evaluate the semantic cohesion of vocabulary, even though this takes place at an unconscious level – language always involves an element of self-reflexiv-

ity. Moreover, folk etymology emphasises the cognitive value of scientific etymological methods (cf. 9.8): the grouping of lexemes into families of common origin is in harmony with the brain's handling of language.

While folk etymology is caused by similarity between two simple forms, ellipsis affects complex forms (such as compounds and phraseologisms) by abbreviating them: in this case, semantic change is thus triggered by contiguity between the initial longer and the resulting shorter forms. More specifically, only the resulting abbreviated form actually undergoes semantic change, as it absorbs the meaning of the original phraseologism, adding it to those it already possesses (cf. 9.5). Below are some examples (cf. 9.3.1, *voiture*, *diligence*, *cabinet*):

Fr. *bas-de-chausses* “socks” → *bas* “id.”

It. *borsa di studio* “scholarship” → *borsa* “id.”

Sp. *café cortado* “small coffee with milk, cortado” → *cortado* “id.”

Lat. TEMPUS HIBERNUM “winter” → *hiver* “id.”, Sp. *invierno*, It. Pg. *inverno*.

→ Maiden, *Folk etymology and contamination in the Romance languages*, OxfEnc

9.3.6 General observations

1. A large number of metaphors, as well as some metonymies, have a euphemistic or dysphemistic (pejorative) character. In such cases, semantic change involves a strong element of expressivity and contributes to the creation of a more evocative term. Dysphemism is a frequent phenomenon, as shown by the previous examples of PLANGERE (metonymy) and CAPUT (metaphor) as well as by metaphors involving a change of the type ‘animal → human’:

Lat. UNGULA “hoof, claw” → Romance “finger/toe nail” (It. *unghia*, Fr. *ongle*, Sp. *uña*, Pg. *unha*, Rom.n *unghie*)

Lat. GAMBA “paw, foot of an animal” → Romance “leg” (It. *gamba*, Fr. *jambe*)

Note that while these changes imply an increase in semantic connotation, this is always followed later by a weakening of meaning, which begins at the point when speakers are no longer aware of the initial motivation. The term *ongle* no longer evokes the meaning of “hoof, claw” in French and has thus lost its character as an evocative expression.

The process of change is similar for *mon-sieur*, *ma-demoiselle*, *ma-dame*: in Old French, these terms were reserved for the nobility; using them with reference to commoners initially implied a euphemistic element, whose meaning was gradually weakened through use and was subsequently lost during the evolution towards present-day French. The process is thus analogous to grammaticalisation (cf. 5.1.3):

- the change introduces an element of expressivity: e.g. the spoken Latin word *MANDUCARE* “chew” is used as a more expressive form alongside *EDERE* “eat”;
- a subsequent phase of routinisation results in the weakening of expressiveness;
- in a process of lexicalisation, the new term replaces the old one, becoming the (new) basic form: *EDERE* was replaced by *MANDUCARE*, cf. Fr. *manger* (< It. *mangiare*), Rom.n *mânca* (cf. the analogous reinforcement of Lat. *EDERE* by means of the prefix *CUM* in Sp. *comer*)

2. Note though that in practice, such categorisation is often not easy. The theoretical distinction between different types of change is relatively clear (identity vs. contiguity vs. similarity), but a reliable identification and description of the initial meaning and the target meaning can prove difficult for previous language stages and knowledge of the historical context is required.

3. Frequently, different mechanisms of semantic change act in sequence; below are two examples of multiple transformations, one of which is unambiguous, while the other allows for two different interpretations:

- Fr. *grève*: spoken Latin **GRAVA* “gravel” (not attested) became O.Fr. *greve* “river-bank” (by metonymy); on the basis of this frequently used term, the toponym *place de la Grève* (in Paris but also elsewhere) was formed (once again by metonymy: it refers to a public square adjacent to the river Seine). This toponym then underwent deonymisation to form the common noun *grève* “strike” (by ellipsis and simultaneous conversion, as a result of a third metonymical change: the square where the Parisian unemployed gathered).
- The Latin word *PLICARE* “bend, fold” has survived throughout much of the *România*: Fr. *ployer*, *plier*, Engad. *plajar*, Friul *pleá*, It. *piegare*, Occ. *plegar*; in the Ibero-Romance languages it has taken on the meaning of “arrive” (Sp. *llegar*, Pg. *chegar*; cf. the secondary meaning of Cat. *plegar* “stop (working)”, thus ‘arrive at a goal’), whereas in Romanian, the opposite meaning of “leave” (*a pleca*) has emerged.

With regard to the change in Ibero-Romance, previous research has assumed that two successive metonymic changes occurred: → * “approach” (when folding a sheet of paper, the two halves approach one another) → “arrive” (the movement of approaching anticipates an arrival).

However, the following interpretation is historically more plausible: “fold” could have been used in a maritime context by ellipsis to mean * “fold up the sails”, which can easily designate the act of “arriving” (by a metonymic change of the type ‘cause-effect’; cf. the corresponding phraseologism in Fr. *mettre les voiles* “to leave” and, in the same vein, the synonym *lever l’ancre* or Rom.n *a ridica ancora*, literally “to raise anchor”). A semantic influence of the contiguous Latin form *APPLICARE* “apply, place (against)” has also been noted.

The meaning of Romanian *a pleca* could also be explained by ellipsis, whereby “fold” may have been used in a society dominated by nomadic culture to mean **“fold the tents”*; hence, by metonymy, “leave” (cf. also Mod.Fr. *faire ses bagages* or even *plier bagage* “leave”).

4. Finally, the lexical meaning of a word can also incorporate the connotative or syntagmatic qualities that are linked to it (cf. 9.2.2 no. 5, *Drei-Ebene-Semantik*), particularly if a word belonging to a given diasystematic variety is borrowed by another: within the new variety, the original connotations become part of its denotation (‘Latinism’, ‘popular word’).

- e.g. *rastaquouère* “stranger with a conspicuous appearance” (< Am. Sp. *rastacuero*, composition consisting of the root *rast(r)a-* from *arrastrar* and *cuero*, initially used to denote leather merchants). The term must have been introduced into familiar Fr. from popular Fr. and is used with pejorative and xenophobic connotations.

9.3.7 Historical onomasiology

1 The purpose of historical onomasiology

Onomasiology is the study of the different forms that are associated with a given concept (cf. 9.3.3). It developed in conjunction with the *Wörter und Sachen* movement in particular (cf. 4.2.1 no. 4; 9.8; 2.2.4 no. 1). Applied to historical questions, it constitutes a subdiscipline of historical semantics: instead of starting from the form of a lexeme and studying the evolution of its meaning across time (termed the ‘semasiological’ method), historical onomasiology takes the concept as a starting point and studies the history of its various denominations.

Historical onomasiology thus analyses:

1. the way in which the forms used to express an individual concept (e.g. ¹HEAD¹), are replaced across the centuries (e.g. CAPUT by TESTA); information linked to the lexemes in question, such as meaning, frequency and context can then be used in order to determine:
2. whether the initial concept has remained stable (as in the case of ¹HEAD¹) or if it has changed over time (e.g. the substitution of O.Fr. *ost* by *armée* in the 15th century coincides with a change in the nature of armies).

Consider two examples of transformation concerning a concept that has remained unaltered:

1. The evolution of words designating the concept ¹MOUTH¹ in the Romance languages: the Latin word *os* is replaced by words signifying ¹CHEEK¹ or ¹THROAT¹:

- Lat. BUCCA “cheek” → Fr. *bouche*, It. *bocca* “mouth” (by metonymy, based on an association with the contiguous ‘MOUTH’, through the notion of ‘EATING’)
- Lat. GALTA “cheek” → Gasc. Béarn. *gaute* “mouth”
- Lat. GULA “throat” → Rom.n *gură* “mouth” (also by metonymy, based on the notions of *yelling*, *producing loud sounds* and *speaking*)

Pop. Fr. *gueule* reflects a separate development, even though it also goes back to the etymon GULA: it emerged through the replacement of the “animal” feature by the “human” feature and implies a dysphemic element, as in the case of the metaphoric use of *griffe* “claw” for *ongle* “nail”, which, in turn, is based on Lat. UNGULA “hoof, claw”).

These transformations show that ‘MOUTH’ is associated with the two notions of *eating* and *speaking*, which thus constitute salient features of the concept.

2. The different words for ‘KILL’ in Latin and the Romance languages: three Latin words convey this concept:

- INTERFICERE does not survive in any Romance language;
- OCCIDERE survives in It. *uccidere* and Rom.n *a ucide*, as well as in O.Fr. *occire* with numerous derivatives; in Mod.Fr. it only survives in the form of the participle *occis*;
- NECARE survives in Aromanian with its original Latin meaning; the meaning of the French hereditary form *noyer* “drown” is the result of specialisation.

In the Romance languages, moreover, new words appear that, according to a figurative description by Wartburg, “are launched like satellites into the centre of gravity of a concept”. Cf. the example of ‘KILL’:

- Fr. *tuer* “kill” takes the place of OCCIRE by means of the following process of semantic evolution: TŪTARI / *TŪTARE “defend” → * “place ashes on the fire” (specialisation) → “extinguish” (metonymy cause-effect) → “kill” (metaphor);
- It. *amazzare* is a new metonymic formation based on *mazza* “large stick” (< MAT-TEA): “beat with a large stick” → “kill”;
- in Sp., *matar* has become established (its origin is unknown, although Arabic *māt* “death” has been suggested as a possible ancestor);
- Rom.n exhibits a Slavonic borrowing: *a omorî*.

In all these cases the concepts remain stable; only the words used to denote them change. In doing so, they sometimes provide clues as to the recurrent associations linked to the concept at a given time.

2 Historical and anthropological interpretation of semantic changes

The analysis of semantic changes requires the linguistic factors described above to be taken into account, in addition to the socio-historical context in which they occur³⁰. As we have seen, the recurrent associations to which semantic change bears witness may allow the identification of both anthropological and historical elements that otherwise remain inaccessible, thus opening a window onto the ‘collective unconscious’ of an illiterate society of earlier times. The examples provided clearly show that the new meaning of a lexeme reflects true semantic associations that must have been current within the community of speakers concerned.

Analogous associations that appear across different languages allow interpretations in terms of historical anthropology, such as:

- the link between the possession of domestic animals and money (in a pastoral society):

Fr. *gagner*, It. *guadagnare* developed from Old Low Franconian **waiðanjan* “to graze, send animals to graze”;

Lat. *PECUNIA* derives from *PECUS* “domestic animal, sheep”;

Engl. *fee* (cf. German *Vieh*) is ultimately based on Old Low Franconian **fehu* “live-stock, cattle” (cf. FEW s.v.);

- the link between work and suffering:

Fr. *travail*, Sp. *trabajo* etc. < Lat. *TRIPALIUM* “instrument of torture consisting of three prongs”; cf. also It. *travaglio* “anguish, suffering; labour pains”;

the etymology of Lat. *LABOR* and the Germanic word *Arbeit* both similarly rely on terms designating pain and suffering (cf. Pfister/Lupis, *Introduzione all’etimologia romanza*, 2001: 115);

- the link between a habitable space and a fortification:

Occ. *bastida* < Germ. **bastjan* “surround by a fence” (FEW 15/1, 76b, s.v.);

for Engl. *town* cf. German *Zaun* “fence” (cf. Pfister/Lupis, *ibid.*, 2001: 112 sqq.).

Independent and parallel processes of evolution such as the above allow the identification of stereotypical associations in specific cultural contexts. In other cases, the constants of semantic change shed light on cognitive associations of a universal type, as illustrated by the animacy hierarchy: the conceptual network is undoubtedly organised in an anthropocentric manner, beginning with the human body and our everyday

³⁰ For certain concepts, it is also necessary to take into account the effect of taboos (consider 9.3.4, ‘WEASEL’, ‘PROSTITUTE’); the erotic taboo is frequent, an example being It. *fica* “female sexual organs”, undoubtedly based on the shape of a fig).

surroundings (relatives, animals, food, shelter, nature and climatic phenomena, professions, etc.).

Historical semantics therefore offers a starting point for studies in historical anthropology. The focus of such research can either be cultural or universal. Within this domain, psycholinguistics and cognitive linguistics complement older contributions made by research in onomasiology and structural semantics.

3 The disappearance of words

Historical onomasiology also deals with the problem of why and how a particular word disappears. It has been shown that the introduction of new terms can be explained as a result of a desire for expressivity or as a consequence of extralinguistic transformations, and the regularly occurring loss of vocabulary is the necessary counterpart of these processes.

The regularities governing word loss are not easy to discern. Certain factors can nevertheless explain the replacement of one word by another:

- a marked phonetic weakness:

APEM > O.Fr. *af* ~ *ef* ~ *es* ~ *e* → *avette* ~ *abeille* “bee” (borrowed from Occitan, cf. 4.2.2);

O.Fr. *ost* → *armée* “army”: further evolution would have reduced [ɔst] to [ɔ]; in this particular case, the extralinguistic motive of the military transformations mentioned above comes into play;

- a relative semantic weakness in the case of concepts that naturally call for expressivity:

IOCARI “to joke” (maintained with the semantically weakened meaning of “to play”) → Fr. *blague* (< Germ.), It. *scherzo* (< Germ.), Sp. *burla* (of unknown origin);

- homophony of words that belong to the same communicative context and that can thus be confused (cf. Blank, *Prinzipien*, 1997: 354 sqq.):

O.Fr. *moudre* “to milk” (< MULGERE) ~ *moudre* “to grind, mill” (< MOLERE), although the issue of MULGERE has survived in northern French dialects;

O.Fr. *noer* “to swim” (< *NOTARE) ~ *noyer*, *neier* “to drown” (< NECARE, cf. (1) above 9.3.1/9.3.7 no. 1); homonymy with the family of NODUS “knot”, which belongs to other contexts, does not play a determining role here;

Spoken Late Lat. AVENA “reins” (< HABENAE) ~ *avena* “oats” (< AVENA);

the classic example of this type is Gasc. *gat* “rooster” (< GALLUS) ~ *gat* “cat” (< CATTUS) → *bigey* “rooster” (< VICARIUS) or even *hasâ* “id.” (< PHASIANUS); cf. Gilliéron / Edmont 1912;

Fr. *stable* ~ *étable*, which possibly illustrates a case of re-Latinisation due to the fact that the two words, which differed in Latin, became homophonous in O.Fr.;
cf. also the example of MATERTA / MATRE and PATRUU / PATRE (9.3.1);

- morpho-semantic isolation:

Lat. SUPERBUS “haughty” (survives as a learned borrowing: *superbe* “superb” in Fr. and *superbo* in It.) → Fr. *hautain*, Sp. *altivo*, which belong to the highly frequent family of Lat. ALTUS → Fr. *haut*, Sp. *alto*;

- (relative) morphological isolation:

O.Fr. *clore* → Mod.Fr. *fermer* (a verb of the 3rd group is almost entirely replaced by a verb of the 1st group).

Often, a word does not truly ‘disappear’ from a language but its frequency and context of use simply change. It may move from one (sociolectal or regional) variety to another. Onomasiological changes thus overlap with changes that occur within the geolinguistic space occupied by a language. In addition, chronological evolutions can intersect with geolinguistic patterns within the dialect space of a given language (cf. 4.2) or within that of all Romance languages (cf. 3.5.3).

4 Application areas of historical onomasiology

In the first half of the 20th century, a large number of specific studies were conducted using the onomasiological approach, such as works analysing the Romance forms which developed from a particular Latin term:

PERSONA: Rheinfelder, *Das Wort “Persona”*, 1928

CAPTIVUS: Haerle, *Captivus – cattivo – chétif; zur Entwicklung des Christentums auf die Terminologie der Moralbegriffe*, 1955

CALERE: Christmann, *Lateinisch “calere” in den romanischen Sprachen: mit besonderer Berücksichtigung des Französischen*, 1958

Other studies are devoted to the evolution of specific concepts from Latin to the Romance languages, e.g.:

the concept ‘FACE’: Renson, *Les dénominations du visage en français et dans les autres langues romanes*, 1962

cf. 9.2.3: Tappolet, Zauner etc.

Studies in this field, however, have become rare in recent years and examples such as that of Kramer, *Die Sprachbezeichnungen Latinus und Romanus* (1998) remain an exception (cf. 9.3.3).

Research in historical onomasiology may adopt a range of different orientations (cf. 9.8), such as:

- the study of transformations affecting the forms used to designate the most frequently used or salient concepts of a language; in other words the study of its basic vocabulary (everyday references / ‘*Lebenswelt*’) or key concepts in Western culture;
- the study of concepts that are representative of specific cultural trends of a given historical period (e.g. the vocabulary of intimacy of the 20th century, the political vocabulary of the French Revolution or the Fascist period, the vocabulary of feudal culture in the Middle Ages);
- the study of specific semantic fields in literary texts (e.g. vocabulary relating to love, adventure or spirituality in a specific circle at a given period);
- the study of scientific or technical vocabulary of different periods (e.g. legal, medical, botanical, mathematical, artistic and architectural vocabulary).

Investigations of this type can be carried out on the *Romània* as a whole, as in the case of Stefenelli’s work *Das Schicksal des lateinischen Wortschatzes in den romanischen Sprachen* (his corpus, however, was chosen based on word frequency rather than on lexical meaning). Studies in semantics may encompass several Romance or non-Romance languages. They can be applied to a specific language or a clearly defined historical period and they can be restricted to an individual variety of a language (such as the Lorrain dialect or popular French) or a defined textual genre (e.g. literary texts). Any chronological and diasystematic segment can constitute the subject of analysis, be it the 16th century as a transitional period between medieval and modern language, the period at which the first Romance texts emerged, or the present-day period. Any combination is legitimate in theory and all rely on the same basic techniques, but all studies involving both historical semasiology and onomasiology are characterised by the combination of the history of words with the history of concepts.

→ Schmitt, *Wörter und Sachen*, LRL 1/1, art. 6
 Söhrman, *Onomasiological differentiation*, OxfGuide 33

9.4 Word formation in the Romance languages

9.4.1 Functions and general characteristics

Word formation encompasses a series of processes that aim to transform and expand vocabulary, and that operate both in parallel with and in conjunction with semantic change (cf. 9.3). They rely on the two principal mechanisms of derivation and compounding:

- derivation forms new words using derivational affixes:
Fr. *maison* + *-ette* → *maisonnette*, *rouge* + *-âtre* → *rougeâtre*; cf. English *red* + *-ish* → *reddish*;
- compounding forms new words by combining two or more lexemes:
Fr. *gratte* + *ciel* → *gratte-ciel*; cf. English *sky* + *scraper* → *skyscraper*.

Word formation has three main impacts on language processing:

- both derivation and compounding enable new concepts to be expressed by combining various existing lexical elements;
- derivation may also result in a change of word class (*réalis-er* (v) + *-ation* → *réalisation* (N)), thus expanding the possibilities of expression in accordance with the principle of linguistic economy;
- moreover, when considered from a historical perspective, derivation can be seen to counterbalance the phonetic reduction of lexemes that results from their natural evolution (cf. 6.7 and 8.3.7 no. 2 for the example of O.Fr. *af*, *e(f)* → *avette*, *abeille*).

The semantic aspects of word formation require particular attention, since the meaning of a constructed word may draw from its constitutive elements in different ways. In some cases, the semantic content of a suffix can easily be identified, and is simply ‘added’ to the meaning of the lexical basis; however, even in such a ‘simple’ scenario, the derivational process generates a new semantic combination. Hence, one may speak of meaning inferred through the affix. Consider the following examples from French:

le X_{verb base} + *-eur* “one who X’s” (regularly) → *le vend-eur* “one who sells, whose profession is that of selling” [Engl. equivalent: *X* + *-er/-or*, cf. *seller*, *vendor*];

la X_{verb base} + *-aison* “the act of X’ing and the result of this act” → *la livr-aison* “the act of delivering; that which is delivered” [Engl. equivalent: *-ation*, cf. *realisation*].

In some cases, however, ‘word meaning’ is very specific and diverges from the so-called ‘word formation meaning’ that can be predicted on the basis of the components of the word and is rather vague and general (cf. 9.4.6).

The most important quality of word formation is its ability to create or enlarge word families (e.g. *jou-er*, *jou-eur*, *jou-euse*, *jou-et*, *jou-able*, *jou-jou*; cf. English *play*, *play-er*, *play-able*, *play-ful*, *play-thing*, etc.), which are immediately recognisable as being semantically coherent. This facilitates the process of memorising a large number of individual words; hence, like polysemy, word formation relieves the burden on the cognitive system (9.2.2 no. 7 (8)). Vocabulary processing is aided by the fact that a single lexical form can have multiple meanings that are linked to one another by given semantic

mechanisms, as well as by the fact that multiple forms share a common base and exploit a limited number of specific derivational mechanisms³¹.

Owing to its predictable and ‘mechanical’ aspect, word formation belongs to the domain of morphology, and many recent studies on the subject of derivation are conceived within the traditions of morphological research. Like inflection, the construction of new words operates on the basis of various ‘preprogrammed’ and thus predictable processes, which can be applied to different base forms. However, due to its essential role for lexis as well as its strong semantic implications, it also belongs to the domain of lexicology. With the passing of time, derived forms often need to be memorised as a whole owing to semantic specification (*cf.* 9.4.6). Word formation therefore has an ambiguous status that situates it between grammar and lexis (*cf.* Laca’s fundamental study, *Die Wortbildung als Grammatik des Wortschatzes*, 1986). Both approaches are necessary: at the moment at which a new lexical form is created, the grammatical aspect is predominant; if it remains in use, it becomes a constitutive part of vocabulary.

→ Rainer, *Derivational morphology*, OxfGuide 28

Bauer, *Word formation*, CambrHist 1, 10

OxfEnc, *Peculiarities of word-formation* (Portuguese, Rio-Torto; Romanian, Grossmann)

RSG 3, sect. XV, *Histoire interne: formation des mots* (roumain, Vasiliu, art. 224b; *Bündnerromanisch*, Liver, art. 229b; *italiano*, Schweickard, art. 232; *français (Europe)*, Glessgen, art. 237; *français (hors d'Europe)*, Queffélec, art. 238; *catalán*, Bruguera, art. 243; *Spanisch*, Geckeler, art. 247; *Portugiesisch (Europa)*, Monjour, art. 253; *Portugiesisch (außerhalb Europas)*, Schmidt-Radefeldt, art. 254; *Konvergenz- und Divergenzphänomene in der Romania*, Rainer, art. 259)

HSK 40/1–3, general aspects and special cases (*Foreign word-formation in Italian*, Iacobini, art. 95; *Word-formation and purism in French*, Braselmann, art. 98; *The origins of suffixes in Romance*, Pharies, art. 108; *From Latin to Romance*, Buchi/Chauveau, art. 111; *From Latin to Romanian*, Rădulescu Sala, art. 112; *From Old French to Modern French*, Rainer/Buridant, art. 113 etc.)

HSK 40/4, sect. XVI, *Romance* (Portuguese, Pöll, art. 144; *Spanisch*, Rainer, art. 145; *Catalan*, Cabré Castellví, art. 146; *French*, Floricic, art. 147; *Ladin*, Siller-Runggaldier, art. 148; *Sardinian*, Pinto, art. 149; *Italian*, Rainer, art. 150; *Romanian*, Grossmann, art. 151)

9.4.2 Derivation in the history of the Romance languages

Morphological derivation involves the combination of free word bases with bound derivational affixes. The latter may cause changes in word class, meaning and/or context of use. The forms affixes take are in most cases relatively stable and they are therefore

³¹ More detailed statistics in this area yield striking results (*cf.* 9.1.3 for the example of the analysis of English vocabulary taken from a selection of newspapers): roughly 45,000 base words are found alongside 140,000 derivatives whose meaning is easily recognisable, leaving only 43,000 derivatives whose meaning cannot be understood without prior knowledge. Hence, words with a transparent meaning constitute 60% of the lexical forms in this case.

easily recognisable, as illustrated by the examples provided below. The lexical base, on the other hand, may display notable variation as a result of varying stem forms within the same stem space. In order to identify the different stem forms of a base word, and thus its stem space, the affixes must be removed. This raises a multitude of questions, such as the following:

- should the French base forms *pan-* (as in *pan-ier* “basket”, *pan-ade* “bread crumbs”, *pan-ification* “bread-making”) and *pain* “bread” be considered as two variants situated within a single stem space, or should it be assumed that these forms belong to two distinct word families (one of which displays a Latinising tendency, cf. Lat. PANIS, PANEM)?
- how should cases of derivatives without pre-existing bases be interpreted? (cf. 9.4.4 no. 4)

Space considerations prevent us from listing all relevant cases here; however, for the sake of illustration, we will briefly comment on some examples of the most frequent types of derivation in the Romance languages.

1 Prefixation

Prefixation occurs when an affix is placed before a word stem (e.g. *im-possible*, *dé-loyal*). This permits changes in meaning, but not of word class. The prefixes AD-, IN-, DE-/DIS-, EX-, RE-, CUM-, SUB-, SUPER-, TRANS-, EXTRA-, frequent in spoken Late Latin owing to their intensifying character, are highly frequent in all the Romance languages. Consider the following examples:

AD- + BATTUERE “hit, strike” → ABBATT(U)ERE > Fr. *abattre* “overthrow”, It. *abbattere*, Occ. Cat. *abatre*, Rom.n *a abate* “divert, distract”

EX- + CAMBIARE “change, exchange” → *EX-CAMBIARE > It. *scambiare* “exchange”, Occ. *escambiar*, Fr. *échanger*, Rom.n *a schimba*

Both examples are of (late) Latin formation, and survive in the majority of the Romance languages. Even if they are recognisable as ‘constructed words’ in their respective languages, they ultimately represent the continuation of a Latin word with no derivational change.

Prefixation involves verbs in particular, although there are also examples of nominal derivatives (e.g. Fr. *nom* “name” → *prénom* “first name”, It. *nome* → *cognome* “last name”, *soprannome* “nickname”).

2 Suffixation

Derivation by means of a suffix, i.e. an affix that is added to the end of the base word, is a more diverse process than prefixation in the Romance languages, carrying more

extensive formal and semantic implications. It can easily lead to changes of word class (nouns, proper nouns, verbs, adjectives → nouns, verbs, adjectives, adverbs) or gender (in the case of nouns).

The semantic content of suffixes is highly variable: in some cases it is limited and monosemic (i.e. with a single meaning), lending newly-constructed words considerable semantic transparency. Many suffixes, however, are polysemic and thus capable of conveying diverse meanings. Below, examples of suffixing are provided, first without, then with change of word class (examples are taken from H. Lüdtke, *Wortbildungslehre*, LRL 2/1 art. 106).

2.1 Suffixation without word-class change

Diminutives:

- ELLU: Rom.n *orășel* “small town”, Sard. *belleddu* “pretty”
- ITTU: It. *vecchietto* “old person”; Cat. *petitet*; Occ. *bosquet* “thicket”, Fr. *jouet* “toy”, *archet* “bowstring (of a musical instrument)”, *coffret* “case, small box” (the suffix is no longer fully transparent in Fr.)
- ITTA: Sp. *casita* “cottage, small house”; O.Fr. *estoilete*; Mod.Fr. *maisonnette* (suffix is transparent)
- INA: It. *manina* “small hand”

Relational nouns (i.e. nouns that have the function of “X in relation to the base word”), referring to persons:

- ARIU: It. *carbonaio* “coalman; charcoal burner”, *carbonaro*, Rom.n *cărbunar*, Fr. *charbonnier*, Cat. *carboner*, Sp. *carbonero*, Pg. *carvoeiro*; learned terms: Fr. *fonctionnaire* “official”, It. *funzionario*, etc.
- ND-ARIA: Fr. *filandière* “spinner (woman)”, *lavandière* “washerwoman”, etc.

Relational nouns referring to objects:

- ARIU: It. *granaio* “granary”, Fr. *grenier*, Cat. *graner*, Sp. *granero*, Fr. *guêpier* “wasp nest”
- ARIA: Sp. *hormiguera* “anthill”, Pg. *floreira* “flower vase”

Collective nouns:

- ETU: It. *faggeto* “beech forest”, Rom.n *făget*; Cat. *vinyet* “vinyard”; Sp. *robleto* “oak forest”; Pg. *olmedo* “elm forest”; O.Fr. *sapoi* “pine forest” (= *sap* “fir” + *-oi*)
- ATICU: Fr. *visage* “face”, *langage* “language”, *feuillage* “foliage”, Pg. *folhagem*; O.Fr. *barnage* “assembly of barons, baronage”; Pg. *folhagem*; It. *coraggio* “courage”
The suffix -ATICU is also used for the formation of deverbal nouns that express the action of the verb in question (Fr. *lavage*, *mariage*, *passage*, It. *passaggio*).
- ALIA: It. *ferraglia* “scrap iron”; Fr. *ferraille*; Cat. *jovenalla* “(the) youth”

Approximation:

var.: It. *bianchiccio* “whitish”, *dolcigno* “mildly sweet”, *belloccio* “cute”, *verdognolo* “greenish”, *ordinariotto* “slightly ordinary”; Fr. *maladif*, *vieillot*, *ellipsoïde*; Sp. *rojizo* “reddish”, *blanquecino*, *azulenco* “blueish”, *amarillento* “pale yellow”, *amarillejo* “yellowish”, *verdoso* “greenish”, *verduco* “id.”, *negruzco* “blackish”

→ Fortin/Rainer, *Evaluative suffixes*, CambrHandb 14

2.2 Suffixation with word-class change

The process of suffixation allows all possible changes between the four lexical word classes:

- nominalisation (V → N, ADJ → N),
- verb formation (N → V, ADJ → V),
- adjective formation (N → ADJ, V → ADJ)
- adverbialisation (ADJ → ADV).

In the section below, examples of deverbal nominalisation will be provided, accompanied by detailed remarks. These will be followed by examples for the remaining types, this time without commentary.

Deverbal nouns (V → N):

-(T)URA (Latin formations): ARARE → ARATURA → Rom.n *arătură* “ploughing”; COOPERIRE “cover” → COOPERTURA → Logud. (= Sard.) *kobertura* “roof”;

-(t)ura (Romance formations): Lat. TEXERE “to weave” → *tessere* → It. *tessitura* “weaving” (1640); Lat. BURRA “heavy woollen fabric” → Sp. Pg. *borra* → *borrar* → *borradura* “erasure” (1571); Lat. D(E)AURARE → Occ. Cat. *dauradura* “gilding” (14th century), Lat. COOPERIRE “cover” → Fr. *couvrir* → *couvert* → *couverture* “roof” (ca. 1160–1170), “cover, blanket” (ca. 1180–), O.Fr. *doreure* (ca. 1167) → Fr. *dorure*.

The Romance successors of the Latin suffix -(T)URA, as well as the learned forms of the Latin suffix itself, remained productive. The dates given for the Romance formations extend from the period of Old French (1167) to the 17th century (note that one should assume that the words documented for the first time in the 12th century must have been formed much earlier; however, more precise dating is only possible in the case of the vernacular words which were introduced into Latin texts of the 9th to 11th centuries, as almost no Romance texts exist for this period.

From a semantic perspective, formations of the type $X_{\text{verb base}} + \text{-(T)URA}$ display relatively little specificity. They denote “that relating in a variable manner to X (depending on the referents and context of use in question)”, or “the action or the result of an action denoted by X”. Semantic differences are highlighted by examples such as Fr. *couverture* “cover” and Sard. *kobertura* “roof”; as is common for nouns, these forms often denote

the results of an action (*dorure*); reference to the action itself may also persist (*arătură*, *borradura*).

-MENTU: MOVERE → *MOVEMENTU (formation not attested in Latin but nevertheless pan-Romance) > It. *movimento* “movement, motion” (13th century), Fr. *mouvement* (ca. 1100), Occ. *mo(u)vemen(t)* (13th century), Cat. *moviment*, Sp. *movimiento* (1250), Pg. *movimento*

The formation *MOVEMENTU is not documented but it may have emerged during the period of Late Latin; however, the Romance lexemes may also constitute parallel later borrowings from Medieval Latin, or represent the result of several parallel but independent formations in various Romance languages. In any event, the first attestations of the forms in question appear at a very early stage; this supports the hypothesis that they were formed before the Romance languages were put into writing. The differences can be explained by chronological discrepancies between languages as regards textual elaboration: in French, there were already a number of important texts in the 12th century, while in Italian, it was not until the 13th century that a significant body of texts was produced; moreover, the current state of chronological research on Pg. and Cat. does not allow reliable dating.

-ANTIA: Fr. *espérance* “hope”, It. *speranza*, Occ. Pg. *esperança*, Sp. *esperanza*, all formed on the base of SPERARE; cf. also Cat. *confiança*

This case displays some similarity with that of -MENTU; the suffix is of semi-learned origin (cf. 6.6 and FEW s.v. SPERARE).

-TION / -SION: It. *realizzazione* “realisation”, Fr. *réalisation*, Sp. *realización*, Occ. Cat. *realizació* (← It. *realizzare* etc.)

The above are examples of Romance formations, which are strikingly similar and which all display a tendency towards Latinisation. This undoubtedly reflects a degree of interdependency among the various Romance languages.

-TOR: Lat. PISCATOR > O.Fr. *pescheor* “fisherman”, Fr. *pêcheur*, It. *pescatore*, Sp. *pescador*, Occ. *pescaire*

These are inherited forms based on a Latin derivative: semantically, $X_{\text{verb base}} + \text{-(T)OR}$ refers to “someone who X’s”. The actions in question may be linked to a one-off event (*vainqueur* “winner”) or to a habit (*buveur* “drinker”); the second-case scenario is at the origin of the extended meaning of “someone whose profession is to X” (*pêcheur*, *entraîneur*; cf. the examples of -iste in 9.4.6).

Deadjectival nouns (A → N):

-IA: It. *allegria* “cheerfulness” (from *allegro* “cheerful, happy”), Rom.n *sărăcie* “poverty” (from *sărac* “poor”), Logud. *iskuría* “lack”, Fr. *courtoisie* “courtliness, courtesy” (from *courtois* “courtly, chivalrous”), Occ. Cat. Pg. *cortesia*, Sp. *cortesía*

These examples illustrate formations of different types (Fr. *courtoisie* is a semi-learnèd lexeme while the Rom.n and Sard. examples are inherited forms), dating from different periods. The suffix displays little semantic specificity (“quality of being X”).

-TAS / -TATE: It. *bontà* “goodness”, Rom.n *bunătațe*, Sp. *bondad*, Pg. *bondade*; Occ. Cat. *falsedat* “falseness”

-TITIA: It. *tristezza* “sadness”, Occ. Cat. *tristesca*, Sp. Pg. *tristeza*, Rom.n *tristețe*, Fr. *justice* (learnèd form), *tristesse* (semi-learnèd form)

Denominal verbs (N → v):

-IDIARE: It. *corteggiare* “court sb.”, Fr. *guerroyer* “wage war”, O.Occ. *domneiar* “reign”, Cat. *tor-rejar* “exceed” (← TURRIS), Sp. *falsear* “falsify”, Pg. *bracejar* “gesticulate, move one’s arms”

-IZARE: It. *polverizzare* “pulverise”, Rom.n *a actualiza* “update” (borrowing; in synchrony, there is nevertheless a relationship between *actual* – (a) *actualiza*), Fr. *tyranniser* “tyrannise”, Cat. *atomitzar* “atomise”, Sp. Pg. *moralizar* “moralise”

Denominal adjectives (N → A):

-ALE: It. *musicale* “musical”

-ANU: Fr. *australien* “Australian”, It. Sp. *australiano*

-IBILE: Fr. *paisible* “calm, peaceful” (← O.Fr. *paix* ~ *pais* “peace, tranquillity”)

-ICU: Fr. *volcanique* “volcanic” (← *volcan*), It. *accademico* “academic”

Deverbal adjectives (v → A):

-ABILE: Fr. *jouable* “playable”, *faisable* “doable, feasible”

-ATORIU: Fr. *diffamatoire* “defamatory, slanderous”, It., Sp. *interrogatorio* “interrogatory, interrogative”, Cat. *convocatori* “convocative”, Pg. *reformatório* “reformatory” (without exception learnèd forms)

-(T)IVU: It. *qualificativo* “qualifying”, Rom.n *adoptiv* “adoptive” (borrowing), Cat. *abusiu* “abusive”, Sp. Pg. *administrativo* “administrative” (learnèd form)

Deadjectival adverbs (ADJ → ADV):

See the formations with the suffix -MENTE (cf. 7.4.2).

3 Conversion and improper derivation

Conversion refers to word class changes that do not involve any formal changes, except, in some cases, a modification of the inflectional suffix.

N → v: It. *pianta* “plant” → *piantare* “plant”, Fr. *sucré* “sugar” → *sucrer* “sugar”, Cat. *espiga* “sprig, ear (of grain)” → *espigar* “to come into ear”, Pg. *espinhar* “prick (on a thorn)”

v → N: Fr. *donner* “to give” → *(une) donne* “(a) deal”, *marcher* “to walk” → *marche* “(the activity of) walking”, It. *arrivare* “to arrive” → *arrivo* “arrival”, *sostare* “to stop, to take a break” → *sosta* “stop, break”

A → v: Sp. *caliente* “warm, hot” → *calentar* “to heat up”

In many cases, the beginning and endpoint of a process of conversion can only be identified when considered diachronically. Speakers at a given time in synchrony can often no longer tell which form is the base and which the derivative (*cf.* also the examples of ‘back-formation’ under 9.4.4 no. 1 below).

Improper derivation can be compared to conversion. This process, also known as ‘zero derivation’, implies the complete absence of formal modification. It includes, for instance, the nominalisation of infinitives, very frequent in Portuguese, Spanish, Catalan, Italian, Romanian and Old French:

- Pg. *o falar* “speech”
- Sp. *el susurrar* “whispering”
- Cat. *el parlar* “speech”
- Fr. *le déjeuner* “lunch”
- It. *il fare* “(way of) doing”
- Sard. *su battiari* “baptism”
- Rom.n *părere* f “opinion”

The two terms ‘conversion’ and ‘improper derivation’ are sometimes considered as synonymous, and in English, it is often impossible to distinguish them from each other at the formal level, due to the absence of inflectional markers (consider *plant* (N) → *plant* (v)).

- Rainer, *Adjectival suffixes: from Latin to Romance*, OxfEnc
- Id.*, *Instrumental and place nouns in the Romance languages*, OxfEnc
- Fábregas/Marín, *Nominalizations in the Romance languages*, OxfEnc
- Regis, *Personal nouns (agent nouns) in the Romance languages*, OxfEnc

4 Change of gender

In the Romance languages, changes of grammatical gender are generally expressed through a change of suffix, which can either take the form of substitution or enlargement.

Substitution, M ~ F:

- ARIU ~ -ARIA: It. *lattaio* / *lattaia*, Fr. *laitier* / *laitière*, Sp. *lechero* / *lechera*, Pg. *leiteiro* / *leitera*

Enlargement, M → F:

- TOR / -TRIX: Fr. *moteur* – *motrice*, OOcc. *trobador* – *trobairitz*
- TOR / -TORIA: Cat. Sp. Pg. *-dor* / *-dora*, Rom.n *-tor* / *-toare*
- + -ISSA: Fr. *maîtresse*, Cat. *mestressa*

Current discussions on the masculine and feminine designations for professions or functions (e.g. *le/la ministre*, *avocat/avocate*, *docteur/doctoresse*, *auteur/auteure*) also belong to the morphological domain of gender change.

→ Rainer, *Sex-denoting patterns of word formation in the Romance languages*, OxfEnc

5 Parasynthesis

Parasynthetic formation involves the simultaneous addition of a prefix and a suffix to a lexical base (e.g. Fr. *in-coll-able* “non-stick”). Most of the examples traditionally considered to belong to this category are actually cases of prefixation in combination with conversion (e.g. Fr. *em-bras-s-er*). Their suffixes are therefore inflectional affixes that express a word class change, not derivational suffixes. Examples concern denominal and deadjectival verbs in particular. Consider the following:

BRACCIU	→ Fr. <i>embrasser</i> , It. <i>abbracciare</i> “embrace”
MAGRU	→ It. <i>dimagrire</i> “slim down” (MAGRU is a variant of Class. Lat. MACER)
LONGU	→ Rom.n <i>a alunga</i> “chase, pursue”, Fr. <i>allonger</i> “grow longer, prolong”
FILIU	→ Cat. <i>afillar</i> “adopt”
CLARU	→ Sp. <i>aclarar</i> “lighten, clarify”
LIGERIU	→ Pg. <i>aligeirar</i> “relieve” (LIGERIU is a variant of protorom. *LEVIARIU, formed on the basis of LEVIS, LEVE)

All these formations occurred after the 8th century, which explains why they are not widespread in a large number of languages. It also explains the existence of parallel yet independent derivatives such as *ab-bracci-are* vs. *em-brass-er*, formed by means of two different prefixes. It should, however, be noted that such independent formations involve the recurrent Latin prefixes AD-, IN- and DE-/DIS-.

There are also formations that involve a prefix and a derivational suffix (such as the example *incollable* “non-stick”, above, or *en-col-ure* “neck, collar”). This type of co-presence, however, may reflect two distinct stages of formation: Fr. *dés-agré-able* “disagreeable”, It. *s-grad-evole*, Sp. *des-agrad-able* and Pg. *des-agrad-ável* all display a stem space that corresponds to Latin GRATU, but in all four cases, the suffixation with -ABILE (Fr. *agréable*, It. *gradevole* etc.) seems to predate the prefixation with EX- or DIS-.

9.4.3 Compounding in the Romance languages

Compounding refers to the combination of two or more autonomous lexical units. The semantic links forged in this way generally remain vague (such as “an X which is in some way related to a Y”). The meaning of the compound thus largely depends on the

respective concepts referred to by the lexical units of which it is composed; it further depends on the context in which it is used.

Romance compounds display a variety of word class combinations, even though the results are essentially nominal in nature. Some compounds, like all Romance derivatives, display clear formal unity; their written representation corresponds to ‘graphic words’ and they are written in a single word, like simple lexemes (e.g. Sp. *caradura*, literally ‘hard face’, ‘cheeky person’). Other compounds, however, are separated by a hyphen (It. *parola-chiave*, Fr. *tire-bouchon*, cf. 9.4.1 above) or even split into several graphic words (‘syntagmatic compoundings’: *pomme de terre*, *machine à laver*, cf. no. 2 below).

→ Forza/Scalise, *Compounding*, OxfGuide 29
 Rainer, *Compounding: from Latin to Romance*, OxfEnc

1 Compounds written as a single word or whose components are separated by a hyphen

[N+N] _N :	Fr. <i>bidonville</i> “slum, shanty town”, <i>chou-fleur</i> “cauliflower”
[N+ADJ] _N :	Fr. <i>chaise-longue</i> “deck-chair”, <i>coffre-fort</i> “strong-box, safe” Sp. <i>caradura</i> “cheeky/impertinent person”
[ADJ+N] _N :	Fr. <i>grand-mère</i> “grandmother” It. <i>gentiluomo</i> “gentleman”
[V+N] _N :	Fr. <i>tournesol</i> “sunflower”, <i>couvre-chef</i> “headgear, hat”, <i>essuie-glace</i> “wind-screen wiper” It. <i>girasole</i> “sunflower” Cat. <i>pica-plets</i> “corrupt lawyer” Rom.n <i>pierde-vară</i> “idler”, literally ‘lose-summer’
[V+INDF N] _N :	Fr. <i>fainéant</i> “slacker”, <i>vaurien</i> “rascal”
[V+V] _N :	Fr. <i>savoir-faire</i> “know-how”
[V+ADV] _N :	Fr. <i>lève-tôt</i> “early bird”, <i>passe-partout</i> “master key” (also used as an adjective meaning “all-purpose”)
[ADV/PRP+N] _N :	Fr. <i>après-saison</i> “postseason”, <i>non-usage</i> “non-use”, <i>survêtement</i> “tracksuit”

Adjectival compounds on the other hand are uncommon; verbal compounds remain scarce and are barely productive today:

[ADJ+ADJ] _A , cumulative:	Fr. <i>sourd-muet</i> “deaf-mute”, It. <i>sordomuto</i> , Sp. <i>sordo-mudo</i> , Pg. <i>surdo-mudo</i>
<i>id.</i> , “between X and Y”:	Fr. <i>franco-allemand</i> , It. <i>franco-italiano</i> , Sp. <i>franco-español</i>
[ADV+ADJ] _A :	Fr. <i>malheureux</i> “unhappy”
[ADV+V] _V :	Fr. <i>maltraiter</i> “mistreat”, Sp. <i>maltratar</i>
[N+V] _V :	Fr. <i>maintenir</i> “maintain”, Sp. <i>mantener</i>

2 Syntagmatic compounds

Syntagmatic compounds of the type *machine à laver* “washing machine” and *pomme de terre* “potato” are the most common type of compound in most Romance languages. The surface structure of such compounds is identical to that of free constituents in natural syntax.

[N+N] _N :	Fr. <i>assurance maladie</i> “health insurance”
[N+PRP+N] _N :	Fr. <i>chemin de fer</i> “railway”
[N+PRP+INF] _N :	Fr. <i>salle à manger</i> “dining room”, <i>machine à écrire</i> “typewriter”
	It. <i>macchina da scrivere</i>

N-ADJ compounds, which are sometimes also termed ‘improper compounds’, are a particularly delicate case:

[N+ADJ] _N :	Fr. <i>vin blanc</i> “white wine”, <i>poisson rouge</i> “goldfish”, <i>bande dessinée</i> “comic strip”, <i>Croix rouge</i> “Red Cross”
	It. <i>scala mobile</i> “escalator”, <i>vino bianco</i>
[ADJ+N] _N :	Fr. <i>haute tension</i> “high-tension”

It is thus important to distinguish the lexicalised compounds given in the above examples from free constituents (*vestito bianco* “white dress”). The impossibility of decomposing a compound form without changing its meaning constitutes evidence of its lexical unity (e.g. Fr. *pomme de terre* “potato”: **Cette pomme est de terre*, literally “This apple is made of earth”, **des pommes jaunes de terre* “yellow apples made of earth”, **des pommes de terre sablonneuse* “apples made of sandy earth”): any characterisation (e.g. by the addition of a qualificative adjective) applies to the compound lexeme as a whole.

In addition to the property of non-decompositionality, cohesion between the elements of a compound is reinforced by several of the following more specific parameters: usage (stability and fixedness), meaning (specificity or figurative meaning in the case of N-ADJ compounds), as well as (to some extent) speaker awareness (the recognition of such forms as units).

3 Semantic aspects of compounding

From a semantic point of view, compound meaning implicates the semantic content of its components in different ways. It is important to distinguish between three types:

- ‘endocentric’ or determinative compounds: one element of the compound determines the meaning of the other:

Fr. <i>gratte-ciel</i> “skyscraper”
It. <i>parola-chiave</i> “keyword”, <i>assicurazione malattia</i> “health insurance”
Cat. <i>peix-espada</i> “swordfish”

Sp. *paso mosca* “flyweight”

Pg. *edifício-garagem* “covered parking lot”

Rom.n *redactor-șef* “editor-in-chief”

In the Romance languages, the first element of a compound is generally determined by the second (determined – determiner), with Germanic languages displaying the opposite pattern, in which the second element is determined by the first (determiner – determined), e.g. *frog-man*, *key-word*, German *Frosch-mann*, *Schlüssel-wort*, etc.

- ‘exocentric’ compounds: the internal determinational relationship is the same as that of endocentric compounds, but the meaning of the compound is metonymic or metaphorical and cannot be deduced directly from its components:

Fr. *casques bleus* “peacekeeping soldiers”, *peau-rouge* (derogatory) “Native American”

- cumulative or ‘copulative’ compounds: the two components contribute equally to the meaning of the compound. When analysed in detail, this type proves to be rather variable, although it requires its two components to belong to the same lexical class:

Fr. *moissonneuse-batteuse* “combine-harvester”, *fille-mère* “unmarried mother”, *porte-fenêtre* “French window”, *Alsace-Lorraine*, *aigre-doux* “bittersweet, sweet and sour”

9.4.4 Other derivational mechanisms

Throughout the history of the Romance languages, a number of other, more specific mechanisms can be identified as following the principles of derivation. Depending on the historical period and the variety in question, these may be more or less widespread.

1 Back-formation or regressive derivation

This process involves the formation of morphologically simple words from derivatives. Consider the following examples:

Fr. *diplomate* N.M “diplomat” (1792) ← *diplomatie* N.F “diplomacy” (1790) ← *diplomatique* ADJ “diplomatic” (1721)

Sp. *auditar* v “audit” ← *auditor* N.M “auditor”; *extraditar* v “extradite” *extradición* N.F “extradition”

2 Affixoids

In this type of derivation, particularly typical of the 20th century, lexical morphemes that may appear as free lexemes are added to other word bases as if they were affixes. Common examples include:

auto- (Fr. *suggestion*, It. *suggestione*, Sp. *sugestión*); *super-*; *télé-* / *tele-*

The use of these forms as free lexemes (e.g. *c'est super*, *la télé*) usually follows a period of frequent use as affixes.

3 Reduplication

Reduplication is common in hypocoristics such as French *fi-fille* ("little girl", a term of endearment), *mé-mère* ("granny"), *jou-jou* ("little toy") or Italian *pian piano* ("slowly"); cf. also the proper names It. *Gigi* (< *Luigi*) or *Totò* (< *Antonio*). It brings about a certain linguistic expressivity or intensification, and is often characteristic of 'nursery language' as well as of borrowings from the latter into informal registers.

4 Derivatives without a base

An affixed form does not necessarily presuppose the existence of a lexical base that can stand freely without an affix. Suffixed forms such as French *mensu|el* "monthly" and *nupt|ial* "nuptial" were introduced as Latinisms, without their respective bases **mensu-* and **nupt-* ever having existed as such in French. Similarly, the original Latin roots of the French forms *bouv|ier* "cowherd", *pan|ier* "basket" or *pel|age* "coat, fur" (BOV-E "ox", PAN-E "bread" and PEL(L)-E "skin, hide") are no longer discernible, even though the suffixes are clearly recognisable (examples taken from Grevisse, § 166 b1; cf. 9.4.2).

In many cases, derived words represent borrowings from Latin, whereas their corresponding base word has been subject to processes of natural or semi-learnèd evolution:

Fr. *marin* "marine, relating to the sea" vs. *mer* "sea"

Fr. *populaire* "popular, relating to the people" vs. *peuple* "people" (semi-learnèd, cf. 9.4.2)

The relationship between free lexeme and derivational base (which are identical in Latin) was thus severed in the Romance languages. Sometimes, borrowings even replace the original base (e.g. in Fr. *oral* "relating to the mouth" vs. *bouche* "mouth").

Moreover, as has already been pointed out, the stem space of a lexeme may acquire new variants during the process of word formation:

in *bijou|t|ier* “jeweller”, *congo|l|ais* “Congolese, relating to the Congo” ou *vou|v|oyer* “address sb. using the *vous* form (i.e. formally)”, the original base – *bijou* “jewel”, *Congo*, *vous* – is extended (in order to avoid the occurrence of hiatus between two morphemes); in *canad|ien* “Canadian” and *mot|ard* “motorcyclist, biker”, the lexical bases – *Canada* and *moto* “motorcycle”, respectively – are shortened.

5 Derivation within compounds

The syntagmatic structure of the French term *conseil|l|er municipal* has an ambiguous semantic status: this compound does not, as one might expect, designate a type of counsellor who is municipal (in nature), but a “member of the *conseil municipal*” (“town council”). This contradictory aspect is strongly perceived by speakers and the use of this type of formation is therefore limited.

9.4.5 Other mechanisms of word formation

Derivation and compounding are the two most frequent mechanisms of word formation. There are, however, other processes, whose productivity, once again, depends on the historical period and the variety in question.

1 Forms of word shortening: clipping, acronyms (and initialisms)

Two relatively recent morphological mechanisms of word ‘shortening’ that have been highly productive since the second half of the 20th century are clipping (or ‘truncation’, illustrated by the French examples *métro[politain]*, *vélo[cipède]*, *prof[esseur]*, *impec[cable]*), and the formation of acronyms. Strictly speaking, acronyms (i.e. abbreviated forms that are pronounced as words) can be further distinguished from so-called ‘initialisms’ (abbreviated forms that are pronounced as a string of initials).

acronym:	Fr. C.A.P.E.S. /ka'pes/ ← <i>Certificat d'aptitude au professorat de l'enseignement supérieur</i> ; cf. its derivative <i>capésien</i> IT. ASL /az'le/ ← <i>Azienda sanitaria locale</i> ‘local health centre / doctors’ surgery’
initialisms:	Fr. C.G.T. ← <i>Confédération générale du travail</i> ; cf. its derivative <i>cégétiste</i> Fr. UFR /yɛfɛʁ/ ← <i>Unité de formation et de recherche</i> It. PC /pi'tʃi:/ ← <i>partito comunista</i> Rom.n C.F.R. /tʃe fe 're/ ← <i>Căile ferate române</i> “Romanian railways”; cf. its derivative <i>ceferist</i> “railway worker”

2 Delocutives

The term ‘delocutive’ is used for lexemes that are based on discourse fragments. The latter often carry strong pragmatic connotations and are capable of evoking particular

contextual settings. Although such formations may involve a number of different types of bases, the most frequent delocutives are based on verbs:

- Verbal: Fr. (*un*) *je-m'en-foutisme* "uncaring attitude", It. *menefreghismo* "id."
 Fr. (*un*) *m'as-tu-vu* "(a) show-off", *va-et-vient* "coming and going", Sp. *vaivén* N.M "id.", Rom.n *du-te-vino* "id.", literally 'go away come'
 Sp. *recibí* "receipt" N.M (< *recibir* P1 PERF)
- Pronominal: Fr. *vouvoyer* (cf. 9.4.4 no. 4), Sp. *vosear*
- Suffixal: Sp. *tico* ADJ "familiar language in Costa Rica" ← *-ico* (a highly productive diminutive suffix in Costa Rica)
- Discursive (formed from elements at the level of discourse), e.g. salutation formulas:
 Fr. *adieu* N.M, Sp. *adiós* N M, It. *addio* N.M
- Cf. also the following Latinisms: Fr. *factotum*, Sp. *factótum*; Fr. *fac-similé*, Sp. *facsimil(e)*

3 Blending

Blending forms new lexemes from elements belonging to two (or more) different words. Although this type of word (trans)formation has a longer history than some of the others, it remains relatively rare, mainly occurring between words that display similar formal and semantic properties (cf. 9.8), as illustrated by the following French examples:

- the noun *trouble* "disorder, confusion" seems to be a blend of Lat. *TURBIDUS* "confused" and *TURBULENTUS* "full of commotion" (which would have resulted in **TURBULUS*);
- the verb *craindre* "be afraid (of)" is arguably the result of blending between Lat. **TREMULARE* (← *TREMERE* V "tremble") and a semantically cognate Celtic verb (cf. Breton *kriedien* "the act of trembling").

Word blends can be compared to the phenomenon of folk etymology (consider the popular French phraseologism *il faut mieux*, which is a reinterpretation of *il vaut mieux* "it is better (to ...)").

Blending may also occur consciously and intentionally, as in the case of so-called 'portmanteau words', widespread since the 20th century: consider Fr. *auto|bus* "bus", *fran|glais* "Franglais, i.e. a mixture of French and English", *resto|route* "(roadside) rest area" or *mo|tel* (a blend of *motor* and *hotel* borrowed from English).

4 Intentional formal distortion

Finally, all languages display phenomena of formal 'distortion' or modification. The processes can be conscious to a greater or lesser extent and often serve the purpose of achieving expressive or creative effects. Classic examples include the creation of euphemisms, for so-called *nomina sacra*, for instance (as in Fr. *parbleu* ! (← *pardieu* ← *par*

Dieu) or English *oh my gosh!* (← *oh my God!*), to avoid pronouncing the name of God as a swear word.

The generalisation of this mechanism has resulted in the French phenomenon of ‘verlan’ (← (*à*) *l’envers* /lɑ̃ˈvɛʁ/), a form of jargon in France since the mid-19th century. As its name suggests, verlan makes use of syllable inversion in order to create new words:

ripou /ʁiˈpu/ ← *pourri* /puˈʁi/ “rotten”

laisse béton /beˈtɔ̃/ ← *laisse tomber* /tɔ̃ˈbe/ “let it go”

meuf /møf/ ← *femme* /fam/ “woman” (with the addition of the word-final vowel)

rebeu /ʁəˈbø/ ← *beur* /bøʁ/ ← *arabe* /aˈrab/ “Arab(ic)” (id.)

→ Pharies, *The playful lexicon in the Romance languages: prosodic templates, onomatopoeia, reduplication, clipping, blending*, OxfEnc

9.4.6 The significance of diachrony in word formation

The ambiguous status of word formation (cf. 9.4.1) becomes particularly apparent when one considers the fact that complex words formed during different periods of history coexist in synchrony. Such lexemes have a variable status with regard to speaker awareness. Different cases can be distinguished:

- complex words of early formation whose composite nature is no longer recognisable in synchrony; e.g. Fr. *soleil* < SOL + -ICULU;
- complex words of early formation whose composite nature may be recognised as such, but which can no longer be easily associated with a base form or meaning; e.g. Fr. *vis-age* or *pan-ier*, in which the suffix, but not the base, can be identified in synchrony (cf. 9.4.4 no. 4);
- complex words of early formation that can be easily identified as such, but whose meaning diverges from that of the original word; e.g. in today’s French, a *luth-ier* is someone who repairs violins, not lutes; as another example, the verbs *engueuler* “to tell off” and *dégueuler* “to throw up” – denominal verbs derived from *gueule* ‘throat; mouth’ – no longer form a pair of antonyms;
- complex words of early formation that can be identified as such and that are semantically close to their base, but that rely on mechanisms of word formation that are no longer productive today; e.g. the intermediate case of the suffix *-ment*, which in Modern French can still be used to form adverbs, but is subject to more restrictions today than it was in the 16th century; hence, its productivity is reduced, even though forms ending in *-ment* are still frequently used;
- complex words that can be identified as such, are semantically close to their base and rely on mechanisms of word formation that remain productive; furthermore, the time of their formation has no effect on their semantic transparency: *mari-age*

[ca. 1135], *lang-age* [ca. 1160] and *pass-age* [ca. 1165] are far older than *régl-age* [ca. 1506] or *bizut-age* “(ritual) initiation (of students)” [1949]; however, they all remain perfectly transparent.

These different categories are situated on a continuum between a ‘lexical’ extreme on the one hand (represented by forms such as the autonomous word *soleil*) and a ‘grammatical’ extreme on the other (including forms such as the couple *régler* – *réglage*). Constructions such as *soleil* are considered to be opaque (i.e. unrecognisable) and fixed, and they must be memorised individually within the mental lexis, whereas forms such as *réglage* remain productive and can consequently be processed as composite forms on the basis of their individual components.

For approximately the past fifteen years, neuropsychological research has proposed a form of ‘double’ processing for complex words that are semantically transparent and are based on mechanisms of formation that remain productive today: during the process of decoding, the brain calls up images from the whole form memory and, at the same time, attempts to segment these forms. Depending on the frequency of the derivative in individual usage, one of the two processes proves to be faster. For production (encoding), in contrast, it is more reasonable to suppose that derivatives are stored in their entirety in the lexical memory.

At the moment of formation of new words, the mechanisms of semantic construction involved are always transparent, as is the process of semantic change itself (cf. 9.4.3). As language evolves, however, the semantic and formal links between a base and a corresponding derived form are weakened or lost entirely. This may be due to various factors:

- semantic change affecting the derived word: e.g. Fr. *beauté* “quality of being beautiful; quality of a beautiful person” → “beautiful woman”;
- change affecting the (external) context of the referent, as in the example of *luthier*, which bears witness to a change in musical taste and instruments;
- the disappearance of base words: e.g. Fr. *liesse* (literary) “exuberant collective joy” appears isolated due to the loss of O.Fr. *lié* “joyous, sprightly” < LAETU.

Moreover, polysemy and the low specificity of affixes, are factors which sometimes hinder the establishment of a direct link between a base word and its derivative. Consider the following example (cf. Glessgen, *Le statut épistémologique du lexème*, 2011: 409 sq.): the electronic version of the French dictionary *Trésor de la langue française* (TLFi) assembles over 3,000 derivatives ending in *-iste*; these are denominal or deadjectival nouns that display a parallel semantic structure, namely “an X (person) who is closely related to a Y (non-personal entity)”. When considered in detail, semantic variation among these terms is nevertheless impressive, as shown by the examples belonging to the subgroup of names of professions:

the French term *journal-iste* (← *journal*, “newspaper”) does not, as its structure may suggest, refer to a “person who sells newspapers”, but to a “reporter who works for the mass media (and not even necessarily the written media, if documentaries are taken into account)”;

a *paysag-iste* (← *paysage*, “landscape”) is either a “landscape painter” or a “landscape architect” (two distinct meanings), whereas a *visagiste* (← *visage* “face”) is a “beautician specialised in enhancing facial features” and not a “portrait painter”;

a *pian-iste* is a “(semi-)professional piano player” and not “someone who builds, repairs or tunes pianos”.

Complete transparency of a complex word presupposes that its exact meaning (‘word meaning’) can be derived directly from the meaning of the base form and from ‘preprogrammed’ patterns of word formation (‘word formation meaning’, cf. 9.4.1). Examples show that such cases are, however, far from typical. Detailed observation thus highlights the difficulties involved in identifying the exact manner in which a new form was created and the manner in which its meaning was constructed at the period in question. This is all the more true for the study of forms belonging to the older stages of a language.

In this chapter, it has been repeatedly shown that the vitality of different mechanisms of word formation is unstable. The same applies at a macroscopic level for different language varieties: specific forms of ‘peripheral’ French, such as the urban *français populaire* or the French spoken in sub-Saharan Africa are currently very productive as regards derivation (and semantic innovation), to a much greater extent than standard or informal French in Europe. The same phenomenon can be observed with regard to Spanish in the Americas, which is much more innovative today than Peninsular Spanish (the opposite was true between the 17th and the 19th centuries).

As a final remark, the major mechanisms of derivation and compounding in the Romance languages are relatively easy to interpret on the surface; however, both the phenomena and their interpretation are considerably more complex when analysed in depth (cf. the observations made by Franz Rainer, *Spanische Wortbildungslehre*, 1993 or Maria Grossmann and Franz Rainer, *La formazione delle parole in italiano*, 2004). The co-occurrence of elements that are to a greater or lesser extent transparent with those that have become opaque, for instance, leads to difficulties in interpreting the global function of word formation in a given language.

With regard to the Romance languages, the study of morphology is further complicated by the presence of numerous Latin borrowings from all periods of history, involving both word bases and affixes, as well as by the presence of Latin affixes that remain recognisable and productive in the Romance period (cf. the examples provided in 9.4.2 no. 2.2).

At the same time, the opportunity of comparing divergent and convergent processes of derivational evolution in multiple Romance languages offers the potential for rich insights into the nature of phenomena of word formation.

9.5 Syntagmatic context: valency, collocation and phraseology

Phraseologisms are halfway between free syntagms and syntagmatic compounds (*cf.* 9.4.3 no. 2). A phraseologism is a grouping of several orthographic words that constitute a recognisable semantic unit, such as French *avoir soif* “be thirsty”, (*ne pas être*) *franc du collier* “(not) be sincere”, or English (*through*) *thick and thin*, *bite the dust*.

Such groupings are heterogeneous and by definition more variable than compounds. This complicates their linguistic description. Difficulties in the description of phraseologisms are already manifest in the multiplicity of terms used to designate them: “multi-word units”, “syntagmatic compounds”, “fixed expressions”, “idiomatic expressions”, “set phrases”, “idioms”, “clichés”, “phrasemes” or “phraseologisms” (this last being the term adopted here).

Phraseologisms are one example of the preferential relationships that exist among lexemes, together with verb valency and collocations. The syntagmatic associations established in this way often become habitual, thus determining the concrete use of a language beyond its syntactic rules and lexical meaning. It is at this level that Coseriu’s ‘usage norms’ come into play (*cf.* 1.2.3).

The links between lexemes that appear in combination with each other may be more or less close, and one may suppose that the level of interaction among them forms a continuum of increasing intensity. On this continuum, we have distinguished three categories that are neither bounded nor uniform: verb valency, collocations and phraseologisms. Compounds displaying the highest possible degree of fixedness may be seen as representing a fourth category.

9.5.1 Valency

Verb valency represents a first level of lexicalised syntagmatic links, as valency frames partially determine sentence construction. The number of arguments admitted by verbs and the prepositions that they govern are restrictions of a syntactic nature, but they may at the same time be linked to individual verbs (*cf.* 8.2.3 and the examples provided):

Fr. *jouer de la flûte* (and not **jouer sur la flûte*), *jouer au foot*

This first level thus involves both syntax and lexis.

9.5.2 Collocations

Collocations are of a more lexical nature, corresponding to quite specific associations between given lexemes in usage:

we say *the water is just right*, rather than *agreeable* or *ideal*, as in French *l'eau est bonne* rather than *agréable* or *à température idéale*, even though these alternative solutions would be equally as coherent from a semantic, syntactic or diasystematic point of view;

cf. also Fr. *faire le naïf* rather than *jouer le naïf* “act innocent”, or the above-mentioned example of *prononcer un discours* instead of *dire, énoncer* or *présenter un discours* (9.1.1 (3)).

Though the use of these words in a given syntagmatic context is commonplace, they nevertheless maintain the autonomy of being able to appear in constructions with other terms. Collocations are characterised above all by their frequency of use and degree of fixedness; such examples of preferential use, however, are not necessarily indicative of strong semantic cohesiveness; moreover, the meaning of a collocation is directly comprehensible on the basis of its components.

9.5.3 Phraseologisms and idioms

Phraseologisms lead to very strongly established lexicalised connections between different lexemes and they display some degree of stability. Such connections may involve different parts of speech (e.g. Fr. *à la bonne heure!* “that’s the spirit!”, *en mourir d’envie* “crave something”). They typically display:

- semantic cohesiveness that can be recognised, if not identified, through substitution tests (e.g. Fr. *à l’intérieur* = *dans*) or through language comparison (e.g. Fr. *bonne chance!* = Sp. *¡suerte!*);
- a certain degree of stability or fixedness: even though phraseologisms often leave room for some variation, this is perceived as unusual by speakers; variation within phraseologisms consequently has stylistic implications, which are exploited in journalistic writing in particular.

Phraseological or syntagmatic models such as French *aide à (...)*, which can be freely reproduced (*aide à l’emploi*, *aide aux personnes âgées*), constitute a particular type of phraseologism.

Phraseologisms may maintain the initial meaning of their components (such as in *à l’intérieur de*), but they more often take on a figurative meaning (one then speaks of ‘idioms’ or ‘idiomatic expressions’). Through linguistic change, they also often diverge from typical syntactic constructions. Consequently, the meaning of these expressions

as a whole may be difficult to deduce from their components alone (consider Fr. *tout à l'heure* “right away” or *le jeu n'en vaut pas la chandelle* “it’s not worth it”, as in English *the game isn’t worth the candle*). In these cases, phraseologisms have the complementary characteristic of being semantically opaque. This parameter is diachronic: at the time when a phraseologism is created, it is transparent to speakers; once fixed, however, the initial motivation for its creation may disappear, either from the referential sphere or from the awareness of speakers. In the case of the example *le jeu n'en vaut pas la chandelle* provided above, lighting in modern theatres is provided by electricity instead of candles (Fr. *chandelles*); today, moreover, plays are no longer called *jeux* in French. This obscures the origins of the idiom: *le jeu* (“the play”) *n'en vaut pas* (“is not worth”) *la chandelle* (“the candles”).

We have seen the same phenomenon in all domains of semantic and formal change, which always display an initial motivation. Therefore, the transparency of the motivation for a phraseologism or the absence thereof simply reflects the vagaries of linguistic history.

As previously mentioned, linguistic terminology varies greatly in this field: what is here referred to as a ‘phraseologism’ may be called a ‘collocation’ elsewhere; others distinguish ‘phraseologisms’ (as semantically transparent expressions) from ‘idioms’ (as syntagms with a figurative meaning), grouping the two under the more general term ‘phrasemes’.

Even the distinction between syntagmatic compounds (cf. 9.4.3 no. 3.2) and certain types of phrasemes is not clear-cut and is more relative than absolute; there is a large transitional zone between the two. Like compounds, phrasemes form complex lexical units displaying a certain degree of stability and unity of meaning, but with somewhat more syntactic and semantic flexibility than compounds.

9.5.4 Observations

In addition to these syntactico-semantic restrictions, there are more general semantic restrictions, such as those inhibiting the mixing of abstract and concrete terms. These are non-specific (stylistic-semantic) constraints, which nonetheless, like phraseologisms, operate at the level of syntagmatic interactions.

The multiple interactive relationships that are established between words are important for the structuring of discourse. The majority of lexemes are used in a relatively limited number of habitual contexts. In order to be convinced of this, one need only search for a specific word in one of the major databases such as *Frantext* or *CREA* (*Corpus de referencia del español actual*). This aspect of lexis is surprisingly poorly perceived by our instinctive linguistic awareness: we are confused by unusual collocations and phraseologisms but we do not immediately recognise them as such, in contrast to other divergent phenomena of lexical usage (e.g. borrowings or new meanings).

These common usage restrictions make the learning and reproduction of foreign languages difficult. They share much of the ‘strangeness’ effect that texts from earlier historical periods have on the modern reader. In language evolution, collocations generally display a faster rate of innovation than the individual forms and meanings themselves. Their flexibility can therefore make them reliable indicators of cultural trends.

This field has been very poorly studied from a diachronic point of view to date, even though there are numerous repertoires of idiomatic expressions (*cf.* Ziltener, *Repertorium der Gleichnisse und bildhaften Vergleiche*, 1972). Comparative and historical studies with a truly linguistic dimension are quite rare (*cf.* e.g. Schweickard, “Il se croit de la côte d’Adam”, 1990 or various articles by Roques, such as “Parler d’autre Martin”, 1999).

9.6 Lexical borrowing in the Romance languages

9.6.1 Definition

The borrowing of lexemes from one language into another is the fourth major mechanism by which the vocabulary may be transformed and enlarged, in addition to semantic change, word formation and phraseology. Through borrowing, languages introduce new lexemes into usage, creating new forms (in the case of formal borrowing) or new meanings (in the case of semantic loans, *cf.* 9.6.3 no. 2).

It is important to consider that every borrowing is preceded and motivated by an underlying extralinguistic cultural model. Borrowing does not occur for language-internal reasons: the extensive introduction of English terms from the domains of computer science and rock and pop music into today’s Romance languages is not due to the ‘beauty’ of these words or their particular functional value, but arises from the prestige of the Anglo-American world in the areas of economy, technology, culture and the military.

Consequently, every instance of cultural contact leaves traces in the form of borrowings in the languages involved. Borrowings are thus simultaneously the effect and the visible proof of socio-cultural influence. Lexis is only one specific domain in which linguistic borrowing and interference occur; these relate more generally to the phenomenon of language contact (*cf.* 3.6). Apart from the extreme case of mixed languages (*cf.* 3.6.2), lexical borrowing is nevertheless much more common than grammatical or phonetic borrowing (which are explained effectively by Gardani, *Contact and borrowing*, CambrHandb 28).

A further observation that can be made is that the external motivation for lexical borrowing is more transparent and precise than it is for semantic change and derivation. Lexical borrowing consequently allows a more detailed study of the effects of external or extralinguistic causes of language-internal transformations.

9.6.2 Contact languages and borrowing in the Romance languages

Numerous languages came into contact with Romance languages, and for the most diverse historical reasons (*cf.* also chapter 10). This area is relatively easy to delimit and has been excellently studied, as shown by the following contributions, which provide examples of borrowing as well as references to further research:

RSG 1, sect. V, art. 48–60	<i>Préhistoire et formation des langues romanes</i>
RSG 2, sect. XII, art. 134–169	<i>Contacts linguistique et migration</i>
LRL 7	<i>Langues en contact, langues des migrants et langues artificielles</i>
HSK 12.2, art. 142–157	<i>Language contacts in Europe [...] – France; Spain-Portugal; Italy-Malte</i>
<i>ibid.</i> , art. 174–176	Romania
CambrHist 2, 6	<i>Contact and borrowing</i>
OxfEnc	<i>Greek in contact with Romance, Romance in contact with Basque, Romance in contact with Semitic, Contact between Spanish and Portuguese, Spanish in contact with South-American languages, with special emphasis on Andean and Paraguayan Spanish, etc.</i>

Volumes 15–20 of the FEW, which are devoted to borrowings, contain a great deal of primary data; for Germanisms, the LEI should also be consulted (*cf.* 9.9.2 no. 3).

Here, a synthetic overview of only the most salient categories of lexical borrowing will be presented (for other situations of language contact, *cf.* 3.6 and 10.4.2).

By far the most frequent borrowings in the Romance languages come from written Latin, which remained the contact language par excellence, except in the case of Romanian (*cf.* 6.6 as well as 9.6.4), for the whole of the Middle Ages and the modern period. In the following sections, the most important contact languages in the history of the Romance languages are presented in chronological order.

1 Languages in contact with Latin

The languages in contact with Latin were spoken by indigenous peoples living within the Roman Empire, before they were suppressed and obliterated by Latin (we speak of linguistic *substrata*, *cf.* 10.2.2). During a phase of at least partial bilingualism, Latin absorbed a number of lexemes from these languages, some of which survived into the Romance languages (sometimes even without leaving any written evidence in Latin). In addition to a limited number of lexemes, numerous toponyms survive (*cf.* 9.7.3). The most important *substrata* are the following:

- Celtic in France left about 250 lexemes (with numerous derivatives) in the different Gallo-Romance varieties, in addition to numerous toponyms (*cf.* FEW).
- Celto-Iberic appears essentially in toponyms of the Iberian Peninsula (Iber. *ili-* “town” is present in place names such as Sp. *llerda* > *Lérida*).

- Etruscan in central Italy survives in some family names and in the names of the inhabitants of a specific locality (so-called ‘gentilics’ or ‘demonyms’).
- Osco-Umbrian produced borrowings such as Lat. *gumia* s.m. “gourmand” (< Umbr. *kumiaf* “loaded, full”), which has been maintained in Sp. *gomia* “scarecrow”.
- Semitic languages, such as Punic or Phoenician, produced some rare Latin terms such as *barca* “boat” or *tunica* “tunic”.
- Greek had a regional impact in southern Italy, as did – to an even greater extent – Basque in south-western France.

By far the most important contact language for Latin was Greek, the second language of the bilingual Empire. The numerous borrowings are nevertheless the effect of cultural contact through the medium of a limited number of cultivated persons rather than the result of everyday contact between populations (as was shown to be the case for the other linguistic *substrata*). More particularly, the Greeks played an essential role as teachers in the Latin-speaking parts of the Empire.

Much later, in the 19th–20th centuries, these Greco-Latin terms assumed particular importance in scientific terminology, which is based on this heritage (cf. 11.1 and 11.1.3). Consider the example of terms used widely throughout the present manual, such as *phonology*, *morphology*, *syntax* or *lexis* or, more specifically, *phylogeny* “the evolution of species” and *monogenesis* “common origins” (cf. 1.2.5).

2 Germanic languages

The Germanic languages were mainly in contact with Late Latin and the emerging Romance languages following Germanic invasions around the 5th century. They influenced the majority of the Romance languages, in particular the dialects of north-eastern Gallo-Romance and northern Italy. These were linguistic ‘*superstrata*’, which arrived after Latin had become established, but did not replace it (cf. 10.3.3 on the influence of the Germanic *superstrata*, which still today is often overestimated in terms of its impact on the evolution of the Romance languages). Germanic borrowings gave rise to several thousand lexemes in the Gallo- and Italo-Romance dialects, which are concentrated above all in the semantic fields of military and court life:

Old Frankish **werra* “battle” > Fr. *guerre*

Germ. **baro* “free man serving in the army” > Fr. *baron* “brave, valiant man”, then “high-ranking nobleman”

Old Frankish **siniskalk* “maître d’hôtel, head steward” > Fr. *sénéchal* “seneschal, justice official, royal agent”

Old Frankish **sparwari* “sparrow hawk” > Fr. *épervier* > It. *sperviere*

Old Low Frankish **waiðanjan* “(take livestock to) graze” > Fr. *gagner* “win, gain” and It. *guadagnare* (cf. 9.3.7 no. 2)

Germ. **manigipō* “great quantity” > Rom. *mintga* (cf. 3.5.2)

3 Arabic

Arabic was particularly important as an *adstratum* language in the Iberian Peninsula between the 8th and 13th centuries (until 1492 in the Caliphate of Grenada). More recently, it has again played the role of a contact language following present-day colonisation and migration (cf. 10.3.3).

Several hundred borrowings found their way into the Romance languages through the medium of Spanish, and to a lesser degree, Italian, during the medieval period, and still survive today. These borrowings concern urban and rural life in particular, as well as the world of spices or medieval sciences:

Ar. *az-za'farān* "saffron" > Sp. *azafrán*; without the article: *za'farān* > It. *zafferano* (cf. 3.6.1)

Ar. (*as*-)*samt* (*ar-ra's*) "(the) zenith" > Sp. *cenit* (cf. *ibid.*)

Ar. *mustá'rib* "person who adapts themselves to Arabic culture" > Sp. *mozárabe* (> Fr. *mozarabe*; cf. 10.4.1 no. 1)

Recent borrowings are especially frequent in French; cf. the following examples:

Maghreb Ar. *b(a)lad* > Fr. *bled* "small isolated villages without resources"

Ar. *ḥāšīš* > Fr. *haschisch*

Ar. *ṭabīb* > Fr. *toubib* "doctor"

4 Slavonic languages

The South Slavonic languages strongly influenced Romanian from the 9th century onwards. Their impact concerns above all Romanian vocabulary, a significant part of which is of Slavonic origin, but also grammar, albeit more sporadically (cf. 3.4.13 and 10.3.2):

the Slavism *vreme* is synonymous with the inherited Latin term *timp* "weather, time"

cf. other Slavonic borrowings: *a iubi* "love", *a citi* "read"

Borrowings from Russian (especially in the 19th and 20th centuries), in contrast, were relatively uncommon; cf. e.g. *bolchevik* > Rom.n *bolşevic*, Fr. *bolchevik*.

5 English and other languages

Particularly since the end of World War II, English has supplied all Romance languages with numerous borrowings, such as Fr. *jazz*, *boomerang* (the latter a word of Australian aboriginal origin), *baseball*, *computer*. We will return to this subject in chapter 9.6.3 below.

Other languages or language groups have had a more sporadic influence on Romance (cf. ch. 10):

- Dutch was present to some extent in France under the Ancien Régime (i.e. during the Renaissance and up until the French Revolution) in maritime vocabulary, etc.;
- Germanic dialects had a strong impact in the north-eastern border zones of France and in the Tyrol;
- Native American languages had some influence on Spanish in America during the colonial period (e.g. Sp. *canoa* “canoe” < Cariban *kanawa*, *tomate* “tomato” < Nahuatl *tomatl*, *hamaca* “hammock” < Taíno (Haiti) *hamaca*);
- Asian and Romani languages are represented sporadically in the Romance languages by sporadic borrowings.

6 Contact among different Romance languages

Finally, intense ‘Romance-internal’ contacts developed at several stages throughout history. Occitan was in contact with Catalan in the Middle Ages (cf. 10.4.1 no. 3); it also generated a significant number of borrowings in French from the 16th century onwards, which were transmitted through the medium of southern regional French. Consider the following examples:

Occ. *abelha* > Fr. *abeille* (cf. 4.2.2)

Occ. *bastida* > Fr. *bastide* (the Occitan term, in turn, is of Germanic origin: **bastjan*, cf. 9.3.7 no. 2; for the phenomenon of ‘distant etymology’, cf. 9.6.3 no. 3 below).

Italian influenced French and the Iberian languages above all in the period between the 14th and 16th centuries (cf. 10.5.1). Borrowings are concentrated in the vocabulary of the arts and architecture, musical, military and banking terminology:

It. *baldacchina* “silk fabric from Bagdad” → “canopy with curtains, which is placed over a bed, throne, etc.” > Fr. *baldaquin*

It. (*dipingere a*) *fresco* “(paint on a) fresh coating” → “mural painted on a coating of fresh mortar, using colours diluted with water” > Fr. *fresque* “fresco”

It. *adagio* “at one’s ease, leisurely” → (music.) “indication of slow movement” > Fr. *adagio* “id.”

As we have seen, Spanish was a vehicle for Arabisms and Native Americanisms in particular; nevertheless, it is also the language of origin of a number of direct borrowings in the 18th century (Sp. *siesta* > Fr. *sieste*, Sp. *zarzaparrilla* “spiny shrub” > Fr. *salsepareille* “sarsaparilla, greenbrier”). Since the 16th century, it has had a strong influence on Galician and – to a lesser extent – on Catalan (cf. 10.5.1).

French has sporadically influenced other Romance languages since the Middle Ages; cf. the examples of Gallicisms in Italian courtly and lyric language, such as the above-mentioned *sperviere* > *épervier* (itself of Germanic origin, cf. no. 2 above). The influence of French in Europe reached its peak in the 18th and 19th centuries; since the end of the 19th century, it has significantly contributed to the development of the vocabulary of Romanian (cf. 10.5.1). The impact of French remains important even in the 21st

century, on Spanish in particular (cf. the exemplary work by Curell, *Diccionario de galicismos del español peninsular contemporáneo*, 2009).

9.6.3 Typology of lexical borrowing

1 Formal borrowings

The most easily identifiable and most common type of borrowing is the formal integration of words from a source language into a target language. The degree of integration may vary depending on the linguistic domains involved. The borrowed form may remain close to that of the source language or it may be adapted to the rules of the target language to a greater or lesser degree.

The following are examples of Anglicisms that represent relatively recent borrowings (appearing from the 18th century onwards) but which are currently widely used in the Romance languages. The typology used here follows Schweickard (*Englisch und Romanisch*, LRL 7, art. 470; for the dates at which they were borrowed into French, cf. Höfler, *Dictionnaire des anglicismes*, 1982).

Phonetic adaptation:

- In general, borrowings are integrated into the pronunciation system of the target language, e.g. *jungle* [ʒœ̃gl] < Engl. *jungle* ['dʒʌŋɡəl] (since 1796, FEW 18, 76a). Variation is nevertheless frequent, particularly in the case of recent borrowings:

Fr. *pipeline* (2nd half of the 19th century) has a French pronunciation [pi'plin] and a second that is closer to the language of origin: ['pajplajn];

cf. also It. *giambo* ['dʒumbo] vs. ['dʒambo] (< Engl. *jumbo*).

Graphic adaptation or adaptation to rules governing graphic-phonetic relationships:

- Adaptation may lead to graphic transformations that simultaneously imply considerable phonetic changes:

Engl. *bowling green* > Fr. *boulingrin* (1663, FEW 18, 34b);

graphic variation is frequent, e.g. *cornère* vs. *corner*, It. *bleffare* vs. *bluffare*, Sp. *bulldog* vs. *bulldog*, Pg. *hóquei* vs. *hockey*.

Inflectional morphological adaptation:

- Inflectional adaptation is inevitable for verbs:

Fr. *boycotter* (since 1880) vs. Engl. *to boycott*

It. *sprintare* vs. Engl. *to sprint*

Sp. *blufear* vs. Engl. *to bluff*

Pg. *lanchar* vs. Engl. *to lunch*

Rom.n *a dribla* vs. Engl. *to dribble*

- For nouns, gender marking is also indispensable; the gender of a term may differ between source and target language. This is apparent in borrowings from English:

Fr. <i>la star</i> “star” (since 1919)	after <i>la vedette</i>
It. <i>la holding</i> “holding company”	after <i>la società</i>
Sp. <i>la jet society</i>	after <i>la sociedad</i>

- Number marking may remain incomplete:

Fr. <i>les boxes</i>	vs. <i>les box</i> (since the end of the 18 th century)
It. <i>i fans</i>	vs. <i>i bar</i>
Sp. <i>los suéters</i>	vs. <i>los mítines</i>

Derivational morphological adaptation:

- Derivational affixes may remain close to the source language or they may be adapted to the typical affixes of the target language:

Fr. <i>agnostique</i> (since 1884)	< Engl. <i>agnostic</i> (the common Fr. suffix <i>-(i)que</i> has been adopted)
Fr. <i>indésirable</i>	< Engl. <i>undesirable</i> (the Fr. prefix <i>in-</i> has been adopted)
It. <i>campeggio</i>	< Engl. <i>camping</i>
Fr. <i>possiblement</i>	< Engl. <i>possibly</i>

Note the existence of ‘false borrowings’, including words such as Fr. *parking* (since 1926) vs. English *car park* (U.S. *parking lot*) or *recordman* vs. Engl. *record holder*, Rom.n *tenisman* (with nativised plural morphology: *tenismeni*) vs. Engl. *tennis player*

Derivation and composition:

- The formation of new derivatives from recent borrowings is infrequent; this process typically denotes an advanced stage of integration of the borrowings into the target language.

Fr. <i>hockey</i> (since 1876)	→ <i>hockeyeur</i>
Fr. <i>franchise</i>	→ <i>franchiser</i>
It. <i>bar</i>	→ <i>barista</i>
Sp. <i>film</i>	→ <i>filmista</i>
(Engl. <i>knife</i> >) Pg. <i>naifa</i>	→ <i>naifada</i>

Semantic change:

Borrowings generally only preserve part of the meaning of the original form; consider Fr. *gay* “homosexual” (first attested in 1952), borrowed from the highly polysemic English word *gay*. Similarly to the formation of derivatives on the basis of borrowings, new meanings develop only with time. Therefore, the early borrowings from Germanic languages during the 5th to 8th centuries often developed complex meanings; e.g. O.Fr. *garnir*, which not only signifies “furnish, equip” (cf. English *garnish*), but also “prepare”, “warn” and “instruct”, for example.

Formation of phraseologisms:

The incorporation of borrowings into new phraseologisms is indicative of a high degree of integration into the new language, as is their involvement in processes of derivation or semantic change; consider Fr. *être publié sur le web* “be published on the web”.

2 Semantic and structural borrowing (‘calque’)

A second type of borrowing parallel to formal borrowing, though less visible, is semantic borrowing or ‘calque’. In this process, a form from the target language receives a new meaning which comes from the model provided by the source language. In contrast, there is no formal innovation in the target language.

There are two fundamental types of semantic borrowing, depending on whether the process concerns single words or compounds and phraseologisms. In the first case, a new meaning is given to a lexical form that shares (at least) one other meaning with a form in the source language; polysemy is thus transposed into the target language:

Fr. *arrière* after Engl. *back* meaning “player in defense position”

Sp. Fr. *cumbre* “political summit” (besides the original meaning “top (of a mountain)”) after *summit*

Close formal proximity between the terms of the source language and target language is often involved in this type of formation:

Fr. *approche* “manner of approaching a subject” after Engl. *approach*

Sp. Fr. *crucial* “decisive” after Engl. *crucial*

Fr. *réaliser*, It. *realizzare*, Rom.n *a realiza* “become aware of” after Engl. *to realise*

In structural borrowing, the individual elements of a compound or phraseologism are reproduced word for word according to the source language and, if necessary, reorganised according to the syntactic structure of the target language:

Fr. *gratte-ciel* after *sky-scraper* (with the determined and determiner in inverted order)

It. *ragazza-squillo* after *call-girl* (*id.*)

Fr. *faire une recherche en ligne* after *online*

Semantic or structural borrowings are evidently less easily perceived by speakers than formal borrowings; consequently, they do not evoke the same value judgements as borrowed foreign words.

3 Observations

Linguistic interference produced by lexical borrowing may have a macroscopic impact on semantic networks in cases where an entire semantic field is restructured (as illustrated by the sports terms exemplified above). It may also lead to transformations in the

preferential mechanisms of word formation (as in the case of the Romance suffixes that correspond to Engl. *-ise* / *-isation* / *-ism* / *-ist*); in exceptional cases, they may even lead to borrowings of a syntagmatic or syntactic type (e.g. Sp. *No le había visto por un año* vs. *No le había visto desde hacía un año*, on the model of Engl. *I hadn't seen him for a year*).

Some borrowings introduce concepts previously unknown in the target language in question; others replace existing words (*cf.* the Gallicism *Trottoir* instead of *Gehsteig* “pavement” in German-speaking Switzerland). Borrowings that fill a semantic gap have been called ‘necessity borrowing’ and redundant borrowings, ‘luxury borrowing’. From a stylistic point of view, however, every borrowing has a specific function for the speakers of the language into which it is adopted; it represents a particular intention and an expressive need.

The evaluative designations (‘necessity’ vs. ‘luxury’) are indicative of the degree to which borrowings evoke emotional reactions. Owing to their strong connotations and emotional charge, they easily lead to controversy or political propaganda and fuel the discussions that constitute linguistic culture (consider the example of modern Anglicisms in French). The example of borrowings also shows how subjective linguistics itself may become and how easily it can acquire a political dimension.

The route of borrowings from one language into another and from there into further languages is not always easy to reconstruct and it is important not to take shortcuts in their description. As an example, most medieval Arabisms in French were borrowed from Spanish, with a few coming from Italian, while many Anglicisms in Modern Spanish and Romanian were borrowed from French. Some words even passed back and forth between languages:

Fr. (*auto-*)*car* comes from the homonymous English word, which, in turn, is a borrowing from the Norman dialectal form *car* (without palatalisation of word-initial /k/, in contrast to the Fr. inherited term *char* < CARRUS)

cf. also O.Fr. *de(s)port* “amusement” > Engl. *sport* > Fr. *sport* and O.Fr. *tenez* > Engl. *tennis* > Fr. *tennis*

In these cases, the direct (or immediate) etymology (*etimologia prossima*) of a term is different from the distant etymology (*etimologia remota*). In cases where speakers are unaware of the distant origin of a loanword, the *etimologia remota* is of purely encyclopedic interest.

9.6.4 Quantitative importance of borrowing in the Romance languages

In comparison with vocabulary produced by means of semantic change and derivation, the contribution made by borrowings to the overall vocabulary of a language is very limited and thus also easier to identify. In the Romance languages, lexical, phonetic and

semantic Latinisms, however, form a category in their own right, as Latin was long considered to be a prestige variety within the same communicative continuum, rather than a foreign language (cf. 6.6). Throughout the history of the Romance languages, Latin has been exploited in the form of derivational affixes, as well as for the formation of numerous internationalisms belonging to modern technical language. Borrowing from Latin into Romance can even be considered as intralinguistic borrowing (i.e. borrowing within the same language).

Borrowings into Romance languages from other sources, in contrast, are restricted in number, as is characteristic for this means of innovation. The statistical relationships that result from attempts at quantification give a clear picture of the contribution of inter- and intralinguistic borrowings into French according to Gougenheim *et al.* (1956: 63 *sqq.*) and Wolf (1991: 50; 182). The result of 1,300 years of linguistic contact may be summarised as follows:

- of the 100 words currently most frequently used, none are loanwords: 92 are inherited forms and only eight are derivatives;
- of the 1,063 most frequent words (including the first 100), the great majority are of Latin origin (either inherited or learned). The exceptions are five Italianisms, one Anglicism and one Hispanism (*amour*, cited as an Occitanism, is of regional French origin), borrowings thus represent 0.7% of this figure ;
- among the 2,500 most common words (including the first 1,063), the foreign contribution is somewhat more noticeable. There are 22 Italianisms, ten Anglicisms, two Netherlandisms and three other loanwords; here, borrowings amount to 1.7%;
- about 5 to 6% of the ‘common’ words of present-day French (= between 50,000 and 70,000 word forms), are borrowings stemming from a great variety of languages. They are characterised by relatively weak polysemy and limited phraseological usage, which further diminishes their role in the target language (cf. 9.6.3 no. 1 above). The other 95% of words are of Latin origin (inherited or learned).

Among the words of Latin origin, it is important to distinguish inherited forms from learned forms and, in the case of each of these types, simple forms from derived forms. Among the simple forms, inherited words which have undergone phonetic evolution (e.g. *père* < PATRE, *mère* < MATRE) are notably less frequent (5%) than intralinguistic borrowings from Latin (25%), the importance of which increased with the centuries (to give one example among thousands, the learned form *liberté*, a medieval borrowing from Lat. LIBERTATEM [1266], became established, rather than the inherited form *livreteit* [attested slightly earlier, in 1190, but relying on uninterrupted oral tradition]). The great majority of forms (65%), however, are derivatives based essentially on inherited Latin words (e.g. *visage*, formed on the basis of *vis* < VISU); this illustrates the importance of derivation.

9.7 Onomastics and deonomastics

9.7.1 Proper nouns and common nominal lexemes

Onomastics is a branch of linguistics whose objective is to study proper nouns in general. Personal names relate to ‘anthroponymy’, place names, to ‘toponymy’.

Proper nouns have linguistic characteristics that are similar to those of common nouns (nominal lexemes), but with a number of restrictions. In the first place, their semantic dimension is more limited: a proper noun must allow the rapid, unequivocal identification of a specific person or place, in order to facilitate its recognition within a discourse. The terms *Louvre*, *Lisbon* or *Michelangelo* denote clearly determined referents. Even frequent names such as *Jack* or *Springfield* are normally easy to identify in a given discourse. In order to guarantee this principle, children belonging to the same family are generally given different names.

A common noun, in contrast, evokes a categorisation that assimilates a defined entity to similar entities: the *man* who happens to be facing us is a man among all others. Proper nouns do not allow generalisations or abstractions of this type. Consequently, while proper nouns have a referent like common nouns, their lexical meaning (“person named *x*”) and the extension of their concept (“ideas associated with the person named *x*”) are greatly reduced (cf. Kleiber, *Problèmes de référence: descriptions définies et noms propres*, 1981).

Similarly, mechanisms of derivation are less varied for proper nouns than for common nouns and both derivation and the formation of phraseologisms are less frequent. Conversely, the syntactic behaviour of proper nouns is closer to that of common nouns, except that the inflection of proper nouns is generally invariable (Fr. *la France*: f.sg., *les États-Unis*: m.pl.) and, in many varieties, they cannot be modified by articles (Fr. *Claude* or It. *Claudio*; cf. nevertheless fam. Fr. *le Claude*, northern and central It. *la Maria*)³².

Moreover, proper nouns cannot be translated without semantic alteration (*François* and *Francesco* do not have the same relationship to one another as *maison* and *casa*, which designate the same concept ‘HOUSE’); nevertheless, they may be borrowed and adapted (Fr. and Engl. *Milan* represent the adaptation of It. *Milano* and Fr. *la Maison Blanche* is a calque of *the White House*).

Despite these differences, the linguistic similarities between proper nouns and common nouns are emphasised by the ease with which words can move from one category to the

³² Place names, on the contrary, can often be determined by a definite article. The article is frequently incorporated into a place name at the time of its formation, where it becomes fixed and degrammaticalised. This occurred in the case of all denominative toponyms formed during the Romance period (i.e. after 700; e.g. *la Villette* or *Levallois* [with agglutination of the article]).

other: a proper noun may be used as a common nominal lexeme without formal transformation (conversion, cf. *poubelle* and *zeppelin* in 9.1.2); it may be transformed into an adjectival lexeme by derivation (cf. the example of *gargantuesque* above); it may also be integrated into a phraseologism, conferring on the latter its recurrent associations (cf. Fr. *boîte de Pandore* / Engl. *Pandora's box*, or examples of wordplay such as Fr. *relax*, *Max* or *cool*, *Raoul*). In contrast, etymologically speaking, proper nouns ultimately derive from common nouns.

A particular type of interdependence characterises ‘antonomasia’, which refers to the transition of proper nouns to common nouns and vice versa. In both cases, the motivation involved relies on salient features: either a character trait that then comes to designate the whole person (e.g. Fr. *le Seigneur* “the Lord” for “God”, Sp. *el Caudillo* for “Franco”) or an individual who is a typical representative of a category (e.g. *un Tartuffe*, Sp. *un narciso*, *un judas*, It. *un cicerone*).

Antonomasia resembles metonymic changes of the type “name of inventor” → “name of invention” (e.g. *zeppelin*) or “name of brand” → “product in general” (e.g. *kleenex*); however, it can reflect a more drastic semantic transformation whereby the change of a proper noun to a common noun is no longer a case of metonymy but of metaphor (a person referred to as a *Tartuffe* merely *resembles* the character from Molière’s play, whereas the word *kleenex* directly refers to the paper tissue itself; cf. 9.3.4).

The transformation of a proper noun into a common noun is referred to as ‘deonymisation’; the discipline that studies such transformations is known as ‘deonomastics’ (cf. Schweickard, *Deonomastik*, 1992 as well as the *Deonomasticum italicum* (= DI), an exemplary lexicographical work by the same author which assembles words derived from proper nouns in Italian). The transformation of a common noun into a proper noun is called ‘onymisation’, and is studied by the discipline of ‘onomastics’ (cf. below).

To summarise, whilst, historically speaking, the onomastic stock of a language ultimately derives from its lexis (cf. 9.1.2), proper nouns also rely on a system of auto-derivation: the name of a person may be based on a place name or another person’s name. Similarly, a place name may have been borrowed from another place name (cf. the diachronic typology established by Chambon, *Sur le système latin de dénomination des localités*, 2002):

- ‘delexical’ proper nouns rely directly on lexical units, which may or may not involve derivation or even compounding, as we have seen;
- ‘deanthroponymic’ proper nouns are formed from personal names, with or without derivation; e.g. Rom.n *Ionescu*, son of *Ion*, cf. the examples in 9.7.2 and 9.7.3 below;
- ‘toponymic’ surnames are formed from place names, with or without affixes; e.g. the Sp. surname *Soriano* is based on the toponym *Soria* (cf. *id.*). Place names can also be formed from other toponyms: this type of formation is typical for localities founded by the inhabitants of older, nearby places;

- finally, proper nouns (cf. 9.4.5 no. 2) – both place names and personal names – can be based on larger units of discourse (e.g. *Garrevaques*, from the imperative Occ. *engarra* “shackle” + *vacas* “cows”).

In the following section, the most important factors relating to the historical formation of personal names and place names in the Romance languages will be presented.

→ RSG 3, sect. XV, *Histoire interne: onomastique (roumain, Tomescu, art. 225; Bündnerromanisch, Liver, art. 229b; Italienisch, Gouvert/Paulikat, art. 233; français, Billy, art. 239; catalan, Bastardas, art. 249; español, Ruhstaller, art. 247; gallego, Boullón Agrelo, art. 250b; Portugiesisch, Weyers, art. 255)*

HSK 11, *Namenforschung/Names studies/Les noms propres*

9.7.2 The formation of personal names in the Romance languages

The Roman Empire had a complex system for the naming of persons: *nomen* – *cognomen* – (*nomen*) *gentilicium* (e.g. *Caius* (= n.) *Julius* (= c.) *Caesar* (= g.)). With the fall of the Roman Empire and the disintegration of its infrastructure (cf. 10.3.1), Latin-speaking Europe adopted a system of single names (of the type *Martinus*, which is the continuation of a *nomen* of the Classical period), which was sufficient in order to maintain less complex social relationships.

This system had an equivalent amongst the Germanic peoples where single names were often composed of two lexical elements (e.g. **gerhard* “spear” + “hard” > Fr. *Gérard*, **gaut-frid*, “name of a Germanic people + peace” > *Geoffroy*). Between the 6th and 8th centuries, the names of Germanic invaders became established through imitation almost everywhere in the *România*, rather like the names of film stars today. In the 10th century, up to 90% of the inhabitants of the former territory of Gaul, the northern part of the Iberian Peninsula and Italy had single names of Germanic origin (*Albert* < NP Germ. *Adalberht* < *adal* “noble” + *-berht* “brilliant, illustrious”; cf. 10.3.3).

Of the Latin single names, only about twenty survived during this period. These are the rare Latin names that underwent phonetic evolution, such as *Étienne*, *Jean* or *Pierre*. The overwhelming popularity of a small number of Germanic names caused a vast reduction of the original stock of anthroponyms. Owing to the limited variety of personal names in existence around the year 1000, the majority of today’s names entered the Romance languages after this date. The anthroponymic stock began to be enriched during the 11th and 12th centuries by means of the adoption of saints’ names or biblical names in learned or semi-learned form (e.g. *Gabriel*, *Matthieu*, *Paul*, *Lucien*).

Socio-cultural changes and demographic development during the 13th to 15th centuries led to the introduction of a double-name system (first name + family name) throughout the *România*. The second name, at first an individual nickname or byname (Fr. *surnom*), eventually became hereditary. The two names draw on different etymological sources:

earlier personal names (= patronyms), place names (= names reflecting a person's origin) and lexemes (= bynames or nicknames). Below are several examples:

patronyms or matronyms (naming after the first name of the mother or father; 'x is the son of y'), with a derivational element, such as Sp. *Sánchez* (derived from the NP *Sancho*) or *Fernández*, or Rom.n *Ionescu* (son of *Ion*); with a preposition, such as It. *De Giorgi(o)* and *Di Maria*, or without formal change, as in Fr. *Martin*, Sp. *García* (< *hija de García*);

names of origin ('x comes from z'): Sp. *Soriano* (derived from the place name *Soria*); Fr. *Duvillier* (< *Villier*); Sp. *Valdés*, It. *Altamura* (without formal change);

byname linked to personal characteristics, particularly names of professions (Fr. *Boulangier*, *Lefèvre* [= blacksmith], Sp. *Herrero*, It. *Ferrari*), functions or social roles (Fr. *Leroi*, *Lévêque*) and physical characteristics (nicknames: Fr. *Legrand*, Sp. *Delgado*, It. *Rossi/Russo*).

The use of terms referring to the human body in personal names has been studied in exemplary fashion by the project *PatRom (Dictionnaire historique de l'anthroponymie romane, 2004–)*. The entry for *OCULUS*, for example, contains the following information: NP Fr. *Deloeil*, Cat. *Ulhet*, Occ. *Ullet*, It. *Bellocchio*, Pg. *Olhalvo* (+ *ALBU-*). This project also clearly illustrates the contributions made by historical anthroponymy to our knowledge of Romance vocabulary (and of its medieval stages in particular).

In Western societies, individual names evolved from the onomastic basis described above during the second half of the second millennium. Numerous derivatives were formed, increasing the diversification of personal names (e.g. *Rossi/Rosso* → *Rossati*, *Rosselli(ni)*, *Rossicci*, *Rossin(i)*, *Rossettini*, *Rossoli(no)*, etc.).

In the *România nova*, specific patterns of variation emerged owing to the integration of indigenous terms and the widespread use of biblical names. Major migrations of the 20th century also brought new changes. The basic principles, however, had been established since the 16th century³³.

Note in this respect that in France a significant proportion of inhabitants (40%) live less than 50 kilometres from the place of origin of their family name (cf. Buchi, *La méthodologie de l'étymologie des noms de famille*, 2001).

³³ Thanks to parish registers, which were systematically introduced in Catholic circles following the Council of Trent (1563), detailed information on Romance personal names is available from the 16th century onwards. Even for the medieval period there are a large number of documentary sources listing at least the tax payers and property owners (cf. 11.1.4).

9.7.3 The formation of place names in the *Romània*

Place names are among the most stable types of linguistic signs in the Romance languages. Once formed, a toponym may survive for millennia, even though, like common nouns, they are subject to phonetic change. As an example, the above-mentioned place name *Antibes* is of Greek origin (< *antípolis*, cf. 9.1.2), but was maintained by the Latin-speaking population as well as by the Occitan-speaking population of the medieval period up until the modern period, when French began to take over. Consequently, it reflects the etymological *stratum* of the time of its emergence.

Place names in the *Romània continua* are extremely varied and thus constitute a valuable field of study for historical linguistics – a field, moreover, which remains largely underexploited; cf. the following examples belonging to different *strata*:

Pre-Romance (more specifically ‘pre-Latin’) toponyms are concentrated in the names of rivers and mountains in particular. Their etymology is thus, by its very nature, uncertain:

Meuse < pre-Celt. **mosa*, considered to be related to the Indo-Europ. root **mad* “soak, flow”

Alps < Lat. *ALPES* < pre-Indo-Europ. **alb-* “mountain, height, boulder”

cf. also *Marseille* < Gr. *Massilia*, a word (designating a waterway) arguably borrowed from Ligurian by the Greeks

Some of the pre-Latin common and proper nouns, after Latinisation, gave rise – in Latin – to toponymic derivatives:

Paris < *Parisiis* (ablative-locative) ← *Lutetia Parisiorum*, main town of the *civitas* of the *Parisi* people, possibly from Gaul. (Celt.) *pario-* “cauldron”

Toul < Gaul. **tullo* “swollen”

Metz < *Mediomatricum*, name of the Gaulish tribe whose main town was Metz, formed by the addition of the Gaul. prefix *mid-* “middle”

Numerous toponyms were formed during the Latin period on the basis of Latin lexemes or personal names:

in France, *Vic* (< *VICUS* “small town”)

cf. also the Celtic formation *Milano* < *Mediolanum* < *MEDIO-PLANUM* “in the middle of the plain (of the Po river)”

Formations from the Germanic period often display interference between Late Latin and the dominant Germanic language. In the Germanic-Romance frontier zone in particular, mixed forms are frequent:

Avricourt < Germ. personal name *Evericus* (= “eponymic” formation, i.e. after the person who gave the place its name) + Late Lat. *cortem*, a lexeme referring to an estate

Various borrowings include:

Arabic names in Spain, e.g. *Alcalá* < Ar. *qal’at* “castle, fortress” or *Gibraltar* < *gibl al-Tāriq* “mountain of Tariq (the leader of Arabic-Berber troops that began conquering Spain in 711 AD)”

Greek names in Italy, e.g. *Napoli* “Naples” < *néa pólis*; *Gallipoli* < *kalè pólis* “beautiful city”

The repopulation of the countryside from the High Middle Ages onwards gave rise to the formation of innumerable place names in Romance (denominal formations after 700 incorporate the definite article):

- many delexical names reflect topographic features, e.g. *Laval* < Lat. *VALLIS* “valley”, *La Pierre* (cf. also Rainer, *Instrument and place nouns in the Romance languages*, OxfEnc)
- with the development of the cult of saints, some localities adopted the name of their patron, e.g. *Saint-Benoît-en-Woevre*, *Sainte-Marie de Gondrecourt*
- others combined common nouns with personal names, e.g. *Château-Thierry* < Lat. *CASTELLU-* + Germanic personal name *Theodoricus*

In the *Romània nova*, the original names of the countries and saint’s names (*Los Angeles* = *Nuestra Señora de los Angeles*) greatly outnumber toponyms of more recent delexical formation (*Buenos Aires*, founded in 1536 as *Santa María del Buen Ayre*) and indigenous toponyms (*Mexico* < *Nahuatl mexika* “Mexican nation”).

Thus toponym formation involves as many lexical as anthroponymic elements; the three types of terms (toponyms, anthroponyms and lexemes) are highly interdependent. Consequently, toponyms also provide insights into the lexis of the period in which they emerged, as well as into underlying phonetic and morphological evolutionary patterns (cf. 10.3.4 no. 2). For linguists, their main interest lies in their contribution to the domain of etymology. However, this multifaceted field of study was neglected during almost one and a half centuries of etymological research, in contrast to the study of common lexemes. The majority of works on toponymy (for Gallo-Romance in particular) suffer from obvious weaknesses, owing to the fact that amateur research has always been more frequent than the work of genuine scholars. The opposite is true for general lexicography; in this domain, a selection of excellent and widely-known dictionaries is available for the Romance languages (cf. 9.9).

The most useful dictionaries of toponyms for Gallo-Romance are given here as examples (in chronological order, cf. FEWGuide 167–174):

Auguste VINCENT, *Toponymie de la France*, 1937.

Albert DAUZAT, *Traité d'anthroponymie française*, 1942; *Dictionnaire étymologique des noms de famille et prénoms de France*, 1951; *Dictionnaire étymologique des noms de lieux en France*, 1963 (enlarged edition co-authored by Charles ROSTAING).

Marie-Thérèse MORLET, *Les noms de personnes sur le territoire de l'ancienne Gaule*, 3 vols, 1968–1985; *Dictionnaire étymologique des noms de famille*, 1997 (based on Dauzat 1951).

Ernest NÈGRE, *Toponymie générale de la France*, 3 + 1 vols, 1990–1991 + 1998.

Andres KRISTOL (dir.), *Dictionnaire toponymique suisse* (DTS), 2005. (It should be noted that the evolution of Romance toponyms in Switzerland followed the same course as it did in neighbouring countries. The virtues of this excellent resource are extolled by Chambon in his extended review, *RLiR* 70, 2006).

The systematic practice of research on place names in their historical context would constitute a valuable contribution to linguistics. Onomastics is the field that offers the greatest potential for innovation within research on the history of the Romance languages (for additional reading, cf. Chambon, *L'onomastique du censier interpolé (ca 946) dans la charte de fondation du monastère auvergnat de Sauxillanges*, 2004; Grémois/id., *Les noms de lieux antiques et tardo-antiques d'Augustonemetum / Clermont-Ferrand*, 2008; id., *Recherches sur la toponymie de l'arrondissement de Lure (Haute-Saône)*, 2023).

9.8 Historical lexicology and etymology in Romance studies

Generally speaking, historical lexicology studies the transformation of vocabulary and lexical innovation over time. It encompasses semantics, derivational morphology and historical phraseology, as well as the history of borrowing, and may incorporate historical onomastics to some extent. It also involves elements of historical phonetics and grammar. It intervenes in various domains: it may focus on a specific field (such as the study of different types of morphological or semantic change, or the historical stratification of loanwords in a language); not confined to the domain of words, it also studies the transformation of concepts throughout history. With regard to both its specifically Romance tradition and its basic principles, historical lexicology is founded on etymology, which today is considered as one of its subdisciplines (cf. 2.2.4).

Etymology studies the relationship between a word or a name in a given language and the corresponding word or name in the mother language (i.e. the language that is the direct ancestor; for example, the etymon of O.Fr. *mere* and Fr. *mère* is Lat. *MĀTER*, in its accusative form *MĀTREM*) or – for borrowings – the contact language. At the same time, this relationship allows various parallel daughter languages to be compared with one another (*mère-madre* or, in the broader context of Indo-European – *māترم-Mutter-mother*). If necessary, the relationship may be further pursued in the direction of the mother language (e.g. Indo-European in the case of Latin); this, of course, is of only marginal interest to specialists of Romance linguistics.

Etymology is a useful tool which serves to categorise lexical changes and even vocabulary in synchrony, since it allows the identification of genetic families of words: the reliable grouping of words according to their common origin facilitates their

semantic delimitation as well as the identification of dominant semantic nuclei. Such categorisations are justified, since the transformations within a genetic family correspond to the way in which semantic memory is structured in the brain: semantic and derivational changes are not mere inventions of linguists, but the reflection of true cognitive relationships.

The identification of the etymology of large numbers of words also proves useful for synchronic lexicography in that it provides insights which are essential for the analysis of semantic differentiation. The quality of a dictionary of modern usage is thus dependent on a good foundation in historical and etymological lexicography. The standard of lexicological excellence attained by the *Petit Robert* (PtRob) would have been impossible without the FEW (cf. 9.9.1). Similarly, the deficiencies of the DRAE (*Diccionario de la Real Academia Española*), the best usage dictionary for Spanish, are due, among other reasons, to the fact that the documentation provided by the otherwise excellent etymological dictionary by Corominas is not particularly abundant.

The importance of etymology is such that it can be considered as the ‘backbone’ of historical lexicography. Etymology, and etymology alone, provides a reliable interpretative framework within which all diachronic developments – lexical, phonetic or grammatical – can be described with a high degree of accuracy. It is particularly important for lexicology as it allows the precise identification of the means a language possesses for the transformation of its vocabulary. Each new form relies on another, older form. Only onomatopoeia may appear spontaneously at any time, enriching the inventory of lexemes; they thus form a separate category (cf. the recent fundamental study by Georgescu, *La regularidad en el cambio semántico. Las onomatopeyas en cuanto centros de expansión en las lenguas románicas*, 2021). In contrast, all other phenomena which constitute perturbations of simple relationships, such as popular etymology, word blending or voluntary transformations such as *verlan*, rely on the previously existing lexical inventory (cf. 9.3.6).

Etymology is the oldest subdiscipline of historical lexicography. Its origins lie both in Antiquity (e.g. Isidore of Seville, *Etymologiae*) and in the Modern Age (e.g. Ménage or Aldrete in the 17th century). With the discovery of the laws of sound change, it became a modern discipline (cf. the etymological dictionary of the Romance languages by Diez 1854, re-elaborated by Meyer-Lübke ³1935), which paved the way for historical semantics and onomasiology, as well as for historical geolinguistics (all three associated with the field of etymology in Wartburg’s monumental *Französisches Etymologisches Wörterbuch*).

The integration of semantics into the methodology of etymological argumentation (which initially concentrated on phonetics) was accompanied by a great deal of debate, in which the onomasiological *Wörter und Sachen* movement, which was highly active in the first half of the 20th century, was involved (cf. 4.2.1 no. 4, 9.5). This movement placed particular (and from a theoretical point of view, disproportionate) importance on the

study of objects, especially those pertaining to rural life (*cf.* the critique by Vârvaro 1968).

Historical lexicology is useful in various domains, both theoretical and practical. It is involved in the creation of scientific terminologies and the elaboration of standardised written languages (such as Romanian in the 19th century, or Catalan, Galician and – more recently – Rumantsch Grischun or Dolomitic Ladin in the 20th century). The analysis of patterns of semantic or derivational innovation contributes to our knowledge of language-internal history. Thus, lexis facilitates the identification of links to sociopolitical or cultural transformations. Finally, historical lexicology allows reflection on the cognitive structures underlying specific types of change.

General interpretative studies such as these require a great deal of detailed work; historical lexicography and etymology have mobilised considerable scientific energy over the past 150 years, creating an exceptional basis for research. However, each investigation of a fundamental nature results in an infinite number of new questions of detail to be resolved. Some figures concerning the evolution of the everyday vocabulary of French have already been presented (*cf.* 9.6.4): 5% of modern words are inherited Latin-based words, 65% are derivatives formed from inherited words, 25% are Latinisms and a final 5% are borrowings from other languages. This very approximate calculation does not take into consideration the percentage of meanings that persist over time, nor does it consider changes in the meaning of a lexeme since its entry into French. Furthermore, it does not take into account the numbers of phraseologisms that were either maintained or that underwent changes. The incorporation of these last two aspects into analyses of lexical change would provide a more accurate idea of the extent of Romance transformations over one and a half millennia, but it would require considerable effort.

From a quantitative point of view, it is illuminating to observe the elements of Latin vocabulary that survived into the Romance languages (*cf.* Stefenelli, *Das Schicksal des lateinischen Wortschatzes in den romanischen Sprachen*, 1992). Of the words that make up the entire Latin basis of the Romance languages (judging from the 8,000 entries in the FEW), 80% are attested in written Latin sources while 20% were confined to oral use (they are known to us today only through phonetic reconstruction). There is thus a fair amount of consistency between written Latin and spoken (Late) Latin.

However, the Latin basis of the Romance languages constitutes a mere 20% of the words attested by Latin sources, that is, 50% of commonly-used Latin words (the remainder do not survive in any Romance language). Nevertheless, of the 1,200 most frequent words in written Latin, two-thirds survive in the various Romance languages (half of them in most if not all Romance languages). The vocabulary of spoken Latin that survived in a given Romance language is thus relatively limited, but its number increases exponentially for frequently used words. Below are some examples:

TOTU / *TOTTU > Sp. Pg. *todo* / Fr. *tout*, It. *tutto*, Occ. Cat. Rom.n *tot*, etc.

OCULU > Fr. *œil*, It. *occhio*, Sp. *ojo*, Occ. *uelh*, Rom.n *ochiu* (the differentiation is purely phonetic)

SANGUE > Fr. *sang*, It. *sangue*, Occ. Cat. *sanc*, Rom.n *sânge* vs. SANGUINE > Sp. *sangre* “blood”

CINERE > It. *cenere*, Fr. *cedre*, Engad. *endra* vs. CINISIA > Dalm. *kanaisa*, Sard. *kijina*, Sp. *ceniza*, Pg. *cinza*, Rom.n *cenușă* “ash” (cf. REW: lexical differentiation due to differing regional Latin forms)

Still considering the 8,000 Latin and pre-Latin bases (including onomatopoeic bases), a little less than a third (2,300) have a wide distribution in the modern *Romània* (with 1,750 Classical words and 500 from spoken Late Latin); for the medieval Romance languages, this percentage is higher.

Thus, a drastic reduction of Latin lexemes can be observed in the passage to the Romance languages, especially for words of lower (literary) frequency. The Romance languages evolved from this basis and then diversified. Furthermore, they enlarged the common word stock by means of the usual mechanisms of linguistic change, acquiring more and more distance from Latin and from each other. Such centrifugal forces were, however, counterbalanced by continuous borrowing from Latin (intralinguistic borrowing), resulting in an ever-increasing degree of resemblance between the written languages and their mother language and sister languages.

Questions of Latin-Romance divergence and convergence in vocabulary have only rarely been pursued from a global perspective (how many Latin words with or without derivation, with or without semantic change, survive in a given Romance language at a given period? How many Latinisms or other borrowings were added? In which semantic fields? With which restrictions in usage?). Although research of this nature would require considerable effort, it would result in a more precise definition of the evolutionary physiognomy of a specific Romance language or of the Romance languages as a whole.

Such questions always underlie the numerous detailed studies on the history of words and concepts (cf. 9.3.7 no. 4). At the same time, they imply comparative and typological considerations; for example, the study on words describing the impact of physical activity on objects (Klein, 1997) shows that Italian and French quite clearly maintain the structure of the semantic fields of Latin, whereas in Romanian and especially Spanish and Portuguese, they have been reorganised. Thus, other rules may govern the evolution of semantic structure than those which determine the maintenance or abandonment of forms (cf. 3.5.2).

→ Buchi/Dworkin, *Etymology in Romance*, OxfEnc

Stefenelli, *Lexical stability*, CambrHist 1, 11

Dworkin, *Lexical change*, CambrHist 1,12; *Lexical stability and shared lexicon*, OxfGuide 32

Pfister/Lupis, *Introduzione all'etimologia romanza*, 2001

Glessgen/Schweickard, *Étymologie romane. Objets, méthodes et perspectives*, 2014

Glessgen, *L'appart des 'Inconnus' du FEW à la recherche étymologique*, 2019

9.9 Research methods: historical and etymological lexicography

Lexicography plays a major role in research on words. Like grammatical scholarship (cf. 8.5), it is of practical use in language standardisation as well as in the organisation of lexical knowledge, which is scattered owing to the exponential increase in knowledge in the modern world. Lexicography – again, like grammatical scholarship – necessarily remains approximate and fragmentary. This is particularly evident even today with regard to the domains of phraseology and variational connotation.

The number of dictionaries currently available for use in lexicological research is worthy of the multitude of words in a language. For Modern French, for example, there are several good quality general dictionaries (the *Petit Robert* [PtRob], the (*Grand*) *Robert* [Rob], the *Petit* and the *Grand Larousse* [PtLar, GdLar] as well as the *Trésor de la langue française du XIX^e et du XX^e siècle*) [TLF]). Parallel to these, there are numerous reverse dictionaries, dictionaries of frequency, pronunciation, synonyms, antonyms, popular French and argot, neologisms, borrowings (Anglicisms), abbreviations and encyclopedic dictionaries, the quality of which varies greatly. Moreover, the use of databases such as *Frantext* (created for the compilation of the TLFi) or the *Monde sur cd-rom* is essential (cf. 11.6.4).

As for diachrony, there are historical dictionaries (dictionaries of older periods that cite words in their context), as well as etymological dictionaries *per se*; cf. the following references, beginning with French.

9.9.1 Historical and etymological dictionaries of French

1 Selected fundamental dictionaries compiled between the 16th and 19th centuries

For the following references, cf. Bierbach and Pellat, *Histoire de la réflexion sur les langues romanes: le français*, RSG 1, art. 21, ch. 3. *Lexicographie* (cf. also 10.5.3 no. 3) and the more detailed FEWGuide, which also provides the URLs, as many dictionaries are now available on the Internet:

ESTIENNE Robert, *Dictionnaire françois[-]latin*, 1539: the first dictionary to be comprised of French entries with definitions in Latin instead of Latin entries with definitions in French.

COTGRAVE Randle, *A Dictionarie of the french and english tongues*, 1611: a bilingual dictionary (Fr.-Engl. / Engl.-Fr.), rich in lexical data and reliable.

RICHELET César Pierre, *Dictionnaire françois, contenant les mots et les choses*, 1680; FURETIÈRE Antoine, *Dictionnaire universel*, 1690; *Dictionnaire de l'Académie*, 1694: the three earliest monolingual French dictionaries.

TRÉVOUX, *Dictionnaire universel français et latin*, 3 vols., 1704: encyclopedic dictionary compiled by the Jesuits of Trévoux. Its successive editions were extensively enlarged, especially those of 1732 (5 vols.) and 1743 (6 vols.) and 1771 (8 vols.).

DIDEROT Denis / D'ALEMBERT LE ROND Jean (ed.), *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers Français*, 35 vols., 1751–80: the monumental *Encyclopédie française* (EncFr), which provides ample lexical material and numerous plates, among other aspects.

LITTRÉ Émile, *Dictionnaire de la langue française*, 4 vols., 1863–72: the model for French lexicography until the second half of the 20th century.

LAROUSSE XIX^e, *Grand dictionnaire universel du XIX^e siècle*, 15 vols. + 2 suppl., 1866–1890.

HATZFELD Adolphe and DARMESTER Arsène, *Dictionnaire Général de la langue française du commencement du XVII^e siècle jusqu'à nos jours*, 2 vols., 1890–1900: to be considered as a model, owing to its well-founded and rigorous semantic distinctions.

2 The most important historical dictionaries

GODEFROY Frédéric, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, 10 vols., 1880–1902 [= Gdf]: includes vocabulary from the entire range of textual genres in Old and Middle French up to the 15th century; based on research conducted directly on manuscripts, both literary and scientific (in libraries) and documentary (in archives).

TOBLER Adolf / LOMMATZSCH Erhard, *Altfranzösisches Wörterbuch*, 11 vols., 1925–2002 [= TL]: only includes vocabulary from literary texts (up to the 14th century), but provides an excellent semantic structure; includes references to corresponding entries in the Gdf and FEW.

BALDINGER Kurt / MÖHREN Frankwalt / STÄDTLER Thomas, *Dictionnaire étymologique de l'ancien français*, 1979–2021, [= DEAF]: covers only part of the alphabet (currently letters E to J) but considers all textual genres up to the 14th century; the documentation for the entire alphabet is now available in the database GallRom. The DEAF is an exemplary work of Romance lexicography with respect to the information it provides on dating and localisation, as well as its definitions. The *Complément bibliographique* by Möhren (DEAFCompl 2021, also within the database GallRom) provides useful insights for the majority of edited sources in Old French. The abbreviations provided by the DEAF, moreover, serve as a reference for the citation of old texts.

STONE Louise / ROTHWELL William, *Anglo-Norman Dictionary*, 1977–1992 [= AND] (enlarged online version, 2nd edition in the course of publication 2005–): describes lexis contained in written French sources from the British Isles.

MARTIN Robert, *Dictionnaire du moyen français* [= DMF]: based on a database of texts and dictionary entries; like the *Trésor de la langue française* (cf. below), it is under the direction of the research unit ATILF (*Analyse et Traitement Informatique de la Langue Française*) and accessible online; it covers the period from 1350 to 1500.

HUGUET Edmond, *Dictionnaire de la langue française du seizième siècle*, 1925–1967 [= Hu]: includes vocabulary from literary texts from the 16th century; only provides approximative definitions comparable to those of the Gdf.

Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècle (1789–1960), ed. by Paul IMBS, 16 vols., 1971–1994 [= TLF]: first point of reference for Modern French; its historical remarks can be a useful complement to those of the FEW (cf. Radermacher, *Le 'Trésor de la langue française'*, [ms.]). Like the DMF, it is based on the database Frantext, elaborated by the ATILF research unit. It also covers the 16th to 18th centuries (this is particularly noteworthy as the 17th and 18th centuries constitute a considerable gap in the historical lexicography of French) and the end of the 20th century (cf. 11.6.4). The electronic version of the dictionary (TLFi) is freely accessible on the Internet.

3 Etymological dictionaries

WARTBURG Walther von, *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, 25 vols., 1922–2002 [= FEW]: includes the vocabulary of French, Francoprovençal, Occitan and Gascon, from their beginnings up to the present-day period, encompassing both written language and dialects; its etymological interpretations also contain references to the other Romance languages, for which the FEW is thus also a significant reference work.

- Structure of the FEW: vols. 1–14 treat Latin, Greek, Pre-Latin and onomatopoeic etyma (*cf.* also vols. 24–25, which contain a revised version of letter A); vols. 15–17 treat Germanic etyma; vols. 18–20, English etyma, etyma of Oriental origin as well as etyma of diverse origin; vols. 21–23 treat words of unknown origin arranged in onomasiological order, according to Hallig and Wartburg's *Begriffssystem*, 1963.
- Sources and abbreviations (with numerous commentaries on the quality of the sources): *FEWCompl* 2010; complements providing the etymology for the words of unknown origin: Baldinger, 1988–2003; for words of Oriental origin: Arveiller, 1999; for the structure of the work, hidden etyma and Slavisms: Buchi, *Les Structures du FEW*, 2003. *Cf.* also the index (ed. Buchi, 2 vols., 2003).
- Indispensable introduction to the use of the fundamental but complex dictionary: FEWGuide.

BLOCH Oscar / WARTBURG Walther von, *Dictionnaire étymologique de la langue française*, ⁶1975 [= BIWbg]: succinct, ample and reliable; facilitates the locating of French forms in the FEW.

REY Alain, *Le Robert. Dictionnaire historique de la langue française*, 2 vols., 1992; new ed. 3 vols., 2010 [= DictHist]: relies to a great extent on the etymological and historical notices of the TLF, which remain the first point of reference and which are easily accessible on the Internet, in the TLFi.

There are a large number of small historical and etymological dictionaries of French, such as Gamillscheg, *Etymologisches Wörterbuch der französischen Sprache* or Dauzat *et al.*, *Nouv. Dict. étym. et hist.*, which do not add anything to the information provided by BIWbg, TLF and the FEW. It should further be noted that the etymological indications of the PtRob are fairly reliable.

9.9.2 Historical and etymological dictionaries of the other Romance languages

1 Romance languages in general

DIEZ Friedrich, *Etymologisches Wörterbuch der romanischen Sprachen*, 1854: of purely historical interest, since it has been surpassed by the REW.

MEYER-LÜBKE Wilhelm, *Romanisches etymologisches Wörterbuch*, ³1935 [= REW]: remains very useful, for non-Gallo-Romance languages in particular.

2 Occitan

RAYNOUARD François, *Lexique roman ou dictionnaire de la langue des troubadours comparée avec les autres langues de l'Europe latine*, 1838–1844, vols. 2–6 [= Rn]: the first historical and etymological dictionary of the Romance languages. The first volume contains a collection of troubadour poetry, the final volume is an index, which is necessary due to the grouping of derivatives under the entries for basic forms. It can still be considered an acceptable reference work despite its poor definitions and the period at which it was written.

MISTRAL Frédéric, *Lou tresor dou Felibrige, ou dictionnaire provençal français embrassant les divers dialectes de la langue d'oc moderne*, 2 vols., 1878 [= M]: very rich, also useful for onomastics and older language stages.

LEVY Emil, *Provenzalisches Supplement-Wörterbuch. Berichtigungen und Ergänzungen zu Raynouards Lexique roman*, 8 vols., 1894–1924 [= Lv]: completes and expands the dictionary of Rn; definitions are often useful and may also be useful as a guide for other medieval Romance languages.

LEVY Emil, *Petit Dictionnaire provençal-français*, 1966 [= LvP]: ample but succinct and reliable.

FEW: cf. 9.9.1 no. 3

STIMM Helmut / STEMPEL Wolf-Dieter, *Dictionnaire étymologique de l'occitan médiéval*, 1996 [= DOM]: only covers the beginning of the alphabet (*a–album*).

BALDINGER Kurt, *Dictionnaire onomasiologique de l'ancien occitan*, 1975–2007 (DAO).

BALDINGER Kurt et al., *Dictionnaire onomasiologique de l'ancien gascon*, 1975–2021 (DAG).

GLESSGEN Martin, *Dictionnaire étymologique d'ancien gascon*, 2023– (DEAG): consists of the DAG and its unpublished materials; accessible through the GallRom website.

→ FEWGuide.

3 Italian

TRAMATER, *Vocabolario universale italiano, compilato a cura della Società tipografica Tramater*, 7 vols., 1829–1840 [= Tramater]: the first historical dictionary for Italian, and for the Romance languages in general, to be compiled according to modern principles; remains interesting even today for technical and scientific language.

TOMMASEO Nicolò et BELLINI Bernardo, *Dizionario della lingua italiana*, 7 vols., 1865–79 [= TB]: large historical reference dictionary, now largely surpassed by the GDLI.

BATTAGLIA Salvatore, *Grande Dizionario della lingua italiana*, 21 vols., 1961–2002 [= GDLI]: currently the principal historical dictionary for Italian.

TLIO – *Tesoro della lingua italiana delle origini*, online dictionary covering the period from the emergence of the written Italo-Romance vernaculars up to 1375; relies on the textual database of the *Opera del Vocabolario Italiano* (OVI; cf. 11.6.4), also freely searchable on the Internet. The database is founded on exemplary philological rigour.

PFISTER Max, *Lessico etimologico italiano*, 1979– (directed by Wolfgang Schweickard for vols. 8–10; directed by Elton Prifti from vol. 11 onwards; currently encompasses large parts of the letters A to G) [= LEI]: first point of reference for etymology not only for the Italo-Romance domain (including Ladin), but also for the other Romance languages, due to the etymological discussion it provides as well as the rigorous semantic structure of its articles; relies on the principles of the FEW, to which its material and interpretations also form a complement. The series of Germanic borrowings has been in progress since 2000 (Elda Morlicchio, *LEI. Germanismi*); the publication of the series containing words of Arabic, Turkish and Persian origin began in 2022 and will be completed in 2025 (Wolfgang Schweickard); deonymic entries are treated in the *Deonomasticon Italicum* (cf. below); cf. LEISuppl for the list of sources and abbreviations, as well as Aprile, *Le Strutture del Lessico Etimologico Italiano*, 2004.

SCHWEICKARD Wolfgang, *Deonomasticon Italicum. Dizionario storico dei derivati da nomi geografici e da nomi di persona. Vol. 1. Derivati da nomi geografici*, 4 vols., 1997–2013 [= DI]: the documentation both of the toponyms (which are the headwords) and the deonymic derivations is exemplary; the treatment of printed sources from the 16th to 19th centuries is of an exceptionally high standard; the bibliographical supplement (2012) containing the references for this extensive repertoire is searchable on the Internet.

BATTISTI Carlo / ALESSIO Giovanni, *Dizionario etimologico italiano*, 5 vols., 1950–57 [= DEI]: short and reliable etymological dictionary with a large number of headwords that also include dialect forms.

CORTELAZZO Manlio / ZOLLI Paolo, *Il nuovo etimologico: DELI – Dizionario etimologico della lingua italiana*, 1999 [= DELIN]: fewer headwords, limited to standard Italian, but provides excellent etymological discussion.

FARÉ Paolo A., *Postille italiane al "Romanisches etymologisches Wörterbuch" de W. Meyer-Lübke, comprendenti le "Postille italiane e ladine" di Carlo Salvioni*, 1972 [= Faré]: Italo-Romance complement to the REW, etymologically reliable and based on dialect forms.

4 Spanish

ALONSO PEDRAZ Martín, *Diccionario medieval español. Desde las Glosas emilianenses y Silenses (s. X) hasta el s. XV*, 2 vols., 1986 [= DME]: a rather unreliable dictionary, both as regards the selection of headwords and the dates and definitions it provides.

MÜLLER Bodo, *Diccionario del español medieval*, 1987–2005 [= DEM]: a work of high quality, but which currently only covers letter A; however, the whole of its material is now accessible online.

GILI GAYA Samuel, *Tesoro lexicográfico. 1492–1726*, 1960 [= TesLex]: a highly useful work which assembles the entries of Spanish dictionaries of the period in question (covering the letters A to E).

Real Academia Española, *Diccionario de Autoridades*, 1726–39 [= DiccAut]: the first historical dictionary for the *România*, still very useful today, particularly for the 16th to 17th centuries.

CUERVO Rufino José, *Diccionario de construcción y régimen*, 8 vols., 1886–1998 [= DCR]: the only historical dictionary for the *România* with a syntactic orientation.

Real Academia Española, *Diccionario histórico de la lengua española*, 1960– [= DH]: a high quality dictionary, although letter A has not yet been completed. In the absence of an extensive historical dictionary for Spanish, the database CORDE, also incorporated within the database *Corpus del Dicionário histórico del español* (CDH, cf. 11.6.4), is a useful substitute.

COROMINAS Joan / Pascual José Antonio, *Diccionario crítico etimológico de la lengua castellana y hispánica*, 6 vols., ²1980–91 [= DCELCH]: an ample dictionary to be used in conjunction with the Catalan DECLC.

COROMINAS Joan, *Breve diccionario etimológico de la lengua castellana*, ³1973 [= BrCor]: based on the DCELCH; remains the best concise etymological dictionary of Spanish.

GARCÍA DE DIEGO Vicente, *Diccionario etimológico español e hispánico*, ²1985 [= DEEH]: a complement to the REW similar to that of Faré, though a great deal less comprehensive and etymologically less reliable.

5 Catalan

ALCOVER Antoni M. / MOLL Francesc de B., *Diccionari Català-Valencià-Balear*, 10 vols., 1926–69 [= DCVB]: provides a good point of reference, although its historical component is rather brief.

Coromines Joan, *Diccionari etimològic i complementari de la llengua catalana*, 9 vols., 1980–89 [= DECLC]: based on the DCELCH, which it complements and corrects.

6 Portuguese

MACHADO José Pedro, *Dicionário etimológico da língua portuguesa [...]*, 5 vols., ²1967.

CUNHA Antônio Geraldo da, *Índice do vocabulário do português medieval*, 1986–88.

GUÉRICOS Rosário Farâni Mansur, *Dicionário de etimologias da língua portuguesa*, 1979.

MESSNER Dieter, *Dictionnaire chronologique des langues ibéroromanes*, vol. 1: *Dictionnaire chronologique portugais*, 1976.

7 Romanian

CIORĂNESCU, Alexandru, *Diționarul etimologic al limbii române*, 2001.

Diționarul limbii române (DLR), București, 1913; Serie nouă, 1986–.

PUȘCARIU Sextil, *Etymologisches Wörterbuch der rumänischen Sprache*, vol. 1: *Lateinisches Element*, ²1975.
 TIKTIN, Hariton / MIRON Paul, *Rumänisch-deutsches Wörterbuch*, 3 vols., ²1985–89.

8 Sardinian, Francoprovençal, Romansh, Ladin, Friulian

WAGNER Max Leopold, *Dizionario etimologico sardo*, 3 vols., 1960–64 [= DES].

GAUCHAT Louis et al. (eds), *Glossaire des Patois de la Suisse Romande*, 1924– [= GPSR]: currently 7 vols., letters A to F.

DE PLANTA Robert (ed.), *Dicziunari Rumantsch Grischun*, 1939– [= DRG]: currently 9 vols., letters A to M.

KRAMER Johannes, *Etymologisches Wörterbuch des Dolomitenladinischen*, 7 vols., 1988–99 [= EWD].

PELLEGRINI Giovan Battista (ed.), *Dizionario Storico Etimologico Friulano*, 1984– [= DSEF].

9.9.3 Final remarks

Historical and etymological lexicography has always held a place of particular importance in Romance studies. In the comprehension of old Romance texts, difficulties relating to lexis and the meaning of words are far more frequent than grammatical problems; this is not surprising due to the close typological proximity of the different Romance languages to one another. Compiling historical and etymological dictionaries requires an enormous investment in terms of working hours. This explains the large number of projects that have been in progress for decades. Even in their incomplete state these dictionaries provide a great deal of varied information on the evolution of the languages in question, as well as on available sources and evolutionary phenomena.

Among the philologies of different languages, the field of Romance studies excels due to its high methodological standards and the diversity of its historical and etymological lexicography. Nevertheless, a satisfactory state of research has only been achieved for the Gallo-Romance and Italo-Romance languages, and even in these domains, much remains to be done and is being done. Spanish and Catalan suffer from a lack of historical dictionaries, despite the high quality of the available etymological works. The situation for Portuguese and Romanian is largely unsatisfactory; the greatest potential for future research lies in these areas. Paradoxically, etymological lexicography is intensely pursued for the ‘small languages’ of the Romance family (9.9.2 no. 8) and it is very reliable.

A certain degree of familiarity with the different works cited above is useful, even if one is chiefly interested in a specific language of a particular period or in literary aspects. Historical lexicography provides a good overall picture of the textual heritage available for the Romance languages, thus serving as a useful point of orientation for the study of older textual genres.

Part 4: **External history of the Romance languages
and varieties**

10 External history of the Romance languages and varieties

10.1 Establishing a framework for the study of external history

10.1.1 The relationship between external and internal history

External history is the essential counterpart of internal history. It places the transformations that occur within the various domains of language into a communicative and variational context, studying the extralinguistic factors that influence the internal history of a language. The two viewpoints are mutually dependent, with regard both to linguistic reality and to how research is conducted in practice.

An analysis of linguistic entities from an internal point of view is necessarily determined by external considerations: the choice of studying several language domains in parallel, of focusing on one in particular (e.g. the phonological system), or of examining a circumscribed aspect of an individual domain (e.g. sibilants), is externally motivated; similarly, any attempt to distinguish historical periods or to delimit textual genres is an external operation. Internal analysis can only begin once chronological and variational choices have been made and a corpus of texts to be studied has been defined; indeed, it is only after these rather demanding issues have been clarified that an adequate internal analysis is possible. Theoretically, at least, this allows a ‘hermeneutic’ approach, i.e. the analysis of the linguistic signs present within a defined body of texts, independently of all context. In practice, however, there is always a back-and-forth between the external and the internal approaches.

These methodological difficulties also apply to the study of present-day languages: our intuitive familiarity with the latter easily lead us to forget that an internal analysis necessarily relies on implicit external preconceptions – be it with regard to the definition of varieties or our judgement of the acceptability or frequency of specific phenomena (cf. 4.1.2). The varieties of a given historical language display internal cohesion, but also differ significantly from one another; modern standard Italian, the Venetian dialect and the popular variety of Italian spoken in Rome are a case in point. The development of these three varieties must therefore be studied separately before a ‘general’ account of the history of Italian can emerge.

It is thus appropriate to view linguistic history as a history of varieties and textual genres, involving both external and internal approaches. In order to select a specific variety of a historical language for analysis, it is first necessary to define the various regional and social varieties which compose the language, and to identify the available sources for these varieties. The external approach is indispensable in delimiting the varieties and languages in use and in determining the conditions of their use; at the same time, only the internal approach allows the precise nature of the linguistic

variation present in the sources chosen to be defined and, by this means, the selection of the latter to be refined.

The most significant obstacle to research on the history of languages is the fact that only the written forms of a language – which are by definition prestigious varieties – can be studied across historical periods. For both internal and external aspects, therefore, diachronic linguistics cannot bypass the history of standard languages. In order to obtain an accurate description of non-standard varieties of a historical language, which are less well represented in writing, the standard variety must necessarily serve as a term of comparison; this is a process known as ‘differential analysis’. It also applies to all historical lexicography and grammatical analysis, although it must be said that the Romance tradition has always placed emphasis on the identification of non-standard elements in historical sources.

Differential analysis is further complicated by the omnipresent discrepancy between the spoken dialects and the written regional varieties (cf. 10.4.3 no. 3), or – especially in Italy – the written varieties of important cities (such as Venice, Genoa or Naples) during both the Middle Ages and the modern period. Such varieties, displaying a rich secular tradition, were subject to processes of pre-standardisation from early on. For each variety, it must be decided whether an independent description is justified or whether differential analysis is sufficient. From a practical and typological point of view, differential analysis has the advantage of being easier to apply and the results it produces are more specific.

In short, external history relies on the assumption that major evolutionary processes taking place at a language-internal level are in constant interaction with extralinguistic factors, such as political, economic, social and cultural development, ideological orientations or aesthetic tendencies. Its goals are to identify and study in greater detail the influence of these factors on the use of a language, to describe the structure of its diasystem and to allow an adequate interpretation of internal changes. Internal and external evolution are inherently linked and they represent two aspects of a single research subject – the changes in language use over time.

10.1.2 The preoccupations of external history

The distinction between internal and external history was first established by Georg von der Gabelentz (*Die Sprachwissenschaft*, 1891; cf. 2.2.4 no. 2). It was next encountered as the main principle underlying the monumental *Histoire de la langue française* by Ferdinand Brunot (1905–1939). Other studies throughout the 20th century focused more specifically on the fundamental aspects of external history, such as the formation of major language regions, beginning with the *Orígenes del español* by Ramón Menéndez Pidal published in 1926 (this was followed, albeit much later, by Baldinger, *La Formación*

de los dominios lingüísticos en la Península Ibérica, 1972 [1958] and Wüest, *La Dialectalisation de la Gallo-Romania*, 1979).

In the 1920s and 1930s, the trend of idealism attempted to specify the interdependency between external factors and internal linguistic data; however, the state of linguistics, like that of other social sciences, was not sufficiently advanced at the time to allow meaningful interpretations. Hypotheses such as that of a link between French verb tenses and the ‘national’ character of the French themselves assumed too close a relationship between external and internal factors and did not consider the internal logic of language, which is much better known today (cf. 2.2.4 no. 2).

Throughout the 20th century, large-scale questions were for the most part treated on an insufficient empirical and interpretative basis and did not always focus on truly essential aspects. A question such as ‘Can the rapid process of sound change in Medieval French be explained as a consequence of the centralised structure of the Carolingian state?’ is perfectly legitimate, but it is formulated in too restrictive a manner (the degree of centralisation in the 9th century was minimal in comparison to that of the 19th century, and the complete absence of mountainous terrain which characterises the lowlands of central-northern France, in addition to the density of the population owing to the fertile land – not to mention its central situation –, must certainly have played a greater role in linguistic evolution). ‘Is the slow progression of phonetic evolution observed in some dialects of Italy a result of the strong presence of Latin in Italy?’ Here, both elements of comparison – the ‘slow progression of evolution’ and the ‘strong presence of Latin’ – are too general. ‘Is the weak differentiation of Romanian a result of the lack of stability in a society dominated by a migratory and nomadic lifestyle?’ This question can in all probability be answered affirmatively (cf. 3.4.12 no. 4).

The exemplary field of study that the *Romània* offers for the external history of its languages, due to its exceptional historical documentation, merits more thorough large-scale and comparative research. It would be of great interest, for instance, to examine whether or not acceleration or deceleration of language change occurred during particular periods and within specific Romance varieties (cf. Vachon, *Le changement linguistique au XVI^e siècle*, 2010, which offers an excellent example of this type of investigation).

Recent studies and manuals aim to improve the scientific basis of external history as an analytical approach, chiefly by means of a more systematic identification of the factors involved, as well as cautious evaluation of their possible impact on language-internal changes. This is the objective of works such as the *Storia della lingua italiana* directed by Serianni and Trifone (1993–1994, 3 vols.), and of the three volumes that continue Brunot’s groundbreaking *Histoire de la langue française* for the years 1880–2000, directed by Antoine, Martin and Cerquiglini (1985–2000), as well as of the three volumes of the *Romanische Sprachgeschichte* (RSG).

The guiding principle for the following chapter comprises three fundamental questions which have already benefitted from a copious amount of research, but whose methodological interest extends beyond the scope of Romance studies.

The first of these concerns the ‘fragmentation of the *Romània*’, which implies the transition from Latin to the Romance languages. This question focuses on when, how, and why Latin evolved into a considerable number of geographically differentiated daughter languages. This was a process of macroscopic language change, accompanied by the ‘territorialisation’ of variation, which is inherent to language (that is, geographical areas becoming associated with specific varieties; it is important to understand that this is the only way in which new languages emerge, *cf.* 1.2.5). An initially unified diasystem split into a number of various sub-systems, each of which subsequently evolved into a separate historical language with its own diasystem. Most of this evolutionary process took place during a period when textual sources were scarce (from the 5th to the 10th century AD); nonetheless, a certain amount of concrete evidence exists, whereas for many other language families, we can only rely on conjecture.

A second major question can be addressed more directly: the means by which the Romance languages were put into writing and standardised from the Middle Ages onwards. During this process, writing in Romance imitated the pre-existing example of Latin, which served as a guide for alphabetic writing with regard to textual models, as well as lexical and syntactic choices. In this respect, the Romance languages reproduced the development of Latin; the latter, when put into writing, imitated the models of its predecessors – Etruscan, Greek and Phoenician (*cf.* 10.2.1). The *Romània* offers great potential for the study of many facets of the development of written and standardised languages which took place between the 9th and the 20th centuries, and which is of great sociological importance for the modern world.

A third question concerns the establishment of complex linguistic diasystems. This evolutionary process which has accompanied the development of modern societies is particularly well illustrated by the linguistic history of the Romance languages of the past five centuries, following the formation of the standard varieties.

10.1.3 Establishing criteria for the study of external history

In order to study the external history of a language, it is essential to establish a general framework as a point of reference for analysis, similar to the distinctions operated at an internal level between phonology, morphology, syntax and lexis (*cf.* 5). For external history, however, this proves more difficult, owing in part to the less immediate nature of the subject matter of external history compared to language-internal features. Moreover, the subdiscipline of external history is more recent and has thus not been studied as extensively as internal history; unlike the latter, it has no real tradition of theoretical research to fall back on.

The general periodisation for the history of the Romance languages suggested above (*cf.* 5.2.1) should therefore be considered as provisional. Similarly, five language-

external factors considered by the author to be of determining influence on language use will be distinguished here:

Factors of low complexity:

1. Geographical area and demographics: in which parts of the world is a specific language spoken? By how many persons?
2. Language contact: which other languages evolved in parallel within the same geographical area occupied by a specific language?

Factors of high complexity:

3. Infrastructure (habitat, channels and means of communication) and socio-cultural, political and economic organisation: within which sociological framework did the language in question evolve?
4. Written culture and linguistic thought: to what extent is writing represented? Which textual genres are well developed? What is the extent of alphabetisation and the role of the media in a given society? To what extent has a tradition of linguistic reflection or a 'linguistic culture' evolved? To what degree has a language been standardised?
5. Diasystem: which varieties can be distinguished at a given time in history? What is the importance of these varieties in communication?

Amongst these factors, only 1 and 3 are truly language-external, and, incidentally, also overlap with each other: geolinguistics, for example, focuses in equal measure on aspects of geopolitics, demographics and sociopolitical organisation. The other three factors are situated at the interface between language use and language form. Written culture and linguistic thought can almost be considered as a separate entity due to their self-reflective nature, distinct from both language-internal phenomena *per se* and language-external factors. Moreover, as far as diachronic research is concerned, written culture is inseparable from the diasystem, since an understanding of the latter can only be gained through the study of written sources and any existing metalinguistic evidence.

Although this classification involves a certain degree of internal contradiction, it nevertheless constitutes an effective analytical tool. The five factors will now be considered in more detail:

1 Geographical area and demographics

Initially, external linguistics deals with the identification of the speakers of a specific language. It must determine the geographical area of the language in question and quantify the number of its speakers. Such factors determine the vitality of a language:

too limited an area and too weak a basis of speakers inevitably leads, sooner or later, to the death of the language, abandoned in favour of a variety spoken by a greater number of persons. There are a considerable number of Romance languages that have disappeared (such as those spoken throughout the entire *România submersa* and most French and Francoprovençal dialects), or that are endangered (these include Occitan, Aragonese, Aromanian, Friulian and arguably even Ladin, Romansh, Aranese and Asturian; cf. 3.4).

The enlargement of the geolinguistic basis of a language through migration and geopolitical expansion also has significant consequences for language-internal evolution: it leads to neutralisation or even language standardisation and elaboration, resulting in the diversification of the diasystem. This is the case for the majority of the Romance varieties that have become established as national standard languages (in particular, Spanish, Portuguese, French and Romanian).

The destiny of Castilian is particularly noteworthy. In the 11th century, the primary dialect of Castilian was spoken in a restricted and mountainous territory in the north of the Iberian Peninsula. Two centuries later, as a result of military expansion and political skill on the part of the princes ruling over their Castilian-speaking subjects, as well as of the vagaries of history, Castilian was spoken across half of the Iberian Peninsula, had been adopted by the royal chancery and had become the language of courtly literature. This expansion was accompanied by a process of elaboration that transformed the primary dialect into Spanish, a language of national importance. In the 17th century, this language was spoken across two thirds of the Americas.

Another example is that of the expansion of French royal domination between the 15th and 19th centuries throughout the present-day territory of France (cf. 10.5.1).

As previously mentioned, it is obvious that such changes within geographical areas are inseparable from sociopolitical transformations (cf. no. 3 below).

2 Language contact

Language contact is correlated with both geographical space and social organisation. It is an omnipresent factor in the history of the *România*, where monoglossia is the exception, in contrast to diglossia and polyglossia. In most cases, language contact is a consequence of migrations. Notable examples include the expansion of Latin throughout the area that was to become the Roman Empire, the European colonisation of the Americas and Africa in particular; and, more recently, the increase in the Spanish-speaking population in the United States or in the Arabic-speaking population in present-day France.

Language change caused by language contact is relatively easy to discern, in contrast to other external influences on language (cf. 3.6 and 9.6). Contact situations may remain superficial, but may also lead to the phenomena of mixed languages and pidginisation that transform entire language systems. Only cases of large-scale language contact will be discussed here; the innumerable situations of interference that charac-

terise the historical and modern stages of the Romance languages will not be taken into account.

3 Infrastructure and socio-cultural, political and economic organisation

It is useful to distinguish geolinguistic and demographic features from infrastructural aspects, which are responsible for a large part of social organisation. These are habitat (rural or urban) and its spatial and technological organisation, existing channels and means of communication, and the mobility of individuals within a society.

As an example, the early and pronounced urbanisation of Italy and – to a lesser degree – the Occitan-speaking domain are thus to be contrasted with the rural *langue d'oïl* territory or the Ibero-Romance area. Nomadism in the pastoral Romanian-speaking society led to yet other patterns of language evolution.

Factors tied to demographics and infrastructure have a direct influence on population density, which, in turn, impacts political organisation. Growing population density plays a determining role in the extreme urbanisation of our era that leads to an overarching transformation of sociological factors and to the restructuring of linguistic varieties.

Socio-cultural, political and economic organisation contributes a further set of parameters affecting language development: clan societies and feudal societies, pre-modern societies and colonies, modern or present-day nations all represent models of sociopolitical organisation that incorporate the above mentioned factors of social infrastructure and population density. Each political or economic model – be it founded on agriculture, trading or industry – has specific communicative characteristics that may accelerate or slow down language change. The most striking example is the emergence of modern European territories in the 16th century, which required long-distance communication and generated phenomena of linguistic elaboration and standardisation throughout the *România*.

Moreover, each human society is invested with a 'metaphysical' heritage (religion and ideology) and an aesthetic heritage (the arts, literature, music, architecture, etc.), as well as institutions such as the Church, schools and universities. In this extensive domain, language historians must consider only those aspects that are likely to be truly relevant to their research: specific socio-cultural settings may give rise to specialised vocabulary or even specialised morpho-syntactic patterns.

In the *România*, as elsewhere, religious culture was the starting point for the emergence of specialised vocabulary and specific textual traditions (e.g. biblical texts, psalms or hagiography). Similarly, the development of the sciences and technology resulted in the elaboration of specific vocabulary and, later, highly specialised terminologies and discourse traditions which had a considerable impact on present-day languages (cf. 4.5.1 no. 1).

Institutions such as the Church or schools also modified the way in which both geographical area and society were structured. The boundaries of dioceses or large urban

areas were capable of influencing the frontiers between dialects and the internal cohesion of regional varieties in particular. As a further example associated with linguistic thought, universities led to an increased presence of written culture in 20th-century Europe, and this necessarily had consequences for language.

4 Written culture, linguistic thought and the diasystem

Written culture plays a key role not only in linguistic change but also in general transformations within modern and present-day societies. It was naturally decisive for linguistic elaboration: education, which in Europe was initially a ‘written education’ (in Latin), was fundamental in the formation of the first written Romance languages. Several centuries later, the furthering of widespread literacy among the population enabled the diffusion of standard languages throughout the Western world.

Written culture includes medieval and modern manuscripts and printed texts as well as media production and audiovisual material. For all historical periods, written culture constitutes the only means of direct access to language itself, forming an intersection between external and internal linguistics – the study of written texts is the only way to observe the language of former periods. The interpretive potential of linguistic data is thus absolutely inseparable from philology, the methodology employed for the analysis of texts from earlier periods. The relationship of linguistics to philology may be compared with that of astronomy to the development of telescopes, of particle physics to the technology and methodological implications of particle accelerators, or of radiology to MRI. Observation is intrinsically linked to the instruments which facilitate it, as Heisenberg’s uncertainty principle (1927) and Schrödinger’s cat (1935) have shown with great clarity.

In the area of linguistic thought and language politics, the interaction between language and its context of use is just as striking. Various developments that should be considered in this respect include:

- scientific or pre-scientific thought on language (including grammatical description and lexicography);
- reflection on the part of individuals or institutions on the state of their own language (its ‘beauty’, its ‘usefulness’ or its characteristic features, for example);
- historiography of national languages;
- a ‘language culture’ involving the discussion of questions related to language by specialists and amateurs alike (e.g. popular literature on linguistics or articles on language in the press).

Furthermore, the involvement of institutions to a greater or lesser extent in the organisation of the state may lead to intensive language politics (through the establishment of language academies, for example); in such instances, one speaks of ‘language planning’.

All these factors have an immediate impact on the shaping of standard varieties, on phenomena of inflectional or derivational regularisation and orthography, as well as on syntactic and lexical elaboration.

Finally, the structure of the diasystem is intimately linked to the degree of elaboration of a language, to the linguistic awareness of its speakers and to the development of textual traditions.

5 Observations on the following chapters

When identifying extralinguistic factors, it is extremely difficult to distinguish those that have a truly quantifiable impact from those that remain without significant or lasting consequence for language. Consequently, a careful selection of potentially relevant circumstances must be made within the broad domain of general historiography. On the whole, however, each of the factors presented above plays its role in shaping the physiognomy of a language and in the composition of its varieties at any given time in history. We will thus attempt to highlight the links between external history and the following three domains: internal history (of the standard and non-standard varieties of a language), philology and the history of thought on language (the latter including the history of linguistics as a discipline and a reflection on its methods).

The material will be presented according to the four major ‘external’ periods used in general historiography and described above, with the establishment of an additional period for the present-day era (cf. 5.2.1):

Ch. 10.2: The age of the Roman Empire (7th century BC to 5th century AD)

Ch. 10.3: The emergence of the *România* (6th to 10th centuries)

Ch. 10.4: The *România* during the Low Middle Ages (11th to 15th centuries)

Ch. 10.5: The modern period (16th to 19th centuries)

Ch. 10.6: The present-day period (1880 until today)

For each period, the five factors defined above will be treated consecutively. This will be followed by an additional chapter on the methodology of philology (10.7). We will conclude with some brief reflections on the contribution of external language history and philology to the study of Romance linguistics.

10.2 The Roman period (~ 7th century BC – 5th century AD)

10.2.1 The shaping of the Roman Empire and Latinisation

In Antiquity, the geographical area occupied by Latin was directly correlated with the expansion of the Roman Empire. In the 7th century BC, Latin was still only spoken in a

very limited zone in central Italy, corresponding to approximately half of the modern region of Lazio, south of the Tiber.

Latin is descended from one of the numerous Indo-European languages widespread in Europe since the 4th millennium BC (along with several other non-Indo-European languages). These emerged in the wake of the last major migratory movements that took place during the Neolithic era. The speakers of these languages belonged to tribal, yet sedentary communities that subsisted by means of agriculture. In the early part of the 1st millennium BC, geographically neighbouring languages of Latin such as Umbrian (Indo-European) and Etruscan (presumably Indo-European) had a much broader distribution throughout the peninsula than Latin. Latin was influenced by these languages (*cf.* 9.6.2 no. 1). When it was put into writing in the 7th and 6th centuries BC, it drew chiefly from Etruscan writing, which, in turn, was influenced by Greek (*cf.* 6.4.1).

Harnessed to the Roman expansion, Latin subsequently prevailed as a language of ‘colonisation’ throughout a large part of the Mediterranean and European world. A series of major conquests followed over half a millennium: between the 5th and the middle of the 3rd centuries BC, the territories of central and southern Italy were either subjected to or allied with Rome. In the second half of the 3rd century, these were joined by Sicily, Sardinia, Corsica, northern Italy (*Gallia cisalpina*) as well as the coastal zones of the Iberian Peninsula. In the 2nd century, Rome expanded its influence over the Iberian Peninsula and incorporated Illyria, Greece, the region of Carthage and southern Gaul. By the 1st century BC, Rome had conquered the whole of Gaul – following Caesar’s Gallic wars (58–50 BC) – and the neighbouring territories up to the Rhine, Rhetia, Asia Minor and Syria, as well as northern Africa, east of Carthage. Later followed the conquest of Britain (43 to 80 AD) and Dacia (107 AD).

It is not the task of language historians to explain this spectacular expansion of what was originally a small clan society and its language. But it is obvious that political expansion was facilitated by infrastructural factors such as the Roman road system, the organisation of public and military life, a hierarchical administration, the justice and fiscal systems, schools, and – last but not least – the use of an alphabetic writing system.

The expansion of the Latin language only partially reached Greece and the Hellenised Orient, which at that time already possessed sophisticated material and cultural infrastructure, and accommodated Latin as a second language only (or, more precisely, Latin served as a *lingua franca* for the purpose of cross-linguistic communication). The administration of the Roman Empire was thus bilingual, divided into a Latin or Latin-speaking part and a Greek part, which included the southern Balkan Peninsula in particular, as well as the entire Orient.

Once it had been conquered, the western part of the Empire was Latinised. For the indigenous populations of these regions, the adoption of Latin was a slow and gradual process. The elite were the first to be initiated into the new culture and language; former Roman legionaries then settled (as colonists) in the countryside, roads were built, pro-

tected by outposts and, little by little, a complex administrative network developed. The new subjects thus became accustomed to the Latin language, first as a *lingua franca*, then as a language of everyday communication, until it finally became established as their mother tongue.

Since the Romans did not attempt to impose their language on their subjects through active language politics, the duration and intensity of Roman colonisation determined the extent of Latinisation in different areas. Some regions of the Empire never fully adopted the Latin language, or it was abandoned several centuries after the fall of the Roman Empire. These include:

- Britain
- *Germania II* and *Rhetia II* (i.e. the Germanic and Celtic provinces situated beyond the Rhine)
- a large part of the Balkan Peninsula
- northern Africa

This even applies to the part of Dacia situated north of the Danube – modern-day Romania – where Latin initially almost disappeared (cf. 10.3.2). In some cases, Latin was abandoned following new invasions, e.g. in the region of Carthage (northern Tunisia) and the Upper Rhine Valley, which had previously been Latinised. In hindsight, however, as far as most of the provinces at the periphery of the Empire are concerned, the loss of Latin should be considered as proof that it had never become established as the mother tongue of the majority of the population there. In these cases, the disappearance of Latin can be understood as the consequence of a more superficial colonisation.

Even within the future *Romània*, the duration and, especially, the intensity of Latinisation varied; e.g. the *Provincia Romana* (present-day Provence) or *Baetica* (present-day Andalusia) were more intensely Latinised than, for example, northern Gaul or Dacia. However, aside from these exceptions, the Latin language was able to establish itself throughout the Empire that would later become the *Romània*.

Perhaps the most remarkable element in this process of Latinisation was the development of a highly structured and hierarchical administration. The organisation of the Empire into provinces and sub-provinces proved a very efficient system of exercising power: on the one hand, strong centralisation resulted in constant interaction between the capital and state officials all over the territory, obliging the latter to account for their actions at all levels; on the other hand, the independent organisation of the provinces, governed by efficient mechanisms of autonomous control, avoided overburdening the central administration. Under these circumstances, Latin as a linguistic model was able to spread in a relatively homogeneous and stable manner.

One aspect that remains to be considered is the nature of the linguistic model of Latin propagated in this manner. The centralised structure and geographical extension of the state necessitated the elaboration of a standardised language that could be

understood throughout the Empire. At first it sufficed for the elite of the population to be capable of writing and understanding this language of high prestige. Subsequently, however, the omnipresence of Latin-speaking officials contributed to its wider diffusion and the existence of a written, normative and elaborated language was inevitably accompanied by the emergence of more or less regionalised non-standard varieties.

10.2.2 Language contact and the fragmentation of the *România*

As the Roman Empire became established, innumerable situations of diglossia and bilingualism must have coexisted with each other (*cf.* Adams, *Bilingualism and the Latin Language*, 2003). Even today, minority mother tongues do not disappear in a single generation, despite school education and the media. Works of 4th century AD writers in the Roman Empire bear witness to the continued survival of former indigenous languages whose peoples had been subjugated to Latin rule: the fact that Saint Jerome (347–420) compares the language of the Gaulish ‘barbarians’ of present-day Trier (Germany) with that of the Galatians (of Celtic origin) living in the vicinity of Ankara (present-day Turkey), suggests that Celtic was still present, at least sporadically, at that late point in history. At the same time Saint Augustine (354–430) reports that the lower classes of his native (Roman province of) Africa spoke Punic, notwithstanding the Latin-speaking elite. There also remained notable minority language ‘islands’ such as Basque or Greek in southern Italy. But in most situations of diglossia, with the exception of those in peripheral provinces and Greece, Latin was able to assert itself in the long term, and it was eventually accepted by the population. In the 5th century AD, shortly before the fall of the Roman Empire, Latin had become the common language of communication and the mother tongue of almost the entire population of the future *România*.

Interference occurred to a greater or lesser extent, depending on the intensity of language contact between Latin and the languages of the colonised subjects: Latin borrowed a number of words from Osco-Umbrian, Celtic and Celtiberian (*cf.* 9.6.2 no. 1). Some of these words entered the general vocabulary of Latin, while others were only adopted by its regional varieties. Furthermore, linguistic *substrata* left numerous traces in the place names of Gaul, the Iberian Peninsula and Italy; this constitutes evidence in favour of continuous settlement in these areas (*cf.* 9.7.3).

The multiple relations between Rome and Greece, the administrative bilingualism of the Empire, the former Greek colonisation of southern Italy and the intense cultural and intellectual pressure exerted by the Greeks on the Latin world led to a strong influence of Greek on Latin. This influence is especially apparent in lexis, particularly in elaborated vocabularies such as those pertaining to medicine, biology or rhetoric. It reflects a form of acculturation (enrichment through the acquisition of new cultural concepts) transmitted by the elite, a ‘cultural *adstratum*’ (*cf.* 9.6.2 no. 1).

Some researchers have emphasised the importance of pre-Latin contact languages for the fragmentation of Latin into the different Romance languages, which occurred after the fall of the Roman Empire. This theory culminated in a model that interpreted the evolution of Latin to Romance as a ‘creolisation’ of Latin (*cf.* the work by de Dardel, *A la recherche du protoroman*, 1996). According to this scenario, the Latin spoken by Roman colonisers was not adopted as such by speakers of the *substrata*, but rather pidginised and creolised in the provinces of the Empire (*cf.* 3.6.3 on creolisation). The contact varieties would then have been partially decreolised, surviving amongst uneducated populations during centuries of Latin domination until the fall of the Empire. At this point, in the face of the growing weakness of standard Latin, these supposedly creole languages would have regained vigour and contributed to the shaping of the Romance languages.

Notwithstanding the fact that spoken language is subject to a great deal of variation and change, this hypothesis is by no means convincing. In the first place, orality and ‘creoleness’ are not identical, and, from a typological point of view, varieties of Proto-Romance are in no way to be equated with creole languages (*cf.* 5.2.2). Furthermore, the regions of the Roman Empire in which *substrata* maintained a strong presence (and could thus in theory have had an impact), such as Britain, Mauritania or *Germania Inferior* (the Roman province situated west of the Rhine), simply rejected Latin after the fall of the Empire rather than creolising it. Finally, if this hypothesis were true, the indigenous languages would have left many more traces in the Romance languages than can actually be observed.

It is therefore a hypothesis that contradicts both historical evidence and evidence from comparable situations today. Despite the solidity of arguments advanced to dismantle it (*cf.* Seidl, *Le système acasuel des protoromans ibérique et sarde: dogmes et faits*, 1995, and Kramer, *Sind die romanischen Sprachen kreolisiertes Latein?*, 1999), the evocative terms of ‘creolisation’ and ‘creolised Latin’ achieved a certain amount of recognition, even amongst researchers who remained relatively unconvinced by the theory due to its weak foundations.

The older ‘*substratum* hypothesis’ is more coherent, but nevertheless explains only a few aspects of the fragmentation of Latin. This theory assumes that pre-Latin languages paved the way for the differentiation of the Roman territories from the Early Middle Ages onwards. According to this scenario it is assumed that spoken Late Latin was regionalised to a large degree due to language contact, the effects of which would still have been felt centuries after the abandonment of the pre-Latin languages.

The little that is known about the regional differentiation of spoken Latin does not lend much weight to this hypothesis. Attempts to detect phonological or grammatical changes brought about by *substrata* have failed for the most part. For example, the voicing and loss of intervocalic plosives in *Gallo-* and *Italo-Romance* has been invoked as evidence of interference between Latin and Celtic, but the geographical area occupied by the potential contact languages only partially corresponds to the territory in

which the phonetic phenomenon can be observed. Possible influence of Celtic on the prosody of spoken Latin – which could have had an effect on phonetic development – can never be proven. It is possible that the *substrata* had some influence on the regionalisation of Latin, but such assumptions remain hypothetical and rather improbable (cf. 10.3.3).

The abandonment of pre-Latin languages may be compared to the disappearance of numerous pre-Hispanic languages in the Americas, where Spanish – like Portuguese or English – encountered numerous indigenous languages that for several centuries were supported by a considerably larger demographic basis than Spanish. Here too, however, the majority of situations of *diglossia* led to the abandonment of indigenous languages in favour of the dominant language, which was sustained by strong super-regional infrastructure. Traces of language contact evidenced by regional forms of Spanish were always eliminated after some time owing to the diffusion of Peninsular Spanish as a model, marked by the prestige of the metropolis and spoken by the elite, the administration, the Church and in schools. In the same way, the traces of regional language interference that were present at the time of the Roman Empire must have been gradually eliminated from Latin.

The *substratum* theory stands in contrast to the ‘*superstratum* hypothesis’, which attributes a more significant role to Germanic-speaking groups with regard to the formation of the Romance languages. From the 1st century AD onwards, Germanic peoples lived within the Roman Empire. During the Late Roman Empire, they occupied important positions in the army – where they were recruited as *auxilarii* (auxiliary troops) – as well as in Roman administration.

Proponents of the *superstratum* theory attribute particular importance to the Germanic nations that invaded and overturned the Roman Empire, which had already been weakened internally, between the 4th and 5th centuries. The Germanic peoples established themselves in large parts of the future *România*:

- the Franks in northern France
- the Ostrogoths, followed by the Lombards, in Italy
- the Visigoths in southern France and Spain
- the Vandals in North Africa, which was still Latin-speaking at the time

Their languages, though dominant, could not assert themselves, doubtless owing to the small number of invaders, the absence of sophisticated administrative infrastructure and differing religious beliefs. Therefore, it may be concluded that the Germanic languages did not play a determining role in the formation of the Romance languages (cf. 10.3.3).

10.2.3 Variation and regionalisation of Latin

Among the varieties of Latin, the high-prestige variety used in writing is by far the one best known to us today. It is also the variety that was least subject to geographical variation, especially after the completion of its standardisation in the 1st century AD. Nevertheless, marked non-standard varieties of Latin also existed.

The diastratic variation of Latin, for example, is very clear and accessible. There is ample evidence for spoken and popular Latin after the 2nd century AD, although the number of sources does not by any means rival those bearing witness to the fixed standard variety (*cf.* the comedies of Plautus; this topic is examined in detail in the third part of Adams' trilogy, *Social Variation and the Latin Language*, 2013). Even the metalinguistic witnesses – consisting of authors and grammarians from the time of the Roman Republic – were more interested in the non-standard varieties of (central) Italy than in varieties of Gaul or Spain. It is probable that diastratic variation was more pronounced in Latin than diatopic variation, even in popular forms of the language. A considerable number of phonetic and phonological, morphological and lexical phenomena that characterised spoken Late Latin – in contrast to standard Latin – appear in the majority of Romance languages. This is the case for the following developments, whose widespread presence in the subsequent *Romània* suggests that features of spoken Latin displayed a certain degree of homogeneity:

- in phonetics: loss of vowel length in stressed vowels and of word-final unstressed vowels (with the exception of *-a*), monophthongisation (OE > *e*, AE > *ε*), syncope (HOMINE > *OM'NE > *homme*, *hombre*, followed by palatalisation: OCULU > *OC'LU > *ueil/oeil*, *ojo*, *occhio*), assibilation of C(E/I) (*tsiel* / *ciel* / *cielo*), loss of word-final consonants *-t* / *-m*;
- in morphology: the progressive disappearance of the Classical synthetic future (replaced by periphrastic forms) and the accusative with infinitive, the introduction of the conditional/future-in-the-past, loss of passive voice and the neuter gender, reduction of declension classes and of cases;
- in lexis: a tendency towards expressiveness in several areas of basic vocabulary (derivation, metaphorisation).

This homogeneity was, however, not total, even if the evidence for diatopic variance is not immediately tangible, as the regional variation which formed the basis for the future territorial differentiation of the Romance languages unfortunately left few written traces. Some of the great Latin writers came from Spain (Seneca, Lucan and Martial), North Africa (Apuleius, Saint Augustine) and – after the 4th century – from Gaul (Ausonius, Sidonius Apollinaris), but only a small number of regional elements are to be found in their writing. A significant part of the available data for regional variation is based on inscriptions, especially funerary inscriptions, which represent a source that is nonetheless standardised and characterised by stereotyped formulations. The

comprehensive study by Adams, *The Regional Diversification of Latin* (2007), however, contains sufficient lexical and, in part, phonetic or even morphological evidence from numerous sources to prove that spoken Latin underwent early regionalisation while remaining in the shadow of the highly standardised written variety. As Adams states, “the regional diversity of the language can be traced back at least to 200 BC” (2007: 685) and it intensified during the Late Latin period.

The regionalisation of Latin is also partly substantiated by a certain number of features found in the medieval Romance languages, which can be reliably traced back to Late Latin by means of the reconstruction method. Relevant features include the different vowel systems of Romance (*cf.* 6.2.3), different types of periphrastic future formation (VOLEO CANTARE, CANTARE HABEO, etc., *cf.* 7.6.4) and lexical differentiation. To return to the example of modern colonial languages, Latin may be compared with Spanish: while the varieties of written Spanish in Spain, Peru and Cuba barely differ from each other, spoken Spanish in the Americas or in Africa displays a high degree of diatopic variation.

The development of regional entities took shape during Late Antiquity as the consequence of various factors affecting cohesion, such as geography, ethnic groups, internal migration, economic interaction and the formation of cultural centres. This process may also have included phenomena of language contact that contributed to the intensification of Romanisation. It is reasonable to assume that an infrastructural fragmentation took place, which, despite its weak linguistic impact, nonetheless foreshadowed and formed the basis for future processes of evolution.

- Seidl, *Les variétés du latin*, RSG 1, art. 49
 Herman, *Du latin aux langues romanes*, 2 vols., 1990, 2005
 Carles, *Trésor galloroman des origines*, 2017

10.3 The emergence of the *Romània* (5th–10th centuries)

10.3.1 From Latin to Romance

In the 5th century, Latin was the mother tongue of almost the entire population living within the territory of the future *Romània*. By the 10th century, their descendants spoke Romance languages that displayed notable regional differentiation and that were clearly distinct from Latin, even spoken Late Latin. By what means and under what circumstances could a language have changed to such an extent that it gave rise to several new, distinct languages?

In the first place, linguistic evolution of this particular type, accompanied by the formation of new linguistic territories, is correlated with extralinguistic transformations that are no less spectacular:

1. The main event which contributed to the acceleration of these transformations was the decline of the Roman Empire. This process was initiated by the ever-increasing fragmentation of the Empire, as well as by internal unrest which was growing ever more virulent. The first significant disturbances occurred between 235 and 284 AD. After the death of Theodosius (395) and Stilicho (408), centralised power lost its authority. The deposition of the last Roman emperor Romulus Augustulus by Odoacer (in 476 AD) sealed the fate of the crumbling Western Roman Empire.
2. The political destruction of the Empire went hand-in-hand with infrastructural disintegration of roads, administration, schools and jurisdictions as well as of written culture in general. Once these had been dismantled, the geolinguistic cohesion of Latin was disrupted and local centrifugal forces had free rein. In some regions, such as Gaul or Tunisia, destructive forces immediately began to make themselves felt, whereas in others, including Spain and the Ostrogoth kingdom of Italy, established institutions persisted for a century or two. Nevertheless, a process of disintegration was underway throughout the former Empire.
3. The high degree of territorial fragmentation and the disappearance of overarching factors which had ensured cohesion were greatly accentuated by the early Justinian Plague, which struck the countries of the western Mediterranean at regular intervals between 541 and 770, depopulating entire regions and substantially reducing the mobility of the population.

Secondly, the linguistic process that appears so spectacular in retrospect in no way represents a rapid and revolutionary change: the process of evolution that transformed a spoken variety that was still close to written Latin into a multitude of new languages occurred over the course of eight or nine centuries. A certain number of transformations are inevitable over such a lengthy period: a text written in French, English or Arabic in 1300 or 1500 distinctly differs from a text written in the same languages around the year 2000, making comprehension difficult for present-day readers.

The elements which distinguished the future Romance languages from written Latin appeared in the spoken variety during the 1st century AD at the latest, at a time when the written language had just achieved standardisation. In order to understand these changes which took place in Latin, it is essential to consider the difference between written and spoken language. Spoken language was used continuously over time and space: children still understood their grandparents without difficulty, and at no time did persons living within a day's walk of each other experience difficulties in communicating. On the other hand, the distinction between the standard variety and colloquial varieties, which already existed in the 1st century AD, continued to increase over time. At a certain point, the illiterate population was no longer able to understand written Latin when read aloud: the chain of so-called 'vertical communication' had been broken (the French term *communication verticale* was introduced by Banniard, *Viva voce*, 1992). The inability of 98% of the population, who were illiterate, to under-

stand Latin marked the severance of the link between written Latin – the language of prestige and culture – and the spoken Romance varieties, as well as the beginning of a true situation of diglossia.

This process was not entirely independent of the changes in written Latin, which underwent a gradual differentiation from Classical Latin, particularly from the 7th century onwards. The chronology is consistent with the periodisation assumed by classical philologists, according to whom the Latin period (as opposed to that of medieval Latin) extends as far as 636, the year of Isidore of Seville's death, which is accordingly considered as the cut-off point by the authors of the principal Latin dictionary, the *Thesaurus linguae latinae* (ThesLL).

To a large extent, the transformations were due to the disintegration of educational institutions and the resulting decline in the knowledge of classical norms. In reaction, movements aiming at achieving a 'Renaissance' of Latin intended to reconcile written Latin and/or its pronunciation with the assumed classical norms of Antiquity began to take shape in the 9th century. But such movements only served to broaden the gap between written Latin (or Latin read aloud) and spoken Latin, and to accentuate the fact that by then, Latin and the Romance languages already constituted two distinct realms.

Due to the absence of sufficient source material, caution should be exercised in assessing the evolution of spoken languages. It is nevertheless plausible that during a given period, entire groups of innovative linguistic features were restructured within new systems. This type of accelerated macroevolutionary trend in the transformation of language may have taken place over a span of a mere three or four generations. According to current hypotheses, the crucial period is situated between the 7th and 8th centuries. Banniard suggests the one hundred year period between 650 and 750 (*cf. La délimitation temporelle entre le latin et les langues romanes*, RSG 1, art. 51), and we have seen that Chambon considers 700 AD, or thereabouts, as a decisive date, as it marks the grammaticalisation of the definite article (*cf.* 5.2.1). The ultimate reasons for this restructuring of spoken language may also have resided in the weakening of the old written norms: once these had been abandoned, spoken language could evolve unfettered.

By 1000 AD, the die had already been cast: the new multitude of speech varieties had become differentiated to a significant degree from the spoken Latin from which they originated, but also from one another, perhaps even to the point that the speakers of these early Romance languages were no longer able to understand each other.

In addition, since the reforms of Latin at the beginning of the 9th century, contemporaries were aware of a gap between a learned language, written and spoken in restricted circles (and used in public on formal occasions) and a mother tongue, almost exclusively spoken, serving as the language of communication and everyday use (*cf.* 10.3.5 no. 4).

10.3.2 Geography: delimitation of the future Romance territories

In the history of the Late Roman Empire, a distinction should be made between its Western and Eastern halves. Germanic invasions of the 5th and 6th centuries AD and, shortly thereafter, the arrival of Slavonic peoples in the east (6th century) and Arab-Berber armies in the west (7th to 8th centuries) concerned, above all, Latin-speaking parts of the Western Roman Empire. The Greek-speaking Eastern part, in contrast, would preserve its political, intellectual and cultural structures until 1453. Additionally, though the Greek language naturally changed during the Byzantine era and the Modern period, vernacular spoken Greek – known as Demotic Greek or *dimotiki* (“language of the people”) – never witnessed internal or geographical diversification comparable to that of Latin. The contrast between East and West thus emphasises the transformations undergone by Latin under the impact of geographical, demographic, infrastructural, political and socio-cultural factors.

As has been mentioned, Latin disappeared from use in peripheral provinces of the Empire between the 5th and 10th centuries, enabling former indigenous languages to regain dominance. These include:

- Celtic and the Germanic language of the Angles (who settled in England in the 5th century) in Britain
- Basque on either side of the western Pyrenees
- other Germanic languages in the interior territories of *Germania*
- Albanian in the province of Illyria
- Berber, and later, Arabic, the language of the new colonisers, in Tunisia

Other territories, on the contrary, were de-Latinised under the impact of invasions: these include what are today the Flemish part of Belgium, the Saarland, Alsace and German-speaking Switzerland as well as the source region of the Danube and specific parts of the Balkans. Latin left few traces in these geographical zones, even though some enclaves, including Tunisian cities and some wine-growing villages of the Moselle (the so-called ‘*Moselromania*’) subsisted for centuries.

Latin thus survived as a spoken language throughout a relatively coherent geographical zone, which encompassed almost the entire Iberian Peninsula, most of the territory of Gaul as well as the Italian Peninsula and the Alpine zones (*cf.* 3.2). The only continental language frontier of this area was situated in the north, where it bordered on the Germanic-speaking zone. The most important border regions on the Romance side were – using modern geographical designations – Wallonia, Romance-speaking Lorraine, Franche-Comté, Romance-speaking Switzerland, Ticino, Grisons, and the linguistic zones of Venetan and Friulian. In the Eastern Alps, the Romance languages were also in contact with Slavonic languages. This frontier took shape during the 5th to 7th centuries, first covering a relatively broad area that included numerous ‘linguistic islands’ and bilingual villages, later forming a more linear frontier that would remain in place

until approximately the year 1000. It would thereafter remain more or less unchanged until the beginning of the 20th century.

The remaining boundaries of the future *Romània* were shaped by the Atlantic and Mediterranean coastlines. Their nature as boundaries must have been reinforced by the Arabo-Muslim expansion that extended over large parts of the Mediterranean basin from the *Hijra*³⁴ onwards, making maritime navigation hazardous until the 10th century.

Within this extensive Romance area, some non-Latin-speaking enclaves have survived until today, including, in particular, the zones occupied by Basque (the northern and eastern boundaries of which once extended further than they do today), Breton (originating from a South-western Brittonic language – more specifically, the variety spoken in Cornwall – and established in Brittany around the 5th and 6th centuries), and Greek (in some parts of southern Italy: Salento, Calabria and Sicily).

Only the future Romanian-speaking area in the East and – to a much lesser extent – the major Mediterranean islands (Sardinia, Corsica, the Balearic Islands and Sicily) remained isolated from the *Romània continua*, which formed a geographical unit from the 6th to the 15th century.

As already mentioned, differentiation within the Romance territory was favoured by the demographic decline that followed the disintegration of the Roman Empire and the Justinian Plague. It may be assumed that a population density distinctly lower than that of the Late Roman Empire persisted for centuries.

As an example, estimated figures for the territory of modern France are as follows: 12 million inhabitants in the 4th century, 5 million at the end of the 5th century, 8 million under Charlemagne's rule (8th c.–814), 7 million around the year 1000, 10 to 12 million again in the 13th century and at the end of the Hundred Years' war (1337–1453), 20 million at the beginning of the 14th and in the middle of the 16th century. During the 7th and 8th centuries, the founding of new villages and hamlets, which can be dated fairly accurately with the help of archaeology and onomastics (i.e. the study of place names), bears witness to a certain degree of demographic revival, which became more significant during the 11th and 12th centuries. One can assume that around 450 AD and 1200 AD, population numbers and density were roughly equal across the whole of the Latin-Romance territory.

Low population density thus played a complementary role, alongside the crumbling infrastructure and the extremely limited role of writing, in furthering macroscopic geolinguistic fragmentation.

Geographical factors that were of most significance for the shaping of new geolinguistic entities seem to have been important mountain chains:

³⁴ The '*Hijra*' denotes the emigration of Muhammad and his followers from Mecca to Medina in 622 AD, which marks the beginning of the Islamic calendar.

- the Pyrenees separate what are today the Ibero-Romance and the Gallo-Romance areas, even though Catalan, which is open to contact with the coastal region of southern France, both by land and by sea, represents an important link in the continuity between the two;
- the boundaries between Occitan, northern dialects of Italian and Francoprovençal, as well as those between Romansh and Ladin are all situated in the Alps;
- the Massif Central plays a role in the formation of the boundary between French and northern Occitan and between northern and southern Occitan;
- the northern border of the Apennines separates Gallo-Italian dialects in the north from dialects of central and southern Italy (this is known as the La Spezia-Rimini line *cf.* 3.3);
- similarly, the frontier between Portuguese and Spanish is marked by mountainous zones in the north.

The extent of the influence of mountains on linguistic frontiers nevertheless varies greatly:

- in some cases, linguistic frontiers coincide precisely with mountain chains, such as that between the dialects of northern and central Italy; in other cases, they pass through the middle of a mountain range, such as that between Spanish and Portuguese;
- sometimes the mountains do not separate two linguistic zones but have given rise to a distinct linguistic zone in their midst (as in the cases of Romansh, Ladin or Francoprovençal); the Pyrenees are surrounded and traversed by a conglomerate of transitional varieties comprising Gascon (including Aranese), southern Occitan, Catalan and Aragonese;
- mountain height is not a determining factor in itself. Varieties of eastern Occitan and dialects of north-western Italy, although separated by the Alps, share a number of typological features that set them apart from other Italo-Romance and Occitan dialects; the mountain range separating the Portuguese and Spanish zones on the Iberian Peninsula is not very high, but is nevertheless broad and even today remains relatively inaccessible;
- finally, frontiers may alter over time: between the 8th and 12th centuries, the boundary separating the *domaine d'oïl* from the *domaine d'oc* appears to have shifted further south, from the Loire to the Massif Central.

Mountain ranges thus do not constitute limits as such, but rather natural fracture zones, around which phenomena of linguistic differentiation may develop over the centuries (*cf.* 3.2, fig. 4 above).

Geolinguistic factors are more difficult to determine in the case of Romanian. Researchers are confronted with a total absence of medieval sources (the earliest texts, gener-

ally written in Cyrillic script, date from around 1500), and the historiography of the language is controversial and tainted by ideology. During the Roman period, only the western half of present-day Romania was Latinised (Banat, Oltenia and Transylvania). Moreover, this Latinisation was short-lived (lasting only from 106 to 271–272), in all probability remained partial, and was further weakened by the arrival of the Slavonic peoples in the 6th century (cf. Dahmen, *Externe Sprachgeschichte des Rumänischen*, RSG 1, art. 66, and Fischer, *Les substrats et leur influence sur les langues romanes: la Romania du Sud-Est*, RSG 1, art. 33).

On the other hand, a large zone south of the Danube had been Latinised more thoroughly and for a longer period of time (until the 5th century). During the Slavonic invasions, this region also received Latin-speaking refugees from north of the Danube. It is thus probable that old Romanian developed chiefly in this zone, between the 6th and 10th centuries (cf. Kramer, *Sprachwissenschaft und Politik. Die Theorie der Kontinuität des Rumänischen*, 1999–2000). Nonetheless, this does not exclude the existence of Latin-speaking and (late) Romanian-speaking groups north of the Danube, even though, initially, these could not have been numerous (cf. Dumniștrăcel/Hreapcă, *Histoire des dialectes dans la Romania: Romania du Sud-Est*, RSG 3, art. 212).

Settlement in Romania later intensified, through migration and demographic growth, on the territory of the present-day country. From the 12th century in particular, Romanian, which was mainly spoken in the mountainous regions, was reinforced by immigration from the south, and in the end, prevailed over Slavonic. In the 16th century, Romanian replaced Old Church Slavonic as a written language, and later became the national language.

The south Danubian varieties of Romanian remained restricted to spoken use, their speakers dispersing in Greece (hence the existence of Aromanian and Megleno-Romanian groups there, cf. 3.4.13). For ideological reasons, the Greek nation long denied the existence of these minority languages, despite there being approximately half a million speakers of Aromanian at the beginning of the 20th century (for a general overview cf. once again Dahmen, RSG 1, art. 66).

Romanian thus displays significant differences compared with the other Romance languages:

- it is geographically isolated from the *România continua*;
- it evolved under the domination of the Eastern Roman Empire and the Byzantine Church, whereas all other parts of the *România* were subjected to the Roman Church;
- it was strongly influenced by Slavonic contact languages (as well as by Greek and Albanian), and was not in direct contact with either Germanic or Arabic languages;
- its current linguistic territory, which is the result of medieval migrations, only became established at a later period; Romanian cannot thus be considered as a language truly belonging to the *România continua*;

- there are no direct medieval sources for Romanian (it will therefore not be treated in the following section [10.3.5] on written culture).

These peculiarities make Romanian a valuable counter-example: when contrasted with the other Romance languages, it enables a better understanding and evaluation of the characteristics of language evolution in the contiguous zones of the central and western *România*.

10.3.3 Migration and language contact: the Germanic and Arabic *superstrata*

Germanic invasions in the territory of the future *România continua* (cf. 10.2.2) have been the subject of many studies, in particular in the 1930s, when pan-Germanic ideology was at its height³⁵. The scholars of the time assumed that the organisation of the Germanic kingdoms and their contact with the languages encountered within the territory of the former Roman Empire led to the regionalisation and differentiation of Late Latin, giving rise to the Romance languages. This so-called ‘*superstratum* hypothesis’, however, could no more hold up to scrutiny by linguists and historians than could the *substratum* theory.

Generally speaking, a powerful Lombard kingdom followed the Ostrogoth period in Italy; in northern Gaul a Frankish kingdom emerged, while southern Gaul was occupied by the short-lived Visigothic ‘Kingdom of Toulouse’ (466–507), which was then displaced to the south towards the Iberian Peninsula (‘Kingdom of Toledo’, 507–711). Population displacements such as these presaged the formation of new territories, for which the ground had nevertheless already been partially laid during the Late Roman Empire. With the exception of some regions of northern France, the Germanic-speaking percentage of the population remained low everywhere. A more detailed observation of the situation necessitates the following precisions:

- Contrary to earlier assumptions, the Burgundians, who played a minor historical role, were not responsible for shaping the geographical area of Francoprovençal. The emergence of Francoprovençal can be better explained by geographical factors and more particularly by the centralising role of Lyon (cf. 10.3.3).
- The density of Frankish settlements in the future *langue d’oil* territory varied according to region. They were particularly numerous in the north-eastern part of present-day France. If the *superstratum* theory were accurate, specific patterns of linguistic differentiation would have emerged, of a far more profound nature than the simple dialectal variation which actually ensued.

³⁵ Cf. the groundbreaking works by Wartburg, *Die Ausgliederung der romanischen Sprachräume*, 1936 and Gamillscheg, *Romania Germanica*, 1934–1936 (cf. also the synthesis on linguistic historiography provided by Schlemmer 1983).

- The Lombard kingdom included territories between the Po Valley and Calabria. The fundamental frontier between Gallo-Italian and central-southern Italo-Romance dialects (the La Spezia-Rimini line) passes through these regions; such a frontier could not have become established, nor could it have been maintained, had the extension of the Germanic kingdom had the linguistic impact that has been attributed to it.

From an internal point of view, the Romance languages concerned were enriched by numerous borrowings, particularly in the case of French and Italian. Wartburg dedicated three of the twenty-five volumes of the FEW to Germanisms, which is a considerable amount, even though it is known that the author of this monumental dictionary placed particular emphasis on the treatment of Germanic etyma (cf. 9.6.2 no. 2). By way of contrast, the number of Germanisms in Spanish remained low despite two whole centuries of Visigothic rule.

Furthermore, as mentioned above, the Germanic languages had a strong impact on anthroponymy (personal names): in the 10th century, 90% of personal names in certain parts of Gaul, Iberia and Italy were of Germanic origin. Such a high frequency cannot be explained by particularly intense linguistic influence, but rather by the desire for the imitation of names borne by members of the ruling classes (cf. 9.7.2).

In the language-internal areas of phonetics, morphology and syntax, in contrast, the influence of Germanic languages is almost non-existent (cf. Tagliavini 1982, § 54; Felixberger, *Sub-, Ad- und Superstrate und ihre Wirkung auf die romanischen Sprachen: Gallo-Romania*, RSG 1, art. 55).

Though the Germanic peoples had a tangible influence on the destiny of the Romance-speaking population, these *superstrata* did not profoundly transform the emergent Romance varieties. The phenomenon of Germanic *superstrata* may be compared to the relationship that was established in northern Africa between French on the one hand, and Arabic or Berber on the other: military invasions led to the domination of a foreign language spoken by a minority group. As in the case of the medieval Germanic languages, the impact of French was not strong enough to eliminate or restructure the indigenous languages. At the most, the latter may have borrowed a number of terms from the dominant languages.

The Arab-Berber conquest of the Iberian Peninsula ensued slightly later, but still during the period of the regional differentiation of spoken Late Latin. It occurred within the broader context of the Arabo-Muslim expansion. Between 711 and 780 AD, the Arabs spectacularly shattered the Visigothic kingdom of Toledo, establishing their own domination throughout almost the entire Iberian Peninsula (*Al-Andalus*). Only the rather mountainous northern fringe remained under the authority of Catholic sovereignty.

The Muslim conquests were of great importance for the history of the Iberian Peninsula, owing primarily to the multiple cultural and linguistic exchanges between the Arab and the Ibero-Romance populations to which they gave rise over the course of

several centuries. From a language-internal point of view, Arabic played the role of a *superstratum* on the Iberian Peninsula. Similarly to the Germanic conquests, however, this did not result in profound modifications to the (Proto-)Romance languages with which it was in contact: Arabic left traces in lexis and in place names, but not, for instance, in personal names (*cf.* 9.6.2 no. 3 and 9.6.3); Berber, a language of low prestige among the conquerors, albeit themselves Berber-speaking for the most part, left no traces.

The most important linguistic legacy of this period is to be observed in the consequences of the ensuing *Reconquista*, during which the varieties spoken in the far north were imposed throughout the Peninsula. The continuity of locally spoken varieties of Late Latin was disrupted by this expansion (*cf.* 10.4.1 no. 1).

10.3.4 Political factors and infrastructure

The Germanic, Arab and Slavonic conquerors left their marks on the Romance territory – less in terms of their languages, but more as a result of the boundaries of the territories under their influence, their political organisation, as well as their material and intellectual structures. These invasions destroyed nearly all factors contributing to geographical cohesion over large areas within the Western Empire, including links to the centralised political and administrative centre, the infrastructures that had enabled travel and commercial exchange across all provinces, as well as the educational system of the elite, founded on a single language.

Some factors of continuity nevertheless withstood this widespread process of disintegration: established rural and urban habitats survived, as did an urban patrician elite in specific regions, albeit sporadically. Some channels of interregional communication (e.g. trade) were upheld, and the persistence and development of a monastic network conserved a rudimentary culture of written Latin. Finally, the increasing power and influence of hierarchically-organised ecclesiastical institutions gradually extended across the entire area of the former Western Empire and were involved in administration at all levels (abbeys, the first rural parishes, dioceses and the new urban and courtly centres that began to develop during this period).

Together with the surviving fragments of the disintegrated Empire, the new political regimes and the Church, whose influence was increasing rapidly, contributed to the shaping of new regional linguistic zones. The influence of political factors on the regional differentiation of the Romance languages varied greatly:

- Geographical delimitation of Romance varieties in the Iberian Peninsula was shaped by the Arab conquests and the Christian *Reconquista* (*cf.* above and below).
- Linguistic proximity between Catalan and Occitan was doubtless favoured by the fact that the area of present-day Catalonia belonged to the Carolingian Empire (constituting a zone known as the *Marca Hispanica* or ‘Spanish March’). Its proximity

to Occitan also explains the particular status of Catalan in the Iberian Peninsula (*cf.* 3.4.10)

- The absence of a rupture in dialectal continuity at the northern limits of the Apennines can probably be explained by the fact that the Lombard kingdoms extended on either side of this natural frontier (*cf.* 10.3.2). It should be noted that the opposite scenario could have led to the development of two separate written languages, as in the case of French and Occitan.
- The important urban centre of Lyon and its network of rivers (the Rhône, the Saône and the Isère in the south, the Ain and the Doubs in the north) probably acted as a focal point for the development of Francoprovençal.

While it seems obvious that political factors may determine the regional differentiation of languages, it is more difficult to discern the possible connections between such external factors and the internal evolution of a language. In this respect, French appears to be an isolated case: compared with the other Romance languages, it is characterised by particularly well-developed phonetic innovation, which, furthermore, began at an early stage (*cf.* 6.5). The rapid pace at which spoken language on the territory of northern Gaul became differentiated from written Latin – at least in terms of phonetics – may have been favoured by the development of infrastructural networks under the Merovingians and Carolingians, which were more dense there than elsewhere in the *Romània* and which were bolstered by the absence of mountainous terrain. The relative ease of circulation resulting from these conditions within the realm could then have facilitated and accelerated the changes.

This interpretation remains hypothetical; were it to be confirmed, however, it would point to two complementary phenomena: on the one hand, a low population density, which (through a relative lack of communication) would have favoured language rifts at certain strategic points; on the other, a degree of infrastructural cohesion that would have increased the dynamics of change within a given territory.

The period between the 6th and the 9th centuries AD is characterised by relative geographical stability of the population. Peasants were tied to their land, and any exchanges primarily took place within short distances. This situation, in turn, contributed to relative stability of new phenomena of language evolution within defined geographical limits.

The interruption of long-distance communication must also have favoured the development of features relating to linguistic immediacy and thus of the tendency towards expressivity which characterised the emergent Romance languages (these frequently displayed diminutive suffixes, morphosyntactic constructions tending towards semantic explicitness and a possible acceleration of sound change; *cf.* the respective sections in Part Two of this manual).

10.3.5 Written culture and linguistic thought

Our interpretations of language development during this period are based on fewer and less reliable sources than those available for other historical periods. Written culture and (meta-)linguistic thought played a less prominent role in society in the 8th century than in the 1st or the 14th centuries. Owing to the scarcity of written sources, particular effort is required in order to detect and interpret the very first signs of the existence of the new Romance languages. Four main categories of data can be taken into consideration:

1. Written Romance *per se*, which provides direct evidence relating to the internal evolution of the languages in question. It indicates an awareness of the existence of Romance languages, and of the fact that they differed from Latin.
2. Pre-textual Romance elements embedded in Latin texts and reflecting innovation (such as innovative graphic elements, new lexemes, place names or personal names).
3. Different ‘registers’ of writing: normative Latin, rustic Latin and vernacular Latin (to which reforms of orthography and pronunciation bear witness).
4. Metalinguistic evidence that points to speakers’ awareness of the diglossic situation (i.e. differences between written and spoken language, with Latin being used for writing and Romance for spoken communication) as well as their awareness of diasystematic variation in writing.

1 Written Romance

The earliest evidence of writing that can no longer be qualified as Latin dates from the 8th to 10th centuries. Two forms of evidence should be distinguished: (1) a very limited number of microtexts or text passages, often short, and (2) fragmentary Romance elements which very frequently appear within Latin texts (*cf.* no. 2 below). The two types of evidence display a similar linguistic physiognomy, but the microtexts stand out more clearly within their surrounding Latin context; for this reason they were identified early on and studied with particular attention.

In general, Romance microtexts can be ascribed to oral situations of communication and they are characterised by strong pragmatic connotations. Though limited in number, reduced in length and displaying inherent difficulties of interpretation, this type of evidence has convinced language historians – more so than any other – of the existence of Romance languages during this historical period. Between the middle of the 8th and the end of the 10th centuries, early forms of French, Italo-Romance, Occitan, Spanish and even Romansh appeared in writing, thus constituting a *terminus ante quem* for the transition from Latin to Romance.

For French and Italo-Romance, there are hardly more than ten original sources, dating from between 750 and 900 AD. These include:

- a parody of the ‘Lex Salica’ (ca. 751–768), which contains Italo-Romance lexemes and expressions inserted into a Latin source text: *et ipsa cuppa, frangant la tota, ad ipso botiliario frangant lo cabo [...]*;
- the Indovinello veronese (“Veronese Riddle”) (ca. 760–780, northern Italo-Romance), strongly Latinising with regard to its graphemic elements but Romanised from a lexical point of view: *separeba boves, alba pratalia araba & albo versorio teneba & negro semen seminaba* “I separated my fingers (“oxen” being a metaphor commonly used amongst scribes to designate fingers), in order to hold the plume (= “plough”, metaphorical use), in order to trace on the blank page (= “plough the field”) the black ink (= “sow the grain”), cf. Hilty, *Das Indovinello veronese ist kein Rätsel*, 2001;
- the *Laudes regiae* of Soissons (ca. 784–789), essentially written in Latin, contains the Gallo-Romance benediction formula *Tu lo(s) iuva* “You (= God) will help him”: the verb (< Lat. *IUVARE*) is rare in Gallo-Romance and represents one of the lexemes that are mainly attested during the early Romance period (cf. FEW 5, 92a);
- the Glosses of Kassel (ms. ca. 810) and, more relevant, the Glosses of Reichenau (written in the early 9th century [ms. 1st half of the 10th century], probably in Picardy), which display strong interference between Latin and Gallo-Romance;
- examples of graffiti in the Catacomb of Commodilla in Rome (mid-9th century, Italo-Romance);
- the Strasbourg Oaths (*Serments de Strasbourg*, [842], possibly written in the 2nd half of the 9th century). Despite their late transmission in a 10th century manuscript, they are conventionally cited as the earliest continuous Romance text, which begins as follows: *Pro Deo amur et pro christian poblo et nostro commun salvament, dist di in avant, in quant Deus et podir me dunat, si salvarai eo cist meon fradre Karlo [...]*;
- the Sequence or Canticle of Saint Eulalia (*Séquence/Cantilène de Sainte Eulalie*, 9th century, Gallo-Romance), which is the first text written in ‘true’ Old French (and, indeed, Romance) conserved in a pre-10th century manuscript: *Buona pulcella fut Eulalia / Bel auret corps, bellezour anima [...]* (cf. the facsimile depicted in 11.2.4);
- fragments of a Gallo-Romance sermon on Jonah (*Sermon sur Jonas*, 1st half of the 10th century), conserved in the municipal library of Valenciennes, MS 521 (10th century);
- the Italo-Romance *Glossario di Monza* (1st half of the 10th century), a Greek-‘Romance’ glossary encompassing 66 terms; some of the ‘Romance’ forms are Latin, while others are Italian.

The earliest textual evidence for other Romance languages dates from the second half of the 10th century:

- Traces of Occitan are present, for instance, in healing requests in the *Breviary of Alaric* (*Breviarium Alarici*), in the *Augsburg Passion song* ('Augsburger Passionslied') and in the so-called bilingual *alba*, containing a repeated refrain in Occitan.
- Evidence of Spanish can be found in the *Nodicia de kesos* ("inventory of cheeses") in Ibero-Romance *kharjas*, a form of poetry written in Arabic or Hebraic characters, and possibly in the *Glossae Aemilianenses* and *Silenses*, although the latter may, however, date from the 11th century.
- The earliest sentence thought to reflect a form of early Romansh is found in the so-called *Würzburg Manuscript*.

Other (micro-)texts in Old French and Old Italo-Romance appear during the same period (including approximately 50 texts in Italo-Romance according to Petrucci, *Il problema delle Origini e i più antichi testi italiani*, 1994, cf. 10.4.3 no. 1). If the earliest Catalan and Sardinian sources, which date from the 11th century, are also taken into account, seven of the principal future Romance languages were already represented in writing during this period (cf. the mostly complete inventory of early texts in the five-volume *Inventaire systématique des premiers documents des langues romanes* (InvSyst), 1997, which does not, however, include texts transmitted by later copies). As for sources of Galician-Portuguese (still a single language at that time, cf. 3.4.11 and 12), the situation is complicated and may still hold some surprises; Romanian, meanwhile, remained in the shadow of Slavonic throughout the whole of the Middle Ages.

These early sources mark the beginning of a real culture of Romance writing. Inserted for the most part into Latin texts, or consisting of glossaries and lists of raw lexical material, these Romance elements are often imbued with marked pragmatic functions (such as in the case of oaths, in particular), and they reflect an oral conception (such as rebuses, graffiti or lists; cf. Koch, *Pour une typologie conceptionnelle et médiale des plus anciens documents [...] des langues romanes*, 1993). They also maintained strong ties to extralinguistic reality (e.g. in the form of legends accompanying illustrations). Moreover, they represented extremely short utterances: vernacular glosses that accompany Latin texts (typically of religious content), inscriptions (mostly engraved in stone), short texts associated with images, legends on seals and coins and other microtexts (cf. InvSyst ch. 1; TypSources fasc. 35 (monument inscriptions and graffiti), 36 (seals and matrices), 20 (heraldry and coats of arms), 21, 42 (coins and medals); cf. also RSG 2, 171: *Les premiers documents en langues romanes*).

From an internal point of view, these linguistic documents are difficult to localise with precision: some contain rather marked regional features (the *Indovinello veronese*, for instance, reflects Venetan, the *Glossary of Monza*, Lombardian or Emilian, and the *Augsburg Passion song*, a variety from south-western France); others are more neutral (such as the *Sequence of Saint Eulalia*, the *Sermon on Jonah* or, to an even greater degree, the *Strasbourg Oaths*).

The relative diatopic neutrality of these sources cannot be accounted for by the still weak dialectal differentiation of the Romance languages, but rather by the fact that writing, by its very nature, tends towards communicative distance, even when reproducing oral texts. Rendering oral Romance texts in writing thus does not reflect a phenomenon of linguistic spontaneity, despite the importance of oral elements and linguistic immediacy. On the contrary, by producing what was already a neutralised, semi-artificial language, scribes paved the way for a traditional culture of writing.

Though the earliest evidence of Romance languages is of paramount importance as it highlights the process by which these new speech varieties were first put into writing, these sources nevertheless only bear witness to a short episode within a period of transformation that lasted for almost a millennium. The much greater number of texts dating from the 12th to 15th centuries brought further essential changes and innovations. The importance accorded to the oldest Romance texts in the educational canon of Romance linguistics since Tagliavini (1949) and his predecessors³⁶ should therefore be critically re-evaluated.

2 Pre-textual Romance elements in Latin texts

Latin sources were the first to provide evidence of the Romance languages before the appearance of complete early Romance texts. From the beginning of the 9th century, Latin texts contained numerous Romance elements in fragmentary form (*cf.* Carles, *L'émergence de l'occitan pré-textuel*, 2011 and *Trésor galloroman des origines*, 2017). Some of these elements concern parts of lexemes (including graphemes representing phonetic changes proper to Romance), while others concern entire lexemes (including the introduction of lexemes that are not attested in Classical Latin) or even larger syntagmatic entities (thus reflecting morphosyntactic or lexical changes in spoken varieties; *cf.* Herman, *Du latin aux langues romanes* I, 1990, and II, 2005).

There have been several projects that aim to compile catalogues of Latin documents from Late Antiquity and the Early Middle Ages:

- For the Late Antique period, the epigraphic collection provided by the *Corpus Inscriptionum Latinarum* (CIL, 1863–) (which includes inscriptions dating from up to the 7th century AD) is indispensable.
- The *Chartae Latinae Antiquiores* (CLA, 1954–1998) groups all original Latin charters dating from before 800 AD, providing diplomatic transcriptions (*cf.* 11.3) and facsimiles of all documents (*cf.* 11.2.3); note that vernacular elements were still almost non-existent during this early period.

³⁶ Above all Savj-Lopez, *Le Origini neolatine* (1929), which Tagliavini used as a basis for his homonymous work.

- The *Patrologia latina* (Migne, 1879–1890), a vast collection of early Christian texts, is an important source for the medieval period; unfortunately, the philological shortcomings of this early work mean that it cannot serve as the basis for a reliable analysis of fragmentary Romance elements.
- The most useful collections include texts from the period after 800 AD, beginning with a series of electronic corpora based on transcriptions of original charters:
 - the *Corpus des chartes originales antérieures à 1121 conservées en France* (cf. Giraud *et al.*, 2010; Courtois and Gasse-Grandjean, 2001; Tock, *Scribes, souscripteurs et témoins dans les actes privés en France*, 2005: 17–55);
 - the *Thesaurus Diplomaticus*, which collects all Latin documents written before 1200 pertaining to the geographical area of present-day Belgium (ThesDipl, available on CD-ROM);
 - *Corpus documentale latinum Cataloniae* (CODOLCAT, 2008).
- Moreover, there is an important series of cartularies, conserved in late copies (16th–17th centuries), but which can be used with caution for the purposes of linguistic analysis (cf. 11.1.5; 11.2.5).

Fragmentary Romance elements mainly occur in place names and personal names. Unlike common nouns, proper nouns often preserved elements of their spoken form within the context of written Latin. This was probably due as much to the necessity for identification and authenticity as to the difficulty of (re-)Latinising proper names. Toponyms provide evidence for most processes of phonetic change as well as numerous phenomena of lexical and derivational change, since proper nouns are often derived from common nouns (consider names such as *Legros*, *Cornu* or *Grassi*; cf. 9.7.2).

This type of onomastic evidence, which remains largely under-exploited by research, is a sign both of the beginning of the dominance of the new Romance varieties and of an underlying awareness among speakers of the diglossic reality: in the absence of an immediate Latin model, scribes chose (on a case-to-case basis) to translate Romance names into Latin, to transcribe them in their original (vernacular) form, or to create mixed forms composed of both Latin and Romance elements.

3 Normative Latin, Rustic Latin and Vernacular Romance

Written Latin was not uniform during this transitional period. Whereas some texts remained relatively close to the models of Classical Antiquity, others complied to a greater degree with the phonetic, morphological and lexical changes taking place in spoken language. The disintegration of the Classical model during the Merovingian period (mid-5th to mid-8th centuries) triggered important reforms of written Latin in the 9th century, known as the ‘Carolingian Renaissance’. These reforms were aimed at re-establishing and reinvigorating Classical Latin, as it was useful for the administration of the Empire.

Parallel to the re-establishment of Classical norms, the Carolingian Renaissance also introduced fragmentary vernacular elements (cf. no. 2 above) and intensified the use of another, more 'Romanised' form of written Latin, the '*scripta latina rustica*' (cf. Sabatini, *Dalla 'scripta latina rustica' alle 'scriptae' romanze*, 1968). The *scripta latina rustica*, moreover, never appears in the form of complete texts, but is inserted into normative contexts, in much the same way as the vernacular elements discussed above. These three language 'registers' thus often appear interwoven.

The sentence *Et de istos missales que hic sunt scripti, abet Bego episcopus misal i.* (984–1010, *Breve de libros*, Clermont-Ferrand), exemplifies the following Romance innovations:

- a phonetic feature of Occitan (the loss of *h*- in *abet*);
- the insertion of a full Occitan word (*misal*);
- the presence of the 'syncretic' oblique case (which groups together several former cases, with the exception of the nominative) after the preposition *de* in *istos missales* and in the personal name *Bego* (rustic Latin);
- a V2-constituent order approaching that of Romance (cf. Modern French "et des missels qui sont nommés ici [...]");
- the presence, alongside these innovative elements, of normative forms (*et, de, que hic sunt scripti, episcopus*).

Cf. the commentary by Chambon, *L'identité langagière des élites cultivées d'Arvernie*, 1998 (the above example is found on p. 404).

The use of 'rustic' Latin was thus comparable to that of the Romance languages, particularly when it appeared in lists. Like the pre-textual vernacular, the variety of 'rustic' Latin reflects increased orality or pragmatism. Scribes consciously manipulated these three different levels or 'registers' of written language, which corresponded to three different degrees of language elaboration: standard or 'normative' Latin, rustic Latin and vernacular Romance.

The return to Classical Latin driven by the Carolingian Renaissance also implied a reform in pronunciation. In the Middle Ages, as in Antiquity, texts were typically read aloud, even in private. The pronunciation of Latin became a particularly delicate matter in public readings before an illiterate audience. Theoretically, adapting the phonetic or even morphosyntactic features of a Latin text to the forms of a spoken Romance language was not a difficult task. Pronunciation reforms aiming at re-establishing Classical norms thus only served to widen the gap between written Latin and spoken Romance. Whereas 'Romanised' Latin texts were certainly understood by a large public during the 6th and 7th centuries, the same did not apply to the Latin re-established on the basis of Classical norms and read aloud with restored pronunciation.

Forms and uses of Latin as well as the intensity of writing differed from region to region throughout the *Romània*. It is generally assumed that Italian intellectuals had the best

command of both Classical and rustic Latin, owing to the large number of monastic schools there; however, in all likelihood there was no fundamental discrepancy between the Italo- and the Gallo-Romance territories, particularly in the period following the Carolingian reforms.

In contrast, the Iberian Peninsula, which was still largely under Muslim domination during the 9th and 10th centuries, appeared to be lagging behind in this respect. During this period, Iberian Latin essentially followed a ‘rustic’ tradition, and it was not until the arrival of the Cluniac monks from Gallo-Romance territory towards the end of the 11th century that Iberian written Latin began to develop. The reforms of Latin took place at a time when the difference between this language and the Romance varieties had become even more obvious than it had been two centuries earlier north of the Pyrenees. Thus ‘rustic’ Latin, which had lost its purpose, ended up merging with the new tradition of Romance writing.

4 Metalinguistic evidence

A certain amount of metalinguistic evidence for the differentiation between written Latin and spoken language (i.e. a situation of diglossia) exists from the Carolingian period onwards. Even the distinction between the two varieties of written Latin is sometimes mentioned.

The oldest – and best known – piece of evidence is found in the recommendations of the Council of Tours (813 AD), which decreed that sermons (*[h]omelia*) were to be held in the *rustica Romana lingua*. In this context, the term *rustica Romana lingua* should be interpreted as meaning “Romance language” (early French in this case; cf. Kramer, *Die Sprachbezeichnungen Latinus und Romanus*, 1998: 37sq.; 50; 52; 163):

Visum est unanimitati nostrae [...] ut [...] omelias quisque aperte transferre studeat in rusticam Romanam linguam aut Thiotiscam, quo facilius cuncti possint intellegere quae dicuntur.
(MGH, Concilia, 2, Conc. aevi kar., c. 17)

“It is our unanimous opinion [...] that [...] every person should affirmatively endeavour to translate the sermons into the Romance or German language, in order that all can more easily understand what is being said.”

This recommendation was certainly implemented, considering that sermons are some of the earliest amongst the oldest Romance documents. The oldest example conserved is the above-mentioned bilingual fragment of a *Sermon on Jonah*, dated to the 10th century (cf. no. 1).

Linguistic awareness appears to have provided the motivation for the use of the varieties in question and even seems to have played a part in their internal evolution. A reform such as the Carolingian Renaissance implies that speakers were conscious of the triglossic situation between written Classical Latin, written ‘rustic’ Latin and a spoken variety of Romance. At the same time, the implementation of the reform served to further increase the gap between Latin and Romance and to reinforce awareness

of the existence of the latter. It also paved the way for the transition of the Romance languages into writing by means of the graphemic elaboration of the fragmentary vernacular (*cf.* Carles, 2011).

10.3.6 Conclusion: the transition from Latin to the Romance languages

The transition from Latin to Romance was a transformational process that affected a language that was essentially spoken, but that was accompanied by the continued use of a more conservative written variety. This combination explains the difficulties encountered in dating and localising the internal linguistic changes. Often, the nature and time span of a linguistic change can only be estimated on the basis of scant evidence.

In order to ascertain language-internal changes that took place during this period, methods of reconstruction (phonetic, morphosyntactic and lexical) can be combined with the study of direct written evidence: the geographical distribution of lexemes in the modern Romance dialects allows the identification of the underlying evolutions (*cf.* 4.2.2); similarly, relative chronology in phonetics and borrowing between Germanic and Romance languages provide evidence of specific changes (*cf.* e.g. Pfister, *Zur Chronologie der Palatalisierungserscheinungen*, 1987).

According to the current state of research, a significant number of internal transformations appear to have taken place between the 6th and 8th centuries, perhaps around 650–750 AD, even though the changes in question actually began earlier and they had not all reached their conclusion by the end of the 8th century. However, the emancipation of the vernacular from Latin did not occur in the same manner and at the same time across the different regions of the *Romània*, with French displaying the most extreme degree of phonetic innovation and (southern) Italo-Romance proving most conservative; in the future Gascon-speaking area, distinctive phonetic features had already appeared around 500, i.e. during the period of regionalised Latin (*cf.* Chambon and Greub, *Note sur l'âge du (proto)gascon*, 2002). It is also possible that the chain of 'vertical' communication between the Latin-speaking elite and the Romance-speaking population was not broken everywhere at the same time (*cf.* Banniard, 1992).

The evolution of language across geographical space is paralleled by its evolution over time. After the disintegration of the structures of the Roman Empire, the vast Latin-speaking territory had lost its linguistic cohesion; as a consequence, the formation of smaller entities was inevitable. The exact nature and degree of interaction between different external factors such as geography, political and religious structures, as well as infrastructure, is impossible to ascertain in detail. It is nevertheless clear that their combined influence created new infrastructural networks between the 5th and 8th centuries, which prefigured the regional differentiation of the Romance varieties.

Roughly speaking, it may be considered that by the year 1000, the first major period of the development of the Romance languages had reached its conclusion – a period characterised by the transition from spoken Latin to the (spoken) Romance languages

and by the development of an awareness among the population of the ensuing diglossic situation.

The second major period began with the appearance of the first independent Romance texts and the elaboration of the Romance languages in writing, which gradually began to replace written Latin. Whereas the first millennium of Romance linguistic history was dominated by orality, the second is marked by the elaboration of writing and by the multiple interactions between the two modalities of language that resulted from this development.

10.4 The Romance languages during the Low Middle Ages (11th – 15th centuries)

10.4.1 Geographical area and migrations

The geographical area occupied by the Romance languages during the Middle Ages was almost entirely limited to Europe. The boundaries of the Romance territory were relatively stable, as were the major geolinguistic frontiers within the area. Even the three dialectal areas of Ibero-Romance, Gallo-Romance and Italo-Romance already closely resembled those of the 19th century (cf. 4.2.1).

Anthropogeography (i.e. the study of man in relation to his geographical environment) is of importance to medieval Romance geolinguistics with regard to three aspects in particular: (1) the major migrations and colonisations of the Christian *Reconquista*, the settlement of the Daco-Romance territory (the territory of present-day Romania) and the Norman conquest of England, (2) a further series of more sporadic migrations and colonisations, and (3) the interplay between population stability and fluctuation within the established linguistic territories.

1 Major migrations and colonisations

The *Reconquista* of Ibero-Romania essentially lasted from the 11th to the mid-13th century, only reaching its conclusion in 1492. It was a reaction to the Muslim conquests, which had brought approximately three quarters of the Iberian Peninsula under Arab domination (cf. 10.3.3). Through a movement of expansion in two phases, Catholic forces gradually reclaimed the Peninsula, first reaching the Tagus river (around 1100 AD), then the Guadalquivir and the Atlantic coast (around 1250). Only the Emirate of Granada (i.e. the eastern part of present-day Andalusia) remained under Muslim rule until its fall in 1492 (cf. 3.4.9, fig. 7).

Following this expansion, the Romance varieties that had spanned the northernmost part of the Peninsula from west to east were exported towards the south. Consequently, the language spoken in Galicia in the northwest gave rise to the language that would later be spoken on the territory of modern Portugal, and western Catalan

became the model for the language spoken in Valencia. The southward expansion of Leonese and Aragonese, in contrast, was quickly halted by Castilian, which would ultimately dominate the entire central band of the Peninsula from north to south.

In reality, all Ibero-Romance expansion varieties played the role of colonising languages, just as Latin had in its time. Owing to the considerable degree of population mixing on the Iberian Peninsula, diatopic variation is noticeably less developed there than in regions with more stable settlement. Contact between the languages of migrants of different origins resulted in considerable linguistic amalgamation.

Thus, dialect variation in the north of the Peninsula is similar to that of Italo-Romance and Gallo-Romance, whereas regional variation in the central and southern regions is considerably limited, concerning above all phonetics, and, to a much lesser extent, vocabulary. Similarly, south of Burgos, we encounter almost exclusively ‘secondary dialects’ (*cf.* 3.4.9 and 4.2.2) or ‘varieties of expansion’. These already possess some characteristics of elaborated languages, as a result of phonological, morphological and lexical neutralisation.

For this reason, the emergence of standard Portuguese and standard Spanish at the dawn of the modern period took place under different circumstances than those which marked the development of standard Italian or French, where differences between dialects and the standard language were much more pronounced.

The same geolinguistic phenomenon may be observed in the area occupied by Romanian. It can be supposed that the pastoral Romanian-speaking population, which had become established south of the Danube, began to extend the radius of its nomadic migrations from the 10th century (*cf.* 10.3.2). On the relatively sparsely populated Balkan Peninsula, these migrations reached the Greek mountains in the south, Istria in the west and present-day Romania in the north, which at the time was under Slavonic rule. It is probably by this means that the Romanian language spread over today’s Romanian-speaking territory in a process that spanned the whole of the medieval period. This expansion must have begun in mountainous zones, where it most likely involved the assimilation of other Romanian-speaking groups that had remained north of the Danube after the fall of the Roman Empire, before reaching the fertile plains that were populated – albeit still sparsely – by Slavonic peoples.

For the same reasons pertaining to migration and reciprocal assimilation, the diatopic variation of Romanian is as little developed as that of the secondary dialects of Ibero-Romance.

The Norman conquest of England, which began in 1066 with the decisive victory of William the Conqueror, Duke of Normandy, at the Battle of Hastings, was the third important migration of the Romance Middle Ages. As a result of this conquest, French became established in England as a written language of prestige alongside Latin, a status it was to hold until the end of the medieval period. The English kings of French descent – the house of Plantagenet (including Richard I “the Lionheart” and his mother

Eleanor of Aquitaine) – also reigned over substantial parts of the Continent, Gascony in particular. Conflict over their continental power resulted in the Hundred Years' War (1337–1453), which also saw the end of the use of French by the English ruling classes. The use of this language would henceforth be restricted to specific contexts until the 16th century, when it definitively lost its status as a prestigious language.

This period had three major consequences for language history: it led to (1) the production of a significant number of texts written in Anglo-Norman during the Middle Ages (the vocabulary of which is largely documented by the AND, *cf.* 9.9.1); (2) a considerable number of borrowings into English vocabulary (*cf.* 3.6.2), where words of Latin or French origin may account for as much as 50% of terms in specific domains of the lexis of present-day English, particularly in cultured language; and (3) the systematic elaboration of French vernacular textual genres between the end of the 11th and the third quarter of the 12th centuries, catalysed by the imitation of existing Anglo-Saxon written traditions (*cf.* 10.4.3 no. 1). By this means, French became the first Romance language in which a wide variety of textual genres was elaborated, a factor which resulted in the predominance of French as a model throughout the medieval *România*.

2 Episodic migration and colonisation

The medieval period witnessed numerous other conquests and migrations that led to the displacement of Romance-speaking groups as well as to situations of language contact, both extra- and intra-Romance:

- The Crusades and the establishment of the Latin Kingdom of Jerusalem (1099–1187) made French a *lingua franca* spoken and written in Asia Minor. This episode nevertheless remained without lasting consequences.
- Of more limited impact, the Norman conquest of southern Italy in the 11th and 12th centuries exported French to the southern part of the Italian Peninsula and Sicily, due in great part to King Roger II of Sicily, grandfather of Frederick II of Hohenstaufen, who would acquire the nickname of *stupor mundi* “wonder of the world” (*cf.* the map of the distribution of Romance designations for “mare” [3.5.3, fig. 10], which shows the introduction of the French word *jument* in Sicily as a result of the spread of courtly culture).
- During the 11th and 12th centuries, settlers and pilgrims from the Occitan- and French-speaking domains journeyed across northern Spain via the road of Santiago de Compostela (among others), playing a significant role in the development of written culture and the elaboration of the Spanish language. Communities of so-called Franks introduced a tradition of Occitan writing, consisting chiefly of legal texts.
- From the 12th to the 15th centuries, Venetian rule over the coastal region of Dalmatia and in the east of the Mediterranean basin considerably enlarged the territory of

the Venetian dialect. The colonisation of Dalmatia had an influence on the Dalmatian language, which closely resembled dialects of north-eastern Italy.

- The activity of Italian merchants in the Mediterranean basin and northern Africa, sometimes accompanied by political and military manoeuvres, led to the sporadic emergence in the 16th to 18th centuries of pidginised varieties of Italian. These minimal languages of communication, although few in number, are collectively known as the Mediterranean *Lingua Franca* (cf. Cifoletti, *Storia della lingua franca*, RSG 1, art. 97; Baglioni, *L'italiano delle cancellerie tunisine*, 2010). *Lingua franca* later became a generic term for languages of communication used among speakers of different mother tongues.
- In the 15th to 16th centuries, Aragonese rule over Sicily and southern Italy intensified the language exchange between the Iberian Peninsula and Italy.
- From the 14th century onwards, sovereigns and merchants from Portugal and Spain began undertaking voyages to the Canary Islands, as well as conquests that would later impose the Portuguese language in Asia and both Spanish and Portuguese in the Americas.

Finally, to some extent, written languages reflected cultural contacts that did not rely on migrations of large groups but rather on the individual movement of literate persons, which resulted in the transposition of elements of their own particular culture to their new surroundings.

It is in this context that the use of French as a written language in northern Italy in the 13th and 14th centuries – so-called ‘Franco-Italian’ – is to be understood. Such texts were written in a defined variety of regional French (often Picard), but with a large number of lexical or phonetic Italianisms. This was an artificial language, restricted to specific textual genres and whose formal appearance differed greatly from case to case (cf. Holtus, *La Versione franco-italiana della “Bataille d’Aliscans”*, 1985; Barbato, *Il francoitaliano: storia e teoria*, 2015; Zinelli, *Espaces franco-italiens*, 2016).

In the same manner, collections of troubadour poetry (so-called ‘*chansonniers*’), were produced and copied in northern Italy, Catalonia and northern France. The influence of these texts on language, however, remained weak (cf. nevertheless the complex multilingual situation described by M. Pfister in his *Untersuchungen zu Girart de Roussillon*, 1970).

All these migrations and contacts had a more or less significant and lasting effect on language, without, however, changing the major language frontiers established throughout Europe at the beginning of the Middle Ages.

3 Stability and fluctuation within established geographical zones

Anthropogeography was also subject to variation within the geographical zones of medieval Romance that had already been established. General mobility began to increase in the 12th century, although it mainly concerned specific groups of the pop-

ulation (merchants, religious and erudite circles and artisans); wars and crusades led to larger migrations and depopulation (consider the Albigensian Crusade in the 13th century, which opposed the French-speaking and the Occitan-speaking populations [immortalised in the *Canço de la cruzada albigès*], or the brutal Hundred Years' War between France and England, which lasted from 1337 to 1453).

The destructive power of wars was equalled, if not surpassed, by that of epidemics (such as the plague in particular, which decimated 30% of the European population in 1348 and the ensuing years). Climate change added to these problems, causing not only food shortages but also population displacements (a cold spell around 1100, for instance, led to populations deserting Alpine zones above 1000 meters; cf. Le Roy Ladurie, *Histoire du climat*, 1967).

Despite destabilising factors such as these, the inhabited areas throughout the *Romània continua* (Gallo-Romance and Italo-Romance and the northern fringe of the Ibero-Romance territory) maintained a linguistic physiognomy characterised by consistent patterns of variation which were determined by anthropogeographical factors (cf. 4.2.2). This geolinguistic order relied on four partially conflicting factors: (1) language stability in a given place, (2) uninterrupted dialect evolution, (3) exchange between neighbouring varieties and (4) the impact of writing.

1. Research in historical linguistics has shown that dialect varieties of the *Romània continua* underwent minimal evolution within contained geographical spaces. Within one and the same territory, the phonological and morphological systems, as well as the vocabulary of modern dialects, can be directly linked to medieval sources. Similarities are striking, despite the elaborated nature of the language in written medieval sources. On the basis of lexemes existing in modern dialects that are restricted to small areas, it is even possible to geographically localise words found in manuscripts and thus to determine the origins of scribes or medieval authors (cf. e.g. the study by Chambon, *Remarques sur la patrie de l'auteur du ms. Brit. Mus., Add. 17920*, 1995).
2. Such continuity naturally did not preclude language development. While numerous similarities in the phonological system or lexis of a variety spoken in a village in Provence or in the Italian region of Umbria may be observed between its 13th- and 19th-century stages, phonological differences, word forms and word meanings nevertheless underwent change. The local dialects of the *Romània* have never ceased to evolve and the divergence between them has therefore increased since the medieval period.

This is particularly noticeable in southern Occitan – especially Languedocien – and Catalan, in that the varieties of these languages resembled one another to a considerably greater degree at the beginning of the 13th century than at the beginning of the 20th century – both in writing and, presumably, in speech. With regard to the medieval period, we can speak of a weakly differentiated South-Occitan-Catalan zone. This stands out all the more in the case of Galician, which was

indistinct from Portuguese until the mid-14th century when Portuguese expanded from Galicia toward the south of the Iberian Peninsula. The differences between the two languages, which were still not particularly pronounced in the 14th century when Galicia came under Spanish rule, would continue to increase.

3. Language differentiation was nevertheless counterbalanced by situations of contact and exchange between population groups and through the force of attraction between neighbouring varieties. Two neighbouring Romance varieties (or even two neighbouring Romance languages) resemble one another to a greater extent than two dialects (or languages) separated by a greater distance. Thus, Catalan is closer to Aragonese than to the dialects of Castile, the Saintongeais dialect (French) is closer to the neighbouring Limousin dialect (Occitan) than to that of Lorraine (French), both in the 13th and in the 19th centuries. The aspect that may have changed over time is the degree of differentiation.
4. Finally, the impact of written Romance languages also played a role in reducing dialect diversification.

4 Conclusion

The Romance zone was stable during the first half of the second millennium. When observed in greater detail, however, this apparent stability was perturbed by more or less salient phenomena of divergence. Nevertheless, cohesion between linguistic zones has a stronger determining influence than sporadic interruptions. The changes that dialects in particular would undergo over the course of almost a millennium would only minimally alter their geographical distribution established around the year 1000.

10.4.2 Infrastructure, socio-cultural development and hierarchical institutions

While Antiquity heralded the cultural blossoming of the Mediterranean countries, the Late Medieval period was a pivotal era for the development of western Europe. The cultural gap between the more advanced southern part of Europe and its northern regions would not be levelled out until after the end of the Middle Ages. The medieval *Romània*, which was in full effervescence, left behind an exceptional cultural heritage, both in terms of religious and secular architecture and its rich and varied textual production.

Major developments took place in the 12th century: agriculture intensified and new land was cleared; urban centres and long-distance trade routes developed. Nevertheless, despite the soaring rise of literature, these texts, written by the elite, which constituted only 2 to 5% of the population, contain little indication as to their origin. They easily lull one into forgetting the hardships of life, the famines, epidemics and wars suffered by the human population during that period of history. Life expectancy was a mere 35 years – a figure which would remain unchanged until the beginning of the 19th century – and the mortality rate of children was extremely high; hierarchical struc-

tures, social as well as those characterising the relationships between men and women or between parents and children, were insurmountable. Human existence on earth was seen as a transient period of suffering on the road to true eternal life in the kingdom of God.

The individual had no intrinsic value, although the first signatures adorning artistic and literary works as well as documents of a legal character marked the tentative beginnings of an awareness of individuality from the 12th and 13th centuries onwards. Even though troubadour poetry and romances such as *Cligès* (ca. 1176) paved the way for the development of the cultural and individual concept of love, it was still, as Huizinga aptly put it with regard to the 15th century, ‘an evil world. The fires of hatred and violence burn fiercely. Evil is powerful, the devil covers a darkened world with his black wings.’³⁷

The history of medieval languages relies on three cultural milieus: urban centres, princely courts and the Church. The Church was the dominant power: omnipresent, with strictly defined hierarchies and a near monopoly on writing, it offered an invaluable guarantee of stability over time and space:

- In numerous villages, 5% or more of the population belonged to the secular and regular clergies. These were completed by secular organisations (the ‘third order’) and brotherhoods organised around the Church.
- Church possessions, which were subject to the law of ‘mortmain’ (“inalienability”), represented extraordinary wealth. This was dispersed when the property of the Church was secularised, having been expropriated by the state either during the French Revolution or later on (in 1835 in Spain, between 1848–1870 in Italy and in 1911 in Portugal). Today only major cathedrals still bear witness to this extreme wealth.
- The production of texts, not to mention literature, could never have achieved the importance it did without the dense networks of ecclesiastical *scriptoria*. Universities emerged within the framework of cathedral schools, just as hospitals did from monastic guesthouses. The Church thus shaped the foundations of cultural and intellectual life in urban centres, as well as those of education and a network of charitable institutions – on the surface, two factors of a very different nature, both nevertheless indispensable for the development of complex societies.
- Parishes covered the whole of the Romance-speaking territory, reaching even the most remote parts of the countryside. Through its omnipresence and its religious and social prestige, the parochial network was able to contribute to the establishment of cross-regional linguistic norms. It has been suggested that the boundaries of dioceses were a determining factor in the shaping of dialect zones (cf. Morf,

³⁷ Huizinga, *The Autumn of the Middle Ages* (1996: 29 [or. 1919]); cf. also Le Roy Ladurie, *Montaillou, village occitan* (1975) for a vivid image of the Romance Middle Ages.

Zur sprachlichen Gliederung Frankreichs, 1911; Jud, *Sur l'histoire de la terminologie ecclésiastique de la France et de l'Italie*, 1934).

- Mendicant orders, in particular Franciscans and Dominicans, who travelled across Europe from the 13th century onwards spreading a religious ideology that was close to the people, also contributed to a preparatory phase of language standardisation.

The importance of urban centres and princely courts varied greatly over time and in different regions of the *România*. Medieval Italy had strong independent cities, some of which included dependent feudal domains. In such cities there were a growing number of persons able to write, such as notaries, but also wealthy artisans and merchants and, more generally speaking, members of patrician circles. The first universities appeared in the 13th century, e.g. in Salerno or Bologna (as well as in Paris and Montpellier), which reduced the ecclesiastical monopoly on writing. Significant exceptions to this network of urban Italian centres were the courts of the Papal States, which covered a large part of central Italy, as well as the Angevin kingdom, and the Aragonese kingdom in Naples and southern Italy between the 13th and 15th centuries.

In contrast, princely courts were dominant in Spain, which already displayed the first characteristics of a centralised state. The *Reconquista* had endowed the country with vast territories. Later, during the reign of Alfonso X 'the Wise' of Castile in the second half of the 13th century, a central administration was introduced. Intellectual circles developed mainly around the royal court. In addition, the Iberian Peninsula had the asset of a triple culture (Christian, Muslim and Jewish), which was responsible for various social and cultural innovations in fields as diverse as agriculture, medicine, urban administration and intellectual creativity (cf. Castro, *De la edad conflictiva*, 1972). The kingdom of Portugal, which began to take on its modern contours in the 13th century, remained in the shadow of Castile.

The Gallo-Romance territory evolved along a path somewhere between these two extremes: southern France can be compared to Italy in this respect, while northern France more closely resembled Castile, with the difference that in France there was an influential nobility with powerful ducal courts (such as the Duchy of Burgundy in the 14th and 15th centuries) as well as a number of smaller princely courts. The royal court, moreover, was still itinerant at the time, as were the princely dynasties, which were not yet attached to fixed territories. In contrast to Castile but in line with Italy, then, the geographical entities of interest to linguists are relatively small, or even limited to the boundaries of an individual city.

In addition to its urban centres, courts and the Church, the medieval *România* displayed a finely-meshed network of linguistic cohesion, which contributed to the homogenisation of spoken language at the regional level. Not only were peasants in contact with representatives of the Church and minor lords, but also with itinerant merchants and professionals such as cutlers or storytellers. Situations of language contact involving these groups, which was typically short-term, introduced new linguistic models to the country-dwellers, eager for any sort of event that would break the monotony of

everyday life. Finally, the relations between urban centres and the surrounding countryside enabled commercial exchange and contact with journeymen, as well as opportunities for communication afforded by specific events such as festivals or marriages. Up until the 18th century, however, the latter remained limited to a range of only a few kilometres, in accordance with the extension of social networks in everyday life.

10.4.3 Written culture and linguistic thought

1 Chronology of written production in Romance

During the Late Middle Ages, writing occupied a position of ever-increasing significance in Western Europe; this included both Latin writing and writing in the vernacular. Highly elaborated forms of the latter already existed. Whereas the prestige of Latin was still considered superior by the elite capable of reading it, Romance texts had the advantage of being comprehensible to the common people.

Before 1100, the number of written sources for Romance was still very limited: the *Inventaire Systématique* (which remains incomplete) lists a total of 50 texts (including those previously mentioned in 10.3.5 no. 1), comprising thirteen in Italo-Romance, eleven in French, ten in Catalan, seven in Occitan, four in Spanish, three in Romansh and two in Sardinian. One can assume this to be roughly a mere 10% of what was actually produced between 850 and 1100 AD (*cf.* 11.3.3).

In the 12th century, written Romansh disappeared, reappearing only three centuries later. Spanish remained under-represented with only five written source-texts (but our knowledge of these sources is insufficient, a state of affairs which also applies to Galician-Portuguese, entirely absent from the above list of documents). In the other territories of the Western *Romània*, however, a Romance writing culture did develop, to which the existence of 29 sources in Italo-Romance, 27 in Sardinian, 22 in Catalan, 97 in French and 182 in Occitan bears witness (*cf.* Petrucci 1994 and 10.3.5 no. 1 above).

The rise of the written vernacular was particularly apparent in the Gallo-Romance domain: in the 12th century, Occitan was mainly used to record acts concerning sales, enfeoffments and loans in the form of the pledging of property (*cf.* the edition of the earliest Occitan charters by Brunel, 1926–1952). It is also important to consider the appearance of troubadour poetry from ca. 1100 onwards. These poems originated as a sophisticated textual genre in written form, even if the transmission of the texts as they are known to us did not begin before the 13th century, when they were collected in codices (known as ‘*chansonniers*’).

The first charter left to us in Old French was written in Douai in the extreme north of France and dates from 1204–1205. Documentary writing in Old French thus began a century later than in Old Occitan. From the 12th century onwards, however, a number of important literary texts were written in the *langue d’oïl*. Until ca. 1160–80, these were entirely produced in England – it should be noted that the elaboration of written French

texts began in this multilingual environment and was only exported to the Continent a century later (cf. Glessgen, *La genèse d'une norme*, 2017). The earliest texts were epic poems ('*chansons de geste*', such as the *Song of Roland*, *Voyage de Charlemagne*, *Chanson de Guillaume*, *Fierabras* or *Aliscans*). These were followed by the *lais* of Marie de France (written in England) and the first 'romance' literature, the latter consisting of the *roman antique* (e.g. the *Roman d'Alexandre*, *de Thebes*, *d'Énéas* and *de Troie*) and the Arthurian romances. Arthurian literature chiefly found its expression in the works of Chrétien de Troyes, the first author of significance active on the Continent, albeit only after a period spent at the royal court in England. His five works *Erec*, *Cligès*, *Lancelot*, *Yvain* and *Perceval* had a strong influence on the evolution of written French and contributed towards a new conception of literature. The earlier branches of the *Roman de Renart* are a further example of 'romance' literature.

From the 13th century, Romance sources multiplied, with the exception of Romanian and Romansh, which were not consistently put into writing until modern times. Textual genres diversified first in Gallo-Romance, then in Ibero-Romance and, towards the end of the century, in Italo-Romance. Literary and documentary as well as religious, scientific and technical texts written in Romance flourished and gradually overtook writings in Latin.

The wealth of medieval sources reached its culmination in the 14th and 15th centuries, which resulted in the transmission of hundreds of millions of original manuscript pages written in Romance (cf. 11.1.4). The contrast between such an abundance of writing and the sixty odd sources which exist for the period between 850 and 1100 is a testimony to the development of a true written culture during the Late Middle Ages. Though of fundamental importance for the history of language and culture in the Romance-speaking world, the rise of Romance writing took place against the backdrop of Latin, which initially maintained its position of supremacy.

2 Latin and the Romance languages through the centuries

Medieval writing in the *Romània* was dominated by Latin texts, which gave rise to countless translations or '*volgarizzamenti*' and served as models for the development of Romance textual genres, as well as for stereotypical formulas or pre-fabricated sections of texts. In reality, the formation and development of the written Romance languages displayed little autonomy, in contrast to Ancient Greek or Latin. Unlike these, the Romance languages can be considered to have undergone a process of acculturation based on and dependent on Latin (cf. 10.2.2), enriched, moreover, by the influence of Arabic on the Iberian Peninsula. The history of the written Romance languages is thus at the same time that of the substitution of Latin by new prestigious varieties.

Between the 9th and the 11th centuries, Latin still fulfilled all the functions of a language used for long-distance communication, though the Romance languages were already well established as spoken languages for the purpose of communicative immediacy. The social elite, who possessed a command of Latin as a written language, were

increasingly distancing themselves from the rest of the population, who were increasingly less capable of understanding it. By the end of this evolutionary process (i.e. the 18th to 19th centuries), the Romance languages – and other modern national languages, such as English in particular – had usurped the place of the old prestigious varieties, relegating Latin to a marginal position even in the domains of history, literature and theology. Between the Middle Ages and the modern period, Latin was gradually abandoned in favour of the Romance languages for the purpose of long-distance communication. This process, which led to the replacement of a single language by a series of new languages over the course of a millennium, is illustrated in simplified form in the following figure:

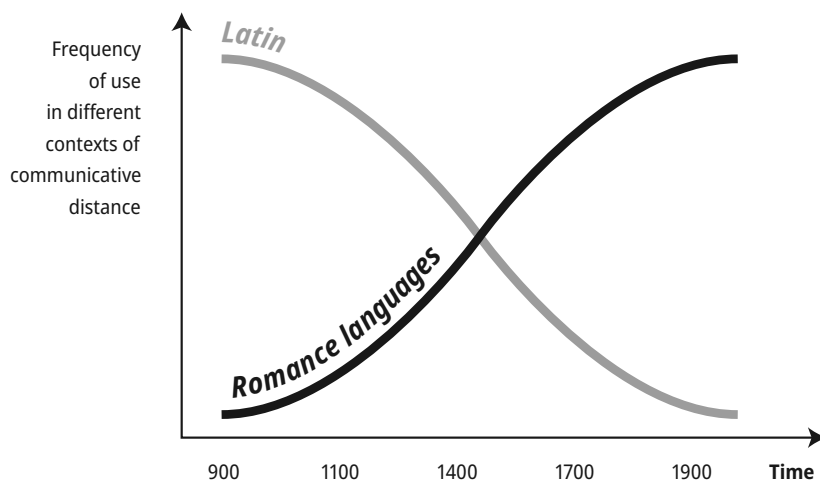


Fig. 31: The replacement of Latin by the Romance languages

The diagram is inaccurate in so far as it omits the fact that Latin retained its function as a linguistic model well into the 20th century; nor does it account for the many and varied manifestations of mutual influence between Latin and the vernacular. Nonetheless, it clearly shows from start to finish the almost mathematical nature of the process by which one language was replaced by another. Concrete manifestations of this process are visible when language change is studied within different textual genres, which represent or reflect the different forms of communication (*cf.* ch. 11.1 for more detail).

Consider the example of judicial and administrative documents: until the 11th century, Latin was practically the only language used throughout the *România*. Romance elements appeared sporadically within Latin documents in the form of lexemes and proper nouns, as well as morphological traits, especially in those sections of the texts which were strongly marked by oral features. Latin nevertheless clearly remained the dominant language.

As we have seen, charters were written in Occitan from the 12th century. The use of French in such texts did not become established until the 13th century, to a greater or lesser extent depending on regions and chanceries: while the earliest documents date from 1204 to 1205, the royal chancery began using French around 1250, and by the end of the century, the majority of chanceries made use of both French and Latin. This co-existence of these languages continued until 1530, when French became the exclusive language of the royal chancery.

In Spain, following the Cluniac reforms, the *scripta latina rustica* first gave way to a form of re-established Latin, which would, in turn, find itself in competition with Spanish from the end of the 12th century onwards. Through reforms of the chancery by Alfonso the Wise, Spanish was firmly established as the language of law and administration around 1260 to 1280. In Italy, the *volgari* began to be introduced above all in the 14th century, due to a greater degree of familiarity with Latin and doubtless also to greater internal linguistic proximity between the two varieties. As in the majority of Romance-speaking countries, Latin was abandoned in judicial-administrative practice in the 16th century, after a long period during which Latin and vernacular texts had existed in parallel, with a great deal of linguistic interference in both directions.

In literature, the first important Romance creations appeared in the 12th and 13th centuries, though many were inspired by existing Latin models. Nearly all Romance texts earlier than 1200 are translations or adaptations of Latin texts. For several centuries, the latter would remain an essential counterpart to Romance literature. Authors such as Dante, Petrarch or Boccaccio, the ‘founders’ of the literary Italian language, also left important treatises and other writings in Latin, as did many other writers of the 16th and 17th centuries. Romance writing in the literary domain, which, in its initial phase, was mixed with Latin, nevertheless gained independence surprisingly quickly and soon brought forth highly refined texts: the poems of the first troubadour, William IXth of Poitiers, the romances of Chrétien de Troyes or the *Scuola siciliana* may be considered to be on a par with the finest literary works of the modern or present-day periods.

The chronology and characteristics of the evolution of scientific and technical texts were different again. Latin was to remain the dominant language for theoretical and scientific discussion until the 18th century (cf. Ernst, *Sprachkontakte: Latein und Italo-romania*, RSG 2, art. 134). It is essentially for this reason that scientific terminologies that have become established in the modern world rely on Greco-Latin vocabulary.

The role of Latin in the development of textual genres was directly correlated with its role in the teaching of written language. Up until the modern period, writing was taught on the basis of the Latin model. Consequently, every person who wrote in a Romance language during the Middle Ages and most literate members of society up until the French Revolution were also competent in Latin (cf. Gerner, *Éducation et histoire des langues: la Gallo-Romania*, RSG 2, art. 107). The multiple interactions between Latin and the Romance languages were thus caused by these bilingual key players, who were constantly exposed to language interference (cf. 6.6). Aspects of interference materialised

both on a small scale, through frequent Latinisation at the graphemic level and lexical or syntactic borrowings, and in more targeted moves towards Latinisation, which went so far as to result in major changes to written French in the second half of the 12th century and to Italian in the 15th and 16th centuries.

Reasons for the progressive abandonment of Latin between the 12th and 18th centuries are not to be sought in a failure to master this language on the part of professional writers (who were normally bilingual), but rather in its incomprehensibility for the majority of the population. The slow rate at which this language was abandoned in writing, despite the fact that it had disappeared from spoken language well before the year 1000, nevertheless suggests that the need for comprehensibility is a modern rather than a medieval concept: at the time, writing symbolised political and economic power, and the distinction of ruling classes was often established on the basis of incomprehension or exclusion. Sacred languages, such as Latin or Sanskrit in the Middle Ages or Classical Arabic in the modern period, generally tended to become incomprehensible to all but the elite.

The appearance of the Romance languages, in contrast, was correlated with situations in which mutual comprehensibility was particularly important. They were first introduced in writing directed at an illiterate public – the texts in question were often religious or of a legal character, but also included poetry and literary texts in general. Later, they contributed to the construction of the socio-cultural and political identity of the elite: the elaboration of scientific prose at the court of Alfonso the Wise, as well as his chancery reforms, served his imperial ambitions, just as the Carolingian reforms served those of Charlemagne and Louis the Pious, (politically more successful). While the reform of the 9th century targeted Latin, that of the 13th targeted the future Spanish language.

Depending on the historical situation, the use of Latin rather than a Romance language, or vice versa, thus held a particular significance, which should be interpreted according to the poles of communicative distance and immediacy on the one hand, and comprehensibility and clarity of content on the other.

3 The *scriptae*

The medieval Romance languages in their written form bear witness to highly-developed regional variation, which constitutes a further dimension of observation, complementary to those of chronology and the distinction between textual genres. Graphemic and grapho-phonetic traditions, morphology and vocabulary all displayed regional features (for the latter, less widely studied, cf. Glessgen / Trotter, *La régionalité lexicale au Moyen Âge*, 2016); in the cases of morphosyntax and syntax, on the other hand, the degree of diatopic variation was very low within a given language, although they did display clear differences according to textual genre.

Within the Hispanic domain, Leonese, Castilian and Aragonese could be distinguished; Eastern Catalan (= Barcelona and the Balearic Islands) could be distinguished from Western Catalan (including Valencian); in Italy, varieties such as Venetan, Lombard, Piedmontese, Tuscan, Roman, Neapolitan or Sicilian took shape; in the *oïl* territory, Walloon, Picard, Anglo-Norman, Lorrain or Burgundian, to give a few examples, displayed their own characteristics; within the Occitan-speaking area, Arverno-Limousin and Dauphinois in the north could be distinguished from Languedocien and Provençal in the south. Each of these regional varieties covered a defined area, as did Gascon, Francoprovençal or the Gallego-Portuguese language, and they were marked by a certain degree of homogenisation which contrasted with the pronounced differentiation displayed by the oral dialects.

In order to better differentiate the medieval written language from the spoken dialects, Louis Remacle and Carl Theodor Gossen introduced the concept of '*scripta*'³⁸. The *scriptae* as regional written languages were a semi-artificial realisation of the spoken Romance languages, elaborated under the influence of textual and lexical models provided by written Latin. The *scriptae* thus took on the appearance of a cultured variety that could be understood over a large geographical area, rather than reproducing the marked features of local spoken varieties.

The tendency towards Latinisation and homogenisation inherent in the *scriptae* necessarily led to the development of similarities among different written regional varieties across larger linguistic areas, increasing the distance between these written varieties and the dialects spoken by their authors. In some cases, this led to strong contrast between writing and pronunciation.

While written regional varieties foreshadowed the standardisation of Romance languages in the 16th and 17th centuries, they did not as yet have a dominant standard form in the Middle Ages: in their beginnings, writings produced in Tuscany, for example, or by the French or Spanish royal chanceries, merely represented regional varieties among others; scribes from 14th century Lorraine tended to rely on models from Burgundy or Picardy as well as from the Île-de-France and the royal court. Moreover, the lack of defined orthographic or grammatical norms favoured a high degree of variation as regards graphic representation and morphology in these texts.

The enormous capacity for variance displayed by writing was further increased by the circumstances of textual transmission – particularly in the case of literary and scientific texts. Manuscripts were copied by scribes who lived during different historical periods and in different places, and who consequently reproduced differing language models. The form of each text, if not its content, was thus adapted according to the usual practices and the intentions of the copyists and their patrons (cf. 11.2.5).

38 For the history of scriptology, cf. Völker, *Skripta und Variation* (2003), as well as Glessgen, *Trajectoires et perspectives en scriptologie galloromane* (2012).

Written medieval varieties, then, were composite languages with a high degree of internal variation, strongly permeated by Latin and moulded within textual traditions. They displayed a clear tendency toward neutralisation, even in the absence of an established unifying model, and at the same time contained obvious regional or even local elements. The Romance *scriptae*, far from being ‘vulgar’ languages, were actually sophisticated semi-artificial varieties of high sociolinguistic prestige.

4 Linguistic culture

During the medieval period a certain degree of interest in the features of language and style began to develop, as well as in reading and writing. Glossaries or rudimentary dictionaries were written, some containing large numbers of words, as well as several treatises on norms, language use and problems related to translation, and a few manuals for language learning.

In the absence of epistemology, which is a characteristic of the 19th- and 20th-century sciences, our medieval predecessors thought in socio-cultural and intellectual terms different from those that would later emerge in modern linguistics. Medieval and pre-modern linguistic culture harbours many surprises, which mostly consist in profound divergences with regard to modern-day sciences, but which sometimes take the form of striking similarities.

Medieval grammatical description of the Romance languages remained extremely limited. The most important textual tradition is that of the French and Occitan *Donats* dating from the 13th and 14th centuries. These texts are adaptations of Latin grammar treatises, attributed to a certain Donatus or Pseudo-Donatus (4th century, cf. RSG 1, art. 21sq.: *Histoire de la réflexion sur les langues romanes: français* [Bierbach and Pellat]; *occitan* [Perugi]). Romance grammatical thought during the medieval period was thus based on grammatical theory developed in Antiquity; this is not surprising since education at the time relied exclusively on Latin (cf. no. 2 above). Moreover, the content of grammatical works was primarily theoretical in nature, oriented towards interpretation rather than the description of language with regard to its use and variation. Dante, who, in his remarkable *De vulgari eloquentia* (ca. 1303), considers variation to be an inherent component of language, constitutes an exception. This revolutionary perspective does not, however, seem to have had any noticeable impact on linguistic thought until it was rediscovered by the Italian Renaissance humanist Trissino in the 16th century (cf. 10.5.3 no. 3; cf. Bossong, *Sprachwissenschaft und Sprachphilosophie in der Romania*, 1990: 57).

Lexicography or, more specifically, glossography, in contrast, underwent significant development in the Middle Ages. It first appeared in the form of glosses (i.e. marginal or interlinear comments) that sporadically accompanied Latin texts: the *Glosas Silenses* and *Aemilianenses* from Northern Spain, the *Kassel Glosses* and the highly significant *Reichenau Glosses* (dating from the end of the 8th century and compiled in Corbie in Northern France) are some of the earliest evidence for the Romance languages (cf. 10.3.5

no. 1); numerous other glosses followed throughout the entire medieval period. These interlinear or marginal comments provided explanations for the Latin used in these texts in order to facilitate its learning. The acquisition of the Latin language thus had a direct impact on the first appearance of Romance vocabulary in writing as well as on its consolidation and expansion.

Later, independent glossaries comprising lists of Latin and Romance words appeared. There is also evidence of Hebraic-Romance, Arabic-Romance or even Romance-Romance glossaries, such as Pietro Berzoli de Gubbio's Provençal-Italian glossary (dating from the beginning of the 14th century; cf. Pfister, *Die italienische Lexikographie von den Anfängen bis 1900*, HSK 5.2, art. 187 and the MRL *Manual of Judaeo-Romance linguistics and philology* by Mensching and Savelsberg). These lists followed a semi-alphabetic or encyclopedic order, the latter based on the authors' desire to establish a common vocabulary for particular concepts.

Some of these texts were of considerable length: for example, the Latin-French *Aalma* glossary (ca. 1380, judging by the first entry) or the *Catholicon* of Lille (15th century, approximately 130 folios in length [i.e. 260 pages]). There are also specialised glossaries, such as those containing medical or pharmacological terms. Others, by far less numerous, are rather short and have the primary function of teaching Latin or another foreign language (cf. Fryba, *Étude et description étymologique et historique du lexique des langues romanes: le français et l'occitan*, RSG 1, art. 32).

Medieval glossography contributed to language elaboration, as did its modern successors, beginning with humanist lexicography (cf. 10.5.3 no. 3). In establishing a vocabulary that could serve as a point of reference, it contributed to language standardisation.

Texts focusing on the learning of 'modern' languages constituted a less widespread genre. German-Venetian works were written in the 15th century as an aid for German merchants who came to trade in Venice. From the point of view of their form, these books closely resembled – and may have been inspired by – encyclopedic glossaries. Among their contents, one can find ready-made sentences such as *Parla forte, che io t'intendo ! – Redd laut, daz ich dich verste!* "Speak loudly so that I can understand you!" or *Dio me ne dia la grazia! – Got geb mir sein Genad!* "May God grant me His grace!" (cf. Rossebastiano Bart, *Vocabolari veneto-tedeschi del secolo XV*, 1983).

Finally, more theoretical works on the nature and specificities of language are also encountered at intervals throughout the medieval period. The most well known of these is the treatise cited above, *De vulgari eloquentia*, composed by Dante Alighieri, which contains highly accurate reflections on regional differences between varieties of language, as well as on linguistic prestige and norms. Just as noteworthy is the line of thought emanating from the translation school established at the court of Alfonso the Wise, which addressed specific issues relating to problems inherent in translation. Scholars of this period mainly reflected on the contribution of translation to the elaboration of Spanish, which at that time was in the phase of pre-standardisation (cf. Bossong, *Sprachtheorie und Sprachphilosophie in der Romania*, 1990).

Moreover, it is not unusual to find mention of, or judgement on, linguistic phenomena in medieval literary or documentary texts: reading (aloud), speaking, perceiving, listening, understanding or remaining silent were aspects of communication that piqued the curiosity of those interested in language³⁹.

The influence of medieval linguistic thought on written language, textual genres and spoken communication, however, remained relatively weak – only glossography may have had a direct, tangible impact on Romance writing. Nevertheless, the different currents of thought inspired by language and writing bear witness to an interest in such questions, at least among the literate elite of the population. These are the first signs of a tradition that would develop during the modern period, ultimately resulting in the emergence of modern linguistics.

10.4.4 Minority languages and varieties

The medieval world was not monolingual. Language contact between socio-cultural groups as well as between individuals became more varied in conjunction with increasing mobility in Europe, beginning in the 11th and 12th centuries. From this point on, different dialectal and regional varieties began to develop within individual languages.

Dialects were local varieties, associated with a particular city or even a village and identifiable as such by all inhabitants of the locality in question. Nonetheless, internal differences among local dialects – the only true mother tongues in the *Romània* – were not as pronounced as they were in the modern period (cf. 10.4.1 no. 3).

Regional varieties reflected written forms of the Romance languages. Inevitably, only the literate minority of the population learned to read and write them, alongside Latin. The illiterate public nonetheless heard them read aloud, for instance in the recital of poems or prose, theatrical representations, public readings and, sometimes, elaborate discourses. These regional varieties possessed a high degree of prestige, situated halfway between local varieties, which were of low prestige, and Latin, which had not entirely lost its status as a sacred language. It is likely that these varieties were spoken in cultivated circles and among the ruling members of society, and that they subsequently became spoken acrolects.

The nature of the interaction between the different varieties was complex. Local varieties used in large urban centres undoubtedly had more prestige than those spoken in rural areas; words or forms used across a broad geographical area must have served to increase the linguistic prestige of a text or discourse in the same way as Latinisms.

Written and spoken Latin was the high-prestige language *par excellence*. The literate population considered it as belonging to the same linguistic continuum as the

39 Cf. Lebsanft, *Hören und Lesen im Mittelalter*, 1982 and *Die eigene und die fremden Sprachen in romanischen Texten des Mittelalters und der frühen Neuzeit*, 2000.

Romance languages, despite the fact that it had become incomprehensible for the illiterate by the 10th or 11th century (cf. 10.3.5 no. 3). The latter were nevertheless regularly confronted with the language ‘of the rich’, i.e. that used by feudal lords, the Church and the tribunals. Latin also contributed to the development of diastatic variation within Romance, both written and spoken: the decision to Latinise word forms or to use words denoting abstract concepts increased the prestige of a language (cf. 4.3.4).

In addition to this continuum consisting of local regional Romance varieties at one extreme and Latin at the other, people were also confronted with varieties of other Romance languages whilst engaging in commerce, as well as during wars and pilgrimages. It is likely that these varieties remained intelligible to non-native speakers; even nowadays, for example, a speaker of Italian or Portuguese is able to understand a speaker of Spanish, provided he or she speaks slowly and distinctly. The network of different varieties was further complicated by the existence of the minority languages mentioned above.

10.4.5 The emergence of the national Romance languages

The concept of the nation-state as we know it did not exist in the medieval world. Though Romance writing developed amongst the elite circles of society and, much like political influence, was naturally associated with the representation of power, it remained within regional boundaries. Between the end of the 19th and the beginning of the 20th centuries, European language historians wrongly assumed that national languages had existed since the Middle Ages (i.e. in the form of de-dialectalised languages displaying minimal grapho-phonetic and morphological variation, which were understood and used by everyone).

France provides a characteristic example of such assumptions. The state is presumed to have existed since Hugh Capet (the first of the Capetian Kings, who reigned during the 10th century) or at the latest since Philip II Augustus (1165–1223, the first to assume the title *Roi de France* [“King of France”] rather than *Roi des Francs* [“King of the Franks”]). Moreover, Paris is considered to have already held the role of a capital city; the language of this centralised kingdom was thought to have been the (popular) spoken language of the Île-de-France, making it the direct ancestor of the French spoken by Voltaire and Pompidou. Although this hypothesis possesses some element of truth, it requires nuancing:

1. From a geopolitical point of view, until the mid-15th century the territory of France scarcely covered half of its present-day geographical area (cf. 10.5.1). Even in the *oïl* domain, the king was first among equals, alongside influential princes such as the dukes of Burgundy or Lorraine. The Île-de-France presumably played a unique role both in the concrete as well as the more abstract spheres of power – one need only consider the polysemy inherent in the term *France*, which designated both the entire *oïl* domain in general and the Île-de-France specifically. The royal court,

however, and by association the royal chancery, were not yet fixed at a specific location, but moved frequently between the *Pays de la Loire* and the region of Paris (it should be noted, however, that Paris played an important role due to the size of its population and the large number of *scriptoria* within its walls and on the outskirts of the city; cf. Lodge, *A sociolinguistic history of Parisian French*, 2004).

2. From a linguistic point of view there was no single written language, even within the boundaries of the *oïl* domain; rather, there was a network of interdependent regional *scriptae* of varying prestige (cf. 10.3.4 no. 4 above). Though it is true that the *scripta* of the royal chancery and the literary language used by esteemed poets who were close to the Court were almost completely devoid of regional features, these elaborated varieties had little in common with spoken language, and even less with the popular language used by the inhabitants of specific regions of France (i.e. the *Pays de la Loire* or the *Île-de-France* region). On the contrary, the neutrality displayed by the *scriptae* was the result of an artificial process of de-dialectalisation and partial Latinisation.

While the written variety used by the royal court and its entourage paved the way for the linguistic model which was to be used throughout pre-modern France, this process essentially took place during the 14th century and involved far-reaching internal changes (cf. Glessgen, *La genèse d'une norme*, 2017). It is therefore legitimate to postulate a link between the French national language and the emergence of the modern nation of France; however, the actual beginning and development of such a process took place towards the end of the Middle Ages.

Although political historians revised the notion of medieval centralisation several decades ago, the nationalist myths that were forged at the end of the 19th century have persisted amongst language historians with surprising tenacity. Language as a powerful vehicle for identity has thus been subject to ideological reinterpretation (cf. 10.6.3).

In the majority of recent handbooks that address the history of the French language, one still encounters the claim that *Francien* (translated from German *Francisch*, a neologism created by Suchier in 1889), the language spoken in the *Île-de-France* region, was the basis for modern (written) French. In reality, the latter is the result of a highly complex process of secular linguistic elaboration, which, in the modern period, gave rise to a semi-artificial, literary and documentary language variety that would only become established as the spoken language much later on. The *Île-de-France* region naturally had its own spoken dialect; however, the variety that was to be labelled '*Francien*' was not a spoken language but rather an elaborated variety created mainly by the royal court and chancery. This explanation is more complex than the traditional interpretation, but it contains no inconsistencies and it is coherent with the evidence provided by textual traditions.

Interestingly, a hypothesis similar to the traditional explanation which prevailed for '*Francien*' appears more plausible in the case of Castilian. After the *Reconquista* at the

end of the 13th century, Castilian became a Romance variety of prestige, having undergone de-dialectalisation to a great extent. Language elaboration under the direct influence of Alfonso the Wise completed the process, providing this variety with an elaborate vocabulary and complex syntax, which was subsequently reinforced by the use of the language in a wide variety of textual genres. At the same time, the intense internal migration that accompanied the reconquering of formerly Muslim territories brought about phenomena of neutralisation, also in spoken language, which led to relative consistency between written and spoken Castilian (or Spanish). Owing to these particular circumstances, Spanish was the first among the Romance languages to acquire the essential characteristics of a modern standard language.

The situation in Italy is different again: the language that would ultimately become established as the national language owes much to the literary language of the *Tre corone* (the three 14th-century authors Dante, Boccaccio and Petrarch), all natives of Tuscany whose texts were used as the basis of an artificial model for written Italian in the 16th century. Historiographical works from the late 19th and early 20th centuries interpreted this as evidence for the dominance of Tuscan during the medieval period, to the detriment of all other varieties in use throughout the Italian Peninsula. Research over the past decades has led to the revision of this interpretation: in fact, many of the highly diversified Italo-Romance dialects underwent a process of written elaboration between the 14th and the 16th centuries, the tradition often continuing into the following centuries. The elaborated forms of the Piedmontese, Venetan, Roman, Neapolitan or Sicilian dialects established themselves alongside Tuscan and they too found their expression in literary as well as legal, religious or medical writing (for northern Italy, cf. Videsott, *Padania scrittologica*, 2009).

It is, however, a fact that Tuscan, the language of the rich, powerful cities and important feudal lords in Tuscany, was used by numerous writers, patrons and influential merchants well versed in the art of writing, and the production of works written in this variety was therefore exceptionally high from the Middle Ages onwards. Thus, over time, it developed a certain influence on other written varieties of Italy (cf. the early study by Ernst, *Die Toskanisierung des römischen Dialekts im 15. und 16. Jahrhundert*) and it replaced many of these from the 16th century onward. The linguistic autonomy of the different *scriptae* of Italy, however, remained very pronounced throughout the 14th and 15th centuries.

The medieval period thus saw the rise of the Romance languages, the establishment of their European territorial basis, as well as the first stage of their elaboration, which nevertheless took place under the heavy influence of Latin. It was also a period of non-dogmatic multilingualism founded on spoken and written regional varieties and characterised by a relative lack of conflict.

10.5 The modern period (1500 – end of the 19th century)

First and foremost, the modern period witnessed the appearance of standardised Romance languages. Not only Romance, but also English, German and other national European languages follow the same chronology. The reasons for such parallels are not to be sought in the internal characteristics of the languages in question but rather in the external transformations that took place in political and socio-cultural spheres in Europe. In the modern period, the need for formal and elaborated long-distance communication increased substantially. These new communicational needs were met by the development of diatopically neutral languages, characterised by minimal internal graphic-phonetic and morphological variation, syntax that allowed complex constructions, as well as a highly-developed vocabulary.

Written sources covering a larger geographical area multiplied during the modern period. Linguistic diversity within the diasystem, too, increased significantly, parallel to standardisation. Finally, the number of studies involving analyses of individual languages or the comparison of different languages saw a sharp rise. In this section, only the elements that have been of most significance for the evolution of the Romance languages will be treated.

10.5.1 Geography, geopolitics and demographics

In terms of demographics, the population of the *România* was subject to strong fluctuations between the 1st and the 17th centuries, although it did not exceed a certain limit. We can assume that the number of Latin speakers under the Roman Empire was roughly equal to the number of those who spoke a Romance language in the 12th or in the first half of the 14th century, before the plague of 1348, amounting to about 30 million (Kramer, *Die Zahl der Sprecher der Sprache Roms in Geschichte und Gegenwart*, 2000).

At most, two million of these were literate, and from the 6th to the 9th centuries this figure decreased even further. Therefore, the direct, written source material covering almost two millennia of Romance linguistic history only concerns 5 to 8% of the total population, with women being particularly poorly represented amongst this number.

Between the 17th and the 19th centuries a demographic expansion took place, catalysed by geopolitical changes, amongst other factors. The distances a single Romance language had to cover were increased by two complementary phenomena: the establishment of the territories occupied by the modern nations and colonial expansion. Both of these deeply marked civilisations on a global scale. The specific manifestations of these general tendencies, however, differed from country to country:

The national territory of present-day France was shaped between the 15th and the 19th centuries:

- after the Hundred Years' War (1337–1453), the king of France ruled Normandy and Picardy, Île-de-France and Champagne, the Centre (Anjou, Orléanais, Berry) and Poitou as well as a large part of Gascony and – after the Albigensian Crusades (1209–1229) – the County of Toulouse;
- in the following century, the royal domain was enlarged by the Duchy of Burgundy (1478), Provence (1482–1486), the Duchy of Brittany (1491–1532) and the County of Flanders (1529);
- this was followed by Béarn and the County of Foix in Gascony (1607), Alsace (1648 [Strasbourg 1681]), Roussillon (1659), Franche-Comté [= the County of Burgundy] (1668–1678); later still, by the Duchy of Lorraine (1766) and Corsica (1768), and finally by Savoy and the County of Nice (1860);
- during the same period, the French colonial empire expanded: from the 16th to the 18th centuries, Canada, Louisiana, the Antilles, French Guiana and several islands in the Indian and Pacific Ocean became part of the kingdom; during the 19th and the beginning of the 20th century, these territories were joined by the Maghreb and part of sub-Saharan Africa.

Not until the dawn of the French Revolution did the territory of metropolitan France correspond approximately to the country as we know it today. It was also under the *Ancien Régime* (1481, Louis XI, to 1789, Louis XVI), and not in the Middle Ages, that the administrative centralisation of the state in the region of Paris began to take place – in reality a phenomenon typical of modern France.

Due to political hegemony of the centre over other areas, the regional languages of the territory of France were marginalised in favour of the national language, a fate similar to that which befell the indigenous languages under the Roman Empire as a result of the spread of Latin. In both cases, it was the desire for imitation, rather than pressure, that led local elites to adopt the language 'of the royalty'. French thus became the only language used by the elites and it extended its monopoly on written language, while the minority languages lost the role of expansion varieties they had had during the medieval period, as well as the resulting prestige. The majority of the population, however, continued to speak their native dialects of French, Francoprovençal, Occitan or Gascon, Basque, Breton, Germanic or Corsican until the First World War (*cf.* Carles/Glessgen, *Les écrits des Poilus*, 2020).

Extra-European expansion, which was a constant preoccupation of the French Crown and of French society in general, led to the emergence of new secondary dialects, such as Canadian French. By studying the population of this country, it has been possible to shed light on the different stages as well as on the mechanisms of formation typical of a colonial language variety (*Cf.* Chauveau/Lavoie, *A propos des origines dialectales du lexique québécois*, 1993; Thibault, *Histoire externe du français au Canada*, RSG 1, art. 76; Martineau/Remysen/Thibault, *Le français au Québec et en Amérique du Nord*, 2022).

Moreover, contact between highly divergent languages leads to phenomena of pidginisation and creolisation, which can mainly be observed on the various islands subjected to colonisation. In the most recent colonies of sub-Saharan Africa and the Maghreb, language change more often resembles secondary dialect formation, and only a few cases can be described as pidginisation (cf. Lafage, *Français parlé et français écrit en pays Éwé (Sud-Togo)*, 1985; Glessgen, *Das Französische im Maghreb*, 1997a).

The significant changes undergone by French in these extra-European communities had little effect on the evolution of French in Europe – in contrast to the situation observed for *Ibero-Romance*. Until the 19th century, the neighbouring regions of French-speaking Switzerland and Walloon-speaking Belgium kept pace with France with respect to most trends in language change.

The political successes enjoyed by the French in Europe nevertheless transformed the variety of the French kingdom into a widespread language used in European courts and higher circles of society in the 17th and 18th centuries. More sporadically, French became a language of domination in the departments east of the Rhine in Germany, following the French Revolution. The impact of the Revolution on linguistic awareness and language politics, on the diasystem, and more particularly on political vocabulary and rhetoric was of great significance (cf. Schlieben-Lange, *Idéologie, révolution et uniformité de la langue*, 1996, and for a more general overview of this period of modern language history, Völker, *Politik, sozioökonomische Entwicklung und Sprachgeschichte: Gallo-Romania*, RSG 2, art. 103).

The territory of Peninsular Spain was constituted in 1492 when the kingdom of Grenada was conquered (an event followed by the edict expelling the Muslims in 1609). The union of the Aragonese and Castilian crowns in 1475 and their political control over Galicia from 1640 onwards allowed written Spanish to gain hold across the entire territory. Once again, Spain witnessed the homogenisation of its territory and its language at an earlier stage than France. The European expansion of the Habsburg Empire outside the Iberian Peninsula, in particular in the Spanish Netherlands, did not lead to specific trends of language evolution, unless the indirect influence resulting from its contribution to the structure and hierarchy of the power of the *Antiguo Régimen* (1514–1833) is taken into account. The latter underwent internal transformations, especially during the time of Carlos III (1759–1788), but its organisational principles remained dominant until the decline of the monarchy after the Spanish-American war in 1898 (cf. 10.5.3 no. 7 for Catalonia/Valencia and Galicia).

In contrast, extra-European conquests radically changed the conditions of language evolution. The first expeditions led the Spanish colonists and conquistadores along the coasts of northern Africa toward the Canary Islands (from 1304–1480), then towards the Caribbean in the modern period (beginning in 1492) and finally to Central and South America (1519 and 1536–1540, respectively). Towards the end of the 16th century, the strategic areas of the Continent were in the hands of the Spanish, who structured the territory by creating *Virreinos* ('viceroyalties': Nueva España 1535, Perú 1553).

The history of this conquest has come to represent the archetypal process of colonisation in the modern imagination, i.e. colonisers with an elaborate military and political organisation (horses, firearms, hierarchies, continual support from the continental base) conquer a pre-modern nation, at the same time introducing diseases to which the indigenous population have no resistance. It should be remembered that over the centuries, the high population density in Europe had produced a great variety of diseases, but also the resistance with which to fight them (*cf.* Diamond, *Guns, Germs, and Steel: The Fates of Human Societies*, 1997, who suggests a connection between the domestication of animals, the elaboration of writing, population density and cultural evolution).

The consequences of this confrontation are well known but nonetheless horrifying. The indigenous population of the Caribbean was diminished within a few decades, while the Indian population of New Spain was decimated from around 18 or 25 million to one million within a single century. Reduction in population was comparable throughout the Americas, including North America, owing to the rapid spread of disease-carrying microbes. This would later favour the establishment of English colonies on the Continent. The wholesale import of African slaves to Latin America and, especially, to English-speaking Protestant America ultimately led to the destruction of the existing structures on both continents.

The percentage of the Spanish population originating from the Peninsula remained extremely low until the 18th century (*cf.* Glessgen, *Historia externa del español en México*, RSG 2, art. 84). Nevertheless, it was the linguistic model of peninsular Spanish that would gradually become established, due to the influence of cities, educational institutions, social and religious structures and, more particularly, the close relationship with Europe, which meant that the Spanish language prospered in spite of the distance.

Thousands of indigenous languages were marginalised by Spanish, which became the official language of the new independent nation-states founded between 1810 and 1830 (Cuba and the Dominican Republic in 1898; Puerto Rico was annexed by the United States at the same time). It was not until the second half of the 20th century that some indigenous languages such as Quechua in Peru (1975) or Guaraní in Paraguay regained their importance and prestige.

The establishment of Spanish in America, associated with strong internal migrations over the centuries, followed the pattern typically associated with the emergence of secondary language varieties. Until the 19th and 20th centuries, diatopic variation of American Spanish was still almost inexistent and did not yet display organised geographical distribution comparable to that of the primary Romance dialects in Europe. Phenomena of differentiation that can be observed between the new national and regional norms (*cf.* 10.6.1), are the result of more recent developments, the majority of which took place after the Second World War.

The Expulsion of Jews was a historical episode of a very different nature, taking place in 1492, the year which also saw the fall of Grenada and (Western) Europe's discovery of America. This radical decision put a sudden end to a century-long conflict, but also to the coexistence of three cultures – Christian, Jewish and Muslim – to which Spain

owed much of its modernity as well as the high degree of elaboration characterising its medieval culture (cf. the classic essay by de Castro, *De la edad conflictiva*, 1961). While from a geolinguistic point of view the main consequence of the expulsion of the Jews was the emergence of Judeo-Spanish (in the Maghreb, Balkans and Israel; cf. Albrecht, *Romanische Migranten- und Vertriebenensprachen: die Sprachen der iberischen Halbinsel*, RSG 2, art. 169), from a socio-cultural point of view, Spain lost an important source of innovative potential. This violent banishment had lasting repercussions for the country that were to make themselves felt throughout the entire modern period.

The history of Portugal and the Portuguese language displays similarities to that of the Spanish kingdom and Spanish; moreover, Portugal was under Spanish rule from 1580 to 1640. The European territory of Portugal was constituted from the very beginning of the modern period and the language displayed relatively weak diatopic variation as a consequence of the *Reconquista*. Its early expansion to Africa (Angola, Mozambique, Guinea-Bissau and Cape Verde), South America (Brazil) and Asia was characterised by the appearance of multiple pidginised varieties, especially in Asia, as well as of several secondary dialects, of which Brazilian is the most prominent. Overseas territories of Portugal gained their independence (e.g. Brazil in 1822) and today constitute the most significant demographic basis of Portuguese. Finally, towards the end of the modern period, Portugal witnessed an economic decline in favour of northern Europe. The foundation of the Republic of Portugal in 1910 gave rise to a period of turmoil and crisis.

A very different geolinguistic scenario characterises Italy, a nation whose culture continued to dominate Europe throughout the modern period. From a linguistic point of view, it had nevertheless inherited regional structures from the medieval period. There were five major centres of power in (pre-)modern Italy – Milan, Venice, Florence, the Papal States and Naples (with Sicily) – in addition to a large number of small states and feudal domains.

In the modern period, this heterogeneous territory was subject to attacks, conquests and the continual usurpation of power by France, Spain (which dominated the kingdom of Naples between 1494 and 1713–1714) and Austria. In contrast to the other previously mentioned Romance-speaking countries, therefore, Italy did not exhibit territorial and political unity before the Napoleonic conquests (1805–1808) and the formation of the Republic of Italy (1861; 1870, with Rome as its capital). For this reason, it did not expand its territory beyond Europe, with the exception of a few short-lived colonies in North and Northeast Africa.

This is all the more remarkable since the formation of standard Italian shows certain similarities to that of the French standard in the 16th–19th centuries. There are thus parallels between the communicative processes which developed in the countries of Europe, even though varying geolinguistic situations ought to have led to more significant differences.

The political diversity of Italy was nevertheless not without consequences: it enables us to understand how major authors could continue to write in elaborated dialects until the 19th and the beginning of the 20th centuries, a completely different situation compared with that of Occitan in France, not to mention Gascon or Francoprovençal. Consider, for instance, the Venetian plays of Goldoni, Milanese poetry by Carlo Porta, the works of the Neapolitan writers Giambattista Basile (17th century), Salvatore Di Giacomo (19th–20th centuries), Edoardo de Filippo (20th century) or, in Rome, the writings of G. G. Belli (18th–19th centuries). It is also the reason for the survival of Italo-Romance dialects to this day: unlike in France (but similar to the situation in Germany), the dialects of Italy did not disappear after Italy became a nation-state at the end of the 19th century.

Finally, the political and linguistic history of Romania is altogether different: feudal states that used Romanian as a courtly language and for writing emerged for the first time in the 16th century. After its emancipation from the Boyar class of landowners and from the neighbouring Turkish and Austro-Hungarian powers, the Romanian State was proclaimed as the union of Wallachia and Moldova (although it did not yet include Transylvania) during the same period as the Italian State (i.e. in 1862; it achieved independence from the Ottoman Empire in 1878).

From the end of the 19th century, the variety of French used in writing contributed to the formation of the modern national language and, in doing so, destroyed many unique aspects of Romanian that had developed over the preceding centuries. For this reason, in addition to the absence of medieval documentation, the linguistic history of Romanian, regrettably, is not as rich as that of the other Romance languages.

The geolinguistic differences between the Romance-speaking countries are therefore considerable, despite the parallel evolution of major linguistic trends. The diversity of concrete situations would become even more apparent if we were to consider Italian- or French-speaking Switzerland, Galicia and Catalonia, the Ladin-speaking valleys of Italy or the regions in which Aromanian is spoken.

During the modern period, moreover, sporadic situations of language contact of various sorts arose even within the European *România*, such as that resulting from the emigration of the Francoprovençal-speaking Waldensians (*Vaudois*) to the regions of Siegen and Hesse (north of Frankfurt, in Germany) in the 17th century for religious reasons.

10.5.2 Infrastructure and socio-cultural evolution

Socio-cultural change as well as geopolitical evolution contributed to the elaboration of modern standard languages. The modern period witnessed a redistribution of socio-economic power and cultural development between southern Europe and central and northern Europe, before the latter became dominant. Moreover, in northern Europe during this period, the capacity for agricultural and artisan production greatly

increased, followed by the development of the industrial sector, with numerous technological discoveries, and the growth of military power. As a result, Europe came to surpass not only its immediate neighbours of Mediterranean Europe and the Near East, but also China, which, after a number of exceptional developments in agriculture, technology, infrastructure and writing, witnessed periods of stagnation.

In modern Europe social classes underwent major upheaval as interaction between them intensified. The feudal pyramid with its relatively simple hierarchy was replaced by more finely differentiated social structures and more elaborate forms of political organisation. A primary role in this evolutionary process was played by large cities, where almost all established princely courts were located, including Paris, Madrid and Naples (all counting between 300,000 and 500,000 inhabitants in the 17th century), Venice, Milan, Palermo and Rome (with 100,000 to 200,000 inhabitants), as well as Lisbon, Seville and Lyon (with at least 100,000 inhabitants).

Several other agglomerations situated outside the *Romània* were equally important during the modern period. At the end of the 17th century, the largest European city was London (with 550,000 inhabitants), followed by Paris (415,000 inhabitants in the mid-17th century); Amsterdam with 172,000 and Moscow with 130,000 inhabitants were on the same scale as the larger Italian cities. Outside Europe, Constantinople had 700,000 inhabitants at this time, and the largest city in the world was Ayutthaya (Thailand), which had almost a million inhabitants around 1700 (*cf.* Lodge, 2004). Nevertheless, the aforementioned cities of the *Romània* were among the world's largest urban centres of the period, in contrast to the situation at the beginning of the 21st century.

These large cities and the courts attached to them displayed a new socio-cultural dynamic, in which architecture, the arts, sciences and literature, as well as numerous professions associated with these fields played a part, as did administrative and military institutions (*cf.* the dated but very vivid description by Sombart, *Liebe, Luxus und Kapitalismus*, 1922). In financial terms, this development drew in part from extra-European sources, especially those of the Americas, which contributed significantly to the wealth of the European world.

Ironically, it was during the centuries of the Renaissance and the Age of Enlightenment that numerous wars devastated Europe, beginning with the religious wars following the Protestant Reformation and the Counter-Reformation. As much as 40% of the European population was decimated during the Thirty Years' War, which took place between 1618 and 1648. All this took place against the backdrop of the Inquisition, which, combined with witch hunts and a system of capital justice, contributed to the staging of an apocalyptic theatre of horrors (*Theater des Schreckens*, van Dülmen, 1988).

It was an era profoundly marked by contradiction and upheaval, inseparable, furthermore, from the factor of population growth, which was scarcely hampered by the wars and epidemics. A wide chasm separates the 16th century – a pivotal era between the Middle Ages and the modern period – from the 18th, during which thought emancipated itself from the Church to become secular for the first time in the Western World.

In the context of this *Companion*, only a very limited number of these evolutions, which are well known and have been described extensively in specialised literature, can be considered. For the purposes of language history, we shall restrict ourselves to emphasising the intensification of different forms of cultural elaboration and the development of princely courts and metropolitan areas as major cultural centres. Both are processes that are intimately linked to language elaboration.

10.5.3 Written culture and language elaboration

1 The establishment of new ‘umbrella languages’

The elaboration of the Romance varieties and standard languages during the modern period was both accompanied and driven by the expansion of the range of their communicative functions. In official and public contexts, these languages prevailed at the expense of both Latin and other Romance varieties and regional languages. This process, which had been in gestation since the Middle Ages, unfolded progressively (*cf.* 10.4.3 no. 3). Textual genres gradually moved from Latin to French, Spanish or Italian, until the old language of prestige was confined solely to religion and the sciences (including philosophy, history, biology, physics, etc.). In the last few decades of the 19th century Latin finally deserted even these last strongholds.

As far as the old regional varieties were concerned, some merged naturally with the new national variety. This is the case for the regional *scriptae* throughout the *oil* domain, where the adaptation to the variety in use at the royal chancery throughout the 15th century did not imply a total break with former writing habits. In Portugal and, to an even greater extent, in the territory occupied by Castilian, regional *scriptae* had not developed during the medieval period, as the linguistic varieties in question were expansion varieties rather than varieties originally based on highly-differentiated oral dialects.

In other cases, by contrast, the replacement of former regional varieties of Romance by the new standard language caused linguistic discontinuity. The abandonment of Occitan, Gascon, Galician and Catalan in the 16th century in favour of French or Spanish are extreme examples (*cf.* no. 7 below). Another deep internal rift emerged when the ‘illustrious’ Tuscan variety (archaic and homogenised) replaced Italo-Romance dialects which had had their own written tradition, such as Genovese, Milanese, Venetian, Roman, Neapolitan and Sicilian, or when Spanish varieties such as Leonese, Asturian and Aragonese were replaced by Castilian. In Switzerland, a French written tradition already existed alongside the indigenous Francoprovençal language during the Middle Ages. Thus, diglossia continued and intensified during the modern era.

As has already been mentioned with regard to written language, in Italy, a number of strong autonomous political powers reached a linguistic compromise very similar to

that of the centralised kingdoms of France, Spain and Portugal (*cf.* 10.5.1). In contrast, the means employed in the elaboration and diffusion of the new ‘national’ languages differed, depending on the country and the period in question. A large number of cultural factors also contributed to language elaboration, although these did not always occur together and were not always all necessary. Rather, a relatively free combination of several of these elements led to the same result throughout the communicative space covered by the four new ‘umbrella languages’, which were French, Spanish, Italian and Portuguese (each of these covered the corresponding geolinguistic territory and served as an exemplary variety within the general architecture of the language, *cf.* 3.1; 4.2.3).

2 Written culture

Written culture underwent significant intensification and diversification in the modern period. The development of textual genres continued those that had emerged during the medieval period, becoming more varied and more specific. The geographical and social basis upon which their production was founded also expanded. Outside the domains of literature (which was the exemplary form that served the purpose of cultural representation of the elite), administration and law, there was above all a development of ever more specialised knowledge. A growing number of texts were translated from Latin or written directly in a Romance language, e.g. in medicine and pharmacology, in architecture, the arts or musicology. Other genres developed directly in the vernacular, such as travel accounts, inspired by discoveries outside Europe, or autobiographical writing, which reflected a new image of the individual in society. In all cases, an unprecedented development of specialised writing can be observed, that at the same time showed a number of similarities to texts of an aesthetic or literary nature.

The social basis of textual production also grew considerably. From the 16th century onwards, for example, a body of Spanish literature from the Americas began to take shape thanks to the writing of both Spanish conquistadors and missionaries of European origin as well as indigenous and mestizo writers (often of noble descent), telling of the encounter between the Old and the New World (*cf.* Bernand/Gruzinski, *Histoire du nouveau monde*, 1991–1993; *cf.* no. 5 for travel narratives; *cf.* Oesterreicher). The 17th and 18th centuries then witnessed the development of writing of a ‘private’ nature by authors who did not belong to the social elite (*cf.* Ernst, *Textes français privés*, 2019). Literary expression thus gradually detached itself from the royal courts and centres of power from whence it had emerged.

Not only did textual genres diversify and increase in number, but their diffusion also broadened thanks to the printing press, which, after a century of trial and error, had become a highly efficient tool. The necessity of producing texts which no longer concerned a mere handful of individuals or a few dozen readers but hundreds or even thousands of people contributed to the homogenisation and neutralisation of language. Major printers thus had a direct and crucial part in the elaboration and diffusion of new linguistic norms in the 16th century (*cf.* 11.2).

Religious writing played a central role in this process, especially the Bible, as a text constituting a moral and cultural point of reference. It is therefore no coincidence that this key text was the very first to be printed – in Latin – by Gutenberg in 1455–1456. New translations of the Bible in the spirit of the 16th-century Reformation greatly contributed to the search for a consensual form of language, not only in the well-known case of the Lutheran Bible, but also, for example, in French translations of the Old and New Testament from the Vulgate version by Lefèvre d'Étaples (1523–1528) and the Greek Bible by Olivetan, prefaced by Calvin (1535, *cf.* the example of the sentence from the *Pater Noster* discussed in 3.5.2; *cf.* also 11.1.2).

For Romance languages with a low number of speakers, such as Romansh, Ladin or Friulian, Bible translations played a fundamental role in the elaboration of modern standard varieties. The first actual text written in Romansh was a Bible translation (*Bible de Bifrun*, 1560; *cf.* *Übersetzungen ins Friaulische; – ins Dolomitenladinische; – ins Bündnerromanische*, RSG 2, 119a-c; *cf.* 11.2.2 no. 1 below). Note that the first text to be printed in Romanian in 1544 was a catechism (*cf.* Tagliavini, *Le origini delle lingue neolatine*, 1982, § 87).

The main consequences of the diversification and intensification of writing were the homogenisation of grapho-phonetic and morphological forms and the enrichment of the vocabulary of various specialised fields. In the 18th century, a remarkable development of terminology took place in the domain of the natural sciences, which was also fuelled by the encyclopedic tendencies of the Enlightenment period⁴⁰ (*cf.* no. 4 below).

16th- and 17th-century texts, moreover, displayed increasing syntactic complexity – a factor that is indispensable for decontextualised and condensed forms of communication. Interestingly, the most elaborate syntax was required by two particular categories of texts: specific literary genres with constraints, such as poetry or, sometimes, theatre, and theoretical discussion. Poetry undoubtedly enjoyed a golden era in the European courts of the modern period; similarly, in the area of theoretical knowledge, the Romance languages advanced significantly compared with Latin. To cite a few key events with regard to the French language, Petrus Ramus (Pierre de la Ramée) was the first to teach in French at the *Collège des trois Nations* in 1529, the *Institution de la Religion chrétienne* was published by Calvin in 1541, and the *Discours de la méthode* by Descartes appeared in 1637. These dates mark the beginning of a new era of theoretical, theological and philosophical discussion, which was expressed by means of the new standard languages.

40 *Cf.* e.g. Buchi, *Les noms des genres dans la flore française de Lamarck* (1778), 1994; Serianni, *Il secondo Ottocento*, 1991; Möhren, *Wort- und sachgeschichtliche Untersuchungen*, 1986 (which deals with the French Encyclopedia of 1765).

3 Linguistic thought: theory, grammar and lexicography

Linguistic thought contributed even more directly to the movements towards language standardisation of the modern period. In the 16th and 17th centuries, discussions revolving around norms and the nature of language as well as grammatical analysis and lexicography witnessed an unprecedented development. Although these topics concern both the external history of language and the discipline of linguistics, they will be treated in this section due to their impact on the evolution of standard languages and thus on the constitution of modern society.

Linguistic thought that resulted in the formulation of explicit, prescriptive norms for the new national languages was manifest in different forms: considerations relating to graphic representation (or orthography), the evaluation of morphological forms or syntactic constructions, grammatical treatises, hypotheses on the origins and history of language, and works of lexicography forged by the Encyclopedic tradition – these last were closely followed by high-quality linguistic and semasiological dictionaries. Reference works produced during the modern period abound for French and Italian, but also for Spanish and Portuguese, as in other fields of specialised knowledge. As this is a relatively well-studied and easily accessible subject, only a number of fundamental works will be mentioned here.

Considering the example of French, the following works were essential to the discourse revolving around language norms and standardisation:

- in the 16th century: Joachim Du Bellay, *La Deffence et Illustration de la langue françoise* (1549), which is based on Sperone Speroni's Italian treatise *Dialogo delle lingue* (1542), a debt which remained unacknowledged;
- Henri Estienne, *Essai sur la Précellence du langage françois* (1579);
- at the beginning of the 17th century, the poetry of François de Malherbe (1555–1628) had a direct influence on the royal court⁴¹;
- shortly thereafter, Claude Favre de Vaugelas, secretary of the *Académie française* (cf. no. 4 below), theorised the idea of language use (French *usage*) (as an inter-subjective factor, rather than absolute) in his *Remarques sur la langue françoise* (1647). According to Vaugelas, the notion of *bon usage* corresponded to 'la façon de parler de la plus saine partie de la Cour' ('the way of speaking of the soundest part of the Court'). Vaugelas was also responsible for the preparation of most of the articles of the later *Dictionnaire de l'Académie* (1694; cf. once again no. 4 below).

⁴¹ Malherbe's role in Paris was to be immortalised in Nicolas Boileau's emblematic phrase "Enfin Malherbe vint [...] / Et réduisit la Muse aux règles du devoir" (*Art poétique*) ('Finally came Malherbe, [...] and reduced the Muse to the rules of duty', after the *Cambridge History of Literary Criticism*, vol. 4, 1997: 564), which is often quoted in works of linguistic historiography.

The above-mentioned treatises have a direct link with grammatical analysis, and the different types of works were often written by the same authors. Precursors in the domain of thought on language include Geoffroy Tory's *Champfleury* (1529), a treatise on calligraphy and typography, and the English scholar John Palsgrave's *Lesclaircissement de la langue françoise* (1530), a teaching grammar of French, written in London. A semi-phonetic system of orthography was developed by Louis Meigret in his famous *Trehtë de la grammeŕe françoëze* (1550). A large number of other works marked this period, including:

- Jacques Dubois (Lat. Iacobus Sylvius), *In linguam gallicam isagōge* (1531)
- Robert Estienne, *Traicte de la grāmaire françoise* (1557; cf. also no. 4 below)
- Pierre de la Ramée (Lat. Petrus Ramus), *Gramere* (1562, ²1572)
- Henri Estienne (Lat. Henricus Stephanus, son of Robert Estienne), *Hypomneses de gallica lingua* (1582)

These works on grammar, which included orthography, morphology and (in part) syntax, reached a new level of quality in the 17th century in the *Grammaire générale et raisonnée* by Antoine Arnauld and Claude Lancelot (1660). The *Grammaire générale* distanced itself from the rather prescriptive character of its predecessors and placed more emphasis on interpretative elements (hence the epithet *raisonnée*, i.e. 'rational'). It played a decisive role for the grammatical theory 'of Port-Royal', which was, in turn, fundamental for modern theories of morphology and syntax (cf. Bierbach and Pellat, *Histoire de la réflexion sur les langues romanes: le français*, RSG 1, art. 21, ch. 4.3).

Such works – descriptive and normative on the one hand, and theoretical and interpretative on the other – had no immediate impact on written language. Nevertheless, all these publications contributed to the complex process of homogenisation and 'normalisation' of writing ('orthography') and morphology, as well as to the development of a more complex syntax. There is no doubt that the extreme elaboration and the artificial regularisation of French were catalysed by the ideology of the centralised state that France had become.

In Italy, linguistic issues were rather different, since there was no 'natural' candidate to assume the role of a standard language comparable to the variety of the royal court and chancery in France. Thus, during the 16th century, a lively debate ensued around the '*questione della lingua*', which sought to determine the general physiognomy of the new system of norms, considered necessary for all.

Grammatical analysis had developed very early on in Italy, beginning with the *Grammatichetta* attributed to the humanist Leon Battista Alberti (1430–1440, but only published in the 20th century). The intellectuals involved in this discussion were numerous: authors such as Baldassarre Castiglione (*Il Cortegiano*, 1528) defended models developed by certain princely courts (such as Urbino or Mantua) or the Papal court. These were joined by proponents of a more 'pan-Italian' movement, including Gian Giorgio Trissino (*Il Castellano*, 1529), and the '*italianisti*', who were in search of cross-re-

gional norms that were not limited to Tuscan. Other authors such as Machiavelli, Gelli, Giambullari or Varchi prescribed the use of the Florentine variety of their own period (hence the names ‘*toscanisti*’ or ‘*fiorentinisti*’ attributed to its proponents).

The language model that would ultimately assert itself was an adaptation of the archaic literary Tuscan language, based on the canon of the *Tre corone*: Dante, Boccaccio (the model for prose) and Petrarch (the model for poetry). This principle was elaborated by Pietro Bembo in his *Prose della volgar lingua* (1525). He received partial support from partisans of a fourth trend, represented by Salviati, who sought a compromise between 14th-century and 16th-century Florentine (cf. Viale, *Studi di storia della lingua italiana*, 1992).

This norm became established surprisingly quickly once its principle had been accepted. A language which no social or socio-regional group was accustomed to using, it was consequently less subject to interference with a spoken language capable of slowing down the acceptance and generalisation of specific suggested forms. Paradoxically, it was thus the particularly artificial character of the Italian norm that enabled its rapid acceptance. Bembo’s model had become established shortly before the middle of the 16th century. In no other Romance-speaking country did the transition from a medieval written language – displaying a high degree of variation – to a modern standardised written language take place within such a short space of time.

Naturally, discussion on grammar, orthography and language theory continued in Italy throughout the 17th and 18th centuries; however, it had lost much of its interest. It was not until the 19th century that it would be revived in the form of the second ‘*questione della lingua*’, which was concerned with the language of the future Italian nation-state.

Even though Spanish underwent considerable internal changes in the 16th century, and despite the fact that the earliest work of grammatical analysis, Antonio de Nebrija’s *Gramática de la lengua castellana* (1492), appeared at a very early date in the environment of the royal court, thought on language in Spain did not have the same importance as it had in France or Italy⁴².

As in France, the basis of the standard variety had already been defined: written Castilian, which had become relatively homogeneous throughout most of the kingdom since before the 15th century. In contrast to France and especially to Italy, however, the internal distance between this language and spoken varieties of Castilian (particularly cultivated language) was not particularly significant, owing to the expansion during the period of the *Reconquista* as well as to intensive internal migrations. Thus, there were no true debates on language normalisation. The only external questions put forth at the time were concerned with rivalries between the norms of the court of Toledo and

42 It should be added that discussions on language remained even more limited in Portugal (cf. Kemmler/Schäfer-Priess, *Geschichte der Reflexion ...: Portugiesisch*, RSG 1, art. 25).

those of the court of Madrid. The predominance of the latter, however, was the result of a political decision rather than a decision made by authorities on language.

The above-mentioned antagonism reflects the intense anti-intellectual climate that reigned in Spain after the expulsion of the Jews in 1492 and lasted until the 18th, if not the 19th century. Numerous works by great language theoreticians such as Juan de Valdés were not published during their lifetime: Valdés' *Diálogos intorno a la lengua*, written in 1536 in Italy, long remained in manuscript form, as did Alberti's *Grammatichetta* (cf. 10.4.3 no. 4 for Dante's *De vulgari eloquentia*).

Even more so than in France, Spanish linguistic thought during the 16th and 17th centuries focused on language origins and change. In this context, the modern reader will sometimes encounter confusing theories: some authors were of the opinion, for instance, that Spanish had evolved from Hebrew or Basque, or that Greek was at the origin of the French language.

At the beginning of the 17th century, however, Bernardo Aldrete (1560–1641) observed phonetic correspondences between Latin and Spanish with a surprising degree of accuracy in his *Del origen y principio de la lengua castellana o Romance que oy se usa en España* (1606), a work which presaged the neo-grammarians theories of the 19th century.

E.g. libr. II, ch. X, Derivacion de los vocablos de Romance, en que se mudan unas vocales per otras:

Y para poder sacar en limpio las derivaciones, es fuerça valernos de lo que el mismo dize, que no deven ser reprehendidos los que buscando el principio, y origen de algun vocablo, ô le añiden, ô quitan letras, para que con mas facilidad puedan alcançar lo que en la diction està escondido⁴³ (cf. Berkenbusch, 1990 and the facsimile reproduced at the address <www.cervantesvirtual.com>, fol. 48v).

Owing to the accuracy of his hypotheses, Aldrete's work surpassed Ménage's etymological deliberations, however remarkable these may have been (cf. no. 4 below). Like the work of so many others, however, it remained without immediate impact.

The role of lexicography in the modern period was as important as that of grammatical description. In the 16th century, bilingual dictionaries held sway. These sought to explain one language by means of another and were primarily conceived as aids for learning Latin or a modern language. Dictionaries based on a Romance language rather than on Latin were an innovation of this period. These include:

- Antonio de Nebrija, Spanish-Latin *Diccionario* (ca. 1495, following the Latin-Spanish dictionary written by the same author in 1492).

⁴³ "Derivation [= development] of Romance words, in which some vowels are exchanged for others: And in order to be able to make out the derivations, we must make use of what he himself says, namely that those who search for the beginning and origin of a word, or add to it, or remove letters, should not be reprehended, so that they can more easily find what is hidden in the diction."

- Robert Estienne, *Dictionnaire françois-latin* (1539, following his *Dictionarium Latinogallicum*, 1538)
- Francesco Alunno, *La fabrica del mondo* (1548)

The use of a Romance language, rather than Latin, as the source language reflected a fundamental change of perspective and of linguistic prestige. This trend culminated in the publication of the first monolingual dictionaries written in a Romance language at the beginning of the 17th century:

- Jean Nicot, *Thresor de la Langue Françoise* (1606)
- Sebastián de Covarrubias, *Tesoro de la lengua castellana o española* (1611, including a large section devoted to etymology)
- Accademia della Crusca, *Vocabolario della Crusca* (1612, cf. no. 5 below).

These monolingual works, however, were initially more restrictive in their choice of words than bilingual dictionaries of the same period, such as Cotgrave's remarkable *Dictionarie of the french and english tongues* (1611) or the *Dictionnaire italien-françois et françois-italien* (1640) and the *Trésor des deux langues espagnole et françoise* (1645) by César and Antoine Oudin (father and son).

A lively interest in etymology emerged during this period, exemplified by Ménage's *Origines de la langue françoise* (published in 1650 with an improved version appearing in 1694), which, although far behind the first modern etymological dictionary by Diez (1854), nevertheless represents a milestone (cf. also the first etymological dictionary to be written for a Romance language, the *Etymologicon françois* [1572] by Jean le Bon (edited by Rézeau, 2023), as well as the accurate etymologies found in the posthumous work of Nicot [1606] which paved the way for Ménage). Throughout the entire 16th and 17th centuries, moreover, lexicography was fuelled by encyclopedic undercurrents (cf. Bierbach, *Grundzüge humanistischer Lexikographie in Frankreich*, 1997). The separation into two distinct traditions (i.e. encyclopedic and linguistic) did not occur until later; a new encyclopedic tradition *per se* would assert itself in the 18th century. For French, see the following works:

- Pierre Bayle, *Dictionnaire historique et critique* (1697)
- the *Dictionnaire de Trévoux* (1704; 1732, etc. cf. 9.9.1)
- the *Encyclopedia* by Diderot and d'Alembert (1765, cf. no. 2 above)

During the same era, the first major semasiological dictionaries were compiled, beginning with the triad of works that appeared in the late 17th century, the origins of which are closely interwoven. These are Richelet's dictionary of 1680 (which, however, still bears the title *Dictionnaire françois, contenant les mots et les choses*), Furetière's of 1690, and the excellent first edition of the *Dictionnaire de l'Académie* (1694, cf. 9.9.1 and no. 4 below), which served as a model both for written language and for future lexicography.

The mere enumeration of 16th to 18th century metalinguistic works which exist for the Romance languages would occupy several volumes (*cf.* RSG 1, art. 18–26 [*Histoire de la réflexion sur les langues romanes*] and 29–36 [*Étude et description étymologique et historique du lexique ...*]; HSK 5, *Wörterbücher*). This is an indication of the intensity with which erudite scholars worked towards language standardisation. Some brought forth exceptional works: Bembo, Robert Estienne, Valdés, Vaugelas, Arnauld and Lancelot or Aldrete, for instance. The majority were content to follow in the footsteps of their predecessors. Some authors remained unknown during their lifetimes, while others became renowned and were emulated. Considered as a whole, their efforts nevertheless represented a socio-cultural factor that played a decisive role in language elaboration.

4 Language planning and language academies

In the modern period, states intervened directly in the formation of standard languages for the first time through language legislation, the establishment of normative institutions, and literacy campaigns (*cf.* no. 6 below). These activities go hand in hand with thought on language and, depending on the historical period, the cultural influence of the royal courts.

Language legislation is an invention of pre-modern states. One of the oldest and most prominent examples is the Ordinance of Villers-Cotterêts (15.VIII.1539), which was part of a series of similar but less extreme and explicit edicts. This law established the monopoly of the French language within the judicial system of the kingdom of France, at the expense of Latin (which was considered to have become incomprehensible) and regional languages (Occitan in particular), thereby promoting a clearly nationalist idea of a ‘French mother tongue’:

[...] Et pour ce que telles choses sont souvent advenues sur l'intelligence des mots latins contenus esdits arrests, nous voulons d'oresnavant que tous arrests, ensemble toutes autres procédures ... soient prononcés, enregistrés et delivrés aux parties en langage maternel françois et non autrement [...]⁴⁴

With the edict of Moulins, issued in 1490 for Languedoc, the tribunals maintained the right to use the *langage françois ou maternel*; this allowed for the use of Occitan, in accordance with the linguistic competence of the inhabitants of the territory belonging to the royal domain for whom French was a foreign language.

Parallel to the development of the state, the importance of language planning has continued to increase right up to the present day, and it continues to play a role in numerous situations (*cf.* 10.6.3).

⁴⁴ “And since such things have so often happened due to the (improper) understanding of Latin words in written decrees, we wish that henceforth all decrees, as well as all other proceedings be pronounced, recorded and delivered to the parties in the French mother tongue and in no other form [...]”.

The most notable influence exerted by the state on language during the modern period was that of the language academies, including the *Accademia della Crusca* in Italy (1583), the *Académie française* (1635), the *Real Academia española* (1713) and the *Academia das Ciências de Lisboa* (1779). These institutions were established either within the context of a private cultivated society, as in Spain and Italy, or by royal decree, as in France (under the patronage of Richelieu) or in Portugal (by Queen Maria I). In Romania, the *Academia română* was not founded until the 19th century (1866), after a period of intense linguistic thought which had its origin as far back as 1780.

Nevertheless, the aims of the academies did not differ fundamentally: they included the compilation of a dictionary, a grammar and, potentially, works on orthography and rhetoric. In all countries, however, the focus was on the dictionary project, which was realised according to different principles and at varying speed:

- the *Vocabolario della Crusca*, consisting of a single volume, appeared in 1612, a mere thirty years after the founding of the Academy. It was based on quotations from respected authors, particularly Tuscan authors of the 14th century, including the *Tre corone* Dante, Boccaccio and Petrarch. It remained the main reference for all Italian lexicography until the beginning of the 19th century, even though its conservative and archaic character was criticised from the time of its publication;
- the *Diccionario de Autoridades* in three volumes was also compiled within thirty-four years (between 1726 and 1739) and it is based on quotations from 15th and 16th century works; for the understanding of Spanish vocabulary from this period it is a masterpiece that is still useful today;
- the publication of the *Dictionnaire de l'Académie française*, begun in 1634, was delayed after the death of Vaugelas (1650), which led to the parallel publication of the dictionaries by Richelet (1680) and Furetière (1690); the two-volume work, eventually published in 1694, reflects the literary vocabulary of the first half of the 17th century with great accuracy, but – like all later editions – it does not provide quotations from the works of established authors (cf. no. 3 above);
- the Portuguese Academy published an extensive volume during the first part of the 18th century, which, however, only went as far as the letter A-; the first complete dictionary to be published by this institution did not appear until 2001 (*Dicionário da Língua Portuguesa Contemporânea*).

Academies were much slower when it came to producing grammars, almost certainly due to the fact that it is initially easier to compile vocabularies than to provide a coherent representation of a morphosyntactic system:

- the *Gramática de la lengua española* of the *Real Academia española* appeared in 1928 (cf. 8.5.1) and remained a useful reference work for 20th-century Spanish until recently;

- the French grammar of the *Académie française*, published in 1932, was, in contrast, significantly inferior with regard to the standards developed at the same time.

Language academies themselves had little influence on language use after the 18th century. However, they played a significant part in the development of the specific traditions of grammatical analysis and lexicography that were intensified and promoted through the prestige of the state.

5 Elite and population

The elaboration of standard languages also relied on the activities of the socio-cultural elite. It was achieved by an intellectual and erudite minority, and the new language forms that it brought forth were destined for a world that revolved around the royal court and urban patriciate. The use of these newly forged languages was linked to situations of linguistic distance and often remained restricted to writing. This was especially pronounced in Italy, where Pietro Bembo himself, the chief architect of a written language based on (archaic) Tuscan, only spoke Venetian. As a parallel from Germany, nearly three centuries later, Goethe typically spoke only his native Frankfurt dialect, except, for instance, when reading his works aloud in public.

In the remainder of the *Romània*, in contrast, such discrepancies did not arise. In France, the elite had become accustomed to using the semi-artificial standard variety in speech from the 16th century onwards.

Cf. the stance taken by Vaugelas: ‘Pour moy j’ay creu jusqu’icy que dans la vie civile, et dans le commerce ordinaire du monde, il n’estoit pas permis aux honnestes gens de parler jamais autrement [que] dans le bon Usage.’ (Literally “As for me, I have always believed that in civic life, and in wordly affairs, honest people were not permitted to speak in any other way than according to good usage.” *Remarques sur la langue française*, Preface, VII.2., after the facsimile accessible on the website *Gallica*; *cf.* also the culture of spoken language in the French salons).

In Spain and Portugal, the written standard must have been close to a prestigious and probably quite widespread spoken variety since the Middle Ages. The new standard Romance languages began to establish themselves as acrolectal varieties in the different countries of the *Romània*, at different periods. The vast majority of the population was not directly affected by language standardisation until the 19th century: intimately linked to writing and its culture, this elitist phenomenon excluded the 95% of the population who were illiterate (*cf.* 10.4.3 no. 2; 10.5.1)⁴⁵. The latter were probably capable of understanding utterances formulated in a neutralised language – no doubt better than

⁴⁵ In 16th-century France, 40,000 out of 20 million subjects (i.e. 2% of the population) were literate; in the 17th century, it is likely that one in 30 million (3.3%) read and spoke the standard language (*cf.* Gerner, *Education et histoire des langues: la Gallo-Romania*, RSG 2, art. 107).

the population of the 13th century had understood Latin – but they were incapable of reproducing it, and written production, understandably, did not reach them at all.

Nevertheless, the elaboration of standard languages or varieties in the 16th and 17th centuries maintains its full importance for the history of language and society. Though initially a phenomenon of limited social reach, the new language forms would gradually assert themselves. At different times throughout the *România*, literacy campaigns – such as that in Normandy during the 18th century – increased the number of standard language speakers; in addition, compulsory or general schooling programmes created a broad basis for these varieties that had by then become ‘national’ languages from the second half of the 19th century onwards. The process essentially achieved its goal after the Second World War, at least in the European *România* (though 20% of the population of a country such as Portugal was still illiterate in 1980). All in all, the influence of the state on standard languages was to have the most noticeable impact through its ‘politics of literacy’ – to a much greater degree than through language planning or the founding of academies.

6 Acrolectal spoken varieties and writing

The relationship between speech and writing is one of the factors that is most difficult to comprehend in the formation of standard languages. It is generally assumed that the 16th century was a pivotal era for the elaboration of standard Romance languages in their written form, and that an array of noteworthy internal changes took place at this time: the evolution from Middle French to Modern French as well as from Old Spanish to Modern Spanish occurred approximately between 1480 and 1630, and that from Old Italian to Modern Italian, between 1500 and 1560 (*cf.* 5.2.1). Texts dating from after this period are considerably less difficult for present-day readers to understand than earlier texts.

This chronology nevertheless only accounts for written evidence of the languages in question, and does not attempt to interpret the status of such evidence within the diasystem of its time. In reality, it is necessary to distinguish between different situations, which is an almost impossible task for the modern observer:

- 16th-century innovations that first appeared in writing, entering spoken language shortly afterwards;
- other innovations specific to writing that remained foreign to spoken language for longer (or that remained permanently restricted to writing);
- in parallel, 16th-century innovations in spoken language, which were catalysed by newly arisen contexts of communication and which later found their way into writing;
- changes in spoken language that predate the 16th century but that did not appear in writing until after this period.

These distinctions may appear theoretical, but they are useful in that they offer a clear picture of the difficulties involved in situating language changes in their precise sociological and communicative context, even within the single domain of acrolectal varieties (cf. 10.5.4 for non-standard varieties).

7 The standard Romance languages

The transformation of medieval written languages into modern standard languages occurred at an early stage in the case of the four languages French, Italian, Spanish and Portuguese. Nevertheless, the other Romance languages that had developed beyond the stage of a local dialect at some point in their evolution also underwent significant change during the modern period. Two opposing processes should be distinguished: partial elaboration, comparable with that undergone by the four national languages cited, or, conversely, a loss of prestige leading to the abandonment of the elaboration process that had been initiated in the medieval period.

The first scenario concerns, above all, Romanian, Romansh and Ladin, taking place in the 16th century, at a time when other Romance languages could already look back on a written tradition of several centuries. It may thus be more aptly described as a quest for a suitable written form rather than an actual process of standardisation. In the case of Romansh and Ladin in particular, the intellectuals who undertook Bible translations and who wrote the first poetry were inspired by the discussions on language which took place during the Italian or French Renaissance.

Romanian, which had always been geographically separated from the rest of the *România*, was a special case. Here, the model of Old Slavonic played an important role and Romanian consequently adopted an expanded Cyrillic alphabet during this period. After tentative beginnings in the first half of the 16th century, a considerable written tradition developed in this language between the mid-16th and mid-19th centuries, which included novels, short stories, poetry, legal and administrative documents as well as scientific and technical texts, representing the typical range of modern textual genres, albeit to a lesser extent compared with France or Italy, but in harmony with cultural evolution in Europe as a whole.

Interestingly, Romanian at that time displayed fewer similarities with other Romance languages and had more in common with the Southern Slavonic languages than it does today. The elaboration of modern Romanian towards the end of the 19th century, which relied on the model of French (or Latin) and introduced the Latin alphabet, would bring this language closer to western Romance ('re-Romanisation'), a process that was fraught with problems arising from the adaptation of Latin characters to the phonological system of Romanian (cf. Dahmen, *Externe Sprachgeschichte des Rumänischen*, RSG 1, art. 66; for Romansh, cf. Darms, *Sprachplanung, Sprachlenkung und institutionalisierte Sprachpflege: Bündnerromanisch*, RSG 2, art. 126c).

Different scenarios characterise Occitan, Catalan and Galician, all three of which had a rich written tradition during the medieval period. Like written varieties of Italy and the *oïl* territories, in the 15th century these languages displayed tendencies toward homogenisation typical of the regional varieties of the late Middle Ages; in short, the stage was set for the development of cross-regional standard varieties. For reasons of political hegemony, however, the process was prematurely interrupted: the whole of the *oc* territory came under French rule from 1453 to 1486; Catalonia passed to the Spanish kingdom in 1516, and the whole of Galicia was annexed in 1640 (*cf.* 10.5.1). These three territories thus experienced the consequences of the elaboration of a ‘national’ language variety that was imposed as a language of prestige. Written production in Occitan, Catalan and Galician declined as a result and the three languages became marginalised.

This had fatal consequences for Occitan and Gascon, despite some attempts at revival during the modern period (*cf.* 10.6.3). For Catalan and Galician, in contrast, important grammatical and lexicographical works were developed, especially from the 18th and 19th centuries onwards. This paved the way for the elaboration of these two languages in the 20th and 21st centuries.

Other Romance varieties with marked internal characteristics, such as Franco-provençal, Sardinian or Friulian, retained their importance as spoken languages, and in some cases attempts were made to put them into writing. Francoprovençal is still used in clearly defined contexts (e.g. in farces), which lends it a specific identity. Similar uses of Gallo- and Italo-Romance dialects and Asturian are found. Between the 16th and 19th centuries, all these varieties were thus able to maintain their status as peripheral, yet living, languages.

10.5.4 Non-standard varieties

Throughout the whole of the modern period, non-standard Romance varieties remained the dominant form of spoken communication in their respective regions. Even within the territories covered by the four large national languages, the mother tongue of almost the entire population consisted of other, non-standard varieties – dialects or, at least, sociolectally marked varieties. Even the young Louis XIII spontaneously spoke a type of familiar or popular French, as can be ascertained from the accurate transcriptions made by his physician Hérouard (*cf.* Ernst, *Gesprochenes Französisch zu Beginn des 17. Jahrhunderts*, 1985).

For Romance languages such as Occitan or Galician, and even more so for Native American languages or creoles, discrepancies between the prestige variety and these mother tongues sometimes led to total mutual incomprehensibility among some speakers, thus facilitating the abandonment (or, in some cases, the survival) of non-standard forms.

In the north of modern-day France, central and southern Spain and Portugal, standard varieties must still have been relatively comprehensible in speech, even for

the rural population. Divergences between the standard and mother tongue varieties nevertheless held some surprises. Take, for example, the invaluable record of an English traveller, named Glass, who claims to have encountered a language close to standard Spanish in Las Palmas, the capital of Gran Canaria, in the 18th century. In the inner part of the island, on the other hand, he was confronted with a language that seemed nearly incomprehensible to him, although it, too, was a variety of Spanish (cf. Glessgen, *Der Beitrag der kanarischen Ostinseln zur hispano-amerikanischen Sprachentwicklung*, 1997b). A truly generalised situation of diglossia must therefore have existed, obscured by written sources, which tend strongly towards the reduction of linguistic diversity.

It is even probable that divergences such as these arose as a result of the formation of standard varieties: since the 16th century, and in ever-decreasing competition with Latin, standard languages fulfilled all the functions of long-distance communication. In the Middle Ages, these functions had been assumed not only by Latin, but also by regional varieties, which had attained a lesser degree of elaboration and were thus closer to the respective mother tongues on which they were based. The gap between spontaneous (familiar) and elaborated (distance) forms of expression was therefore reduced, and it is likely that the two extremes of the continuum were relatively close.

In contrast, the abandonment of regional languages immediately accentuated the distance between the prestigious, cross-regional form, and the colloquial, local spoken varieties. The use of mother tongue varieties was restricted to situations of communicative immediacy and excluded from the sphere of linguistic distance. This process contributed to the progressive geographical restriction of the use of dialectal varieties and, thereafter, to a growing internal differentiation between dialects.

Similarly, in situations where diastratic varieties of low social prestige were in opposition to a high-prestige variety that was becoming increasingly distant and dominant, the former developed increasingly marked linguistic features. Thus, it was in all likelihood the elaboration of standard varieties that provoked the intensification of the diastratic dimension of the Romance languages in the 15th and 16th centuries, which had been little developed up to that point. It can be hypothesised that the internal differentiation of primary Romance dialects reached its peak towards the middle or the end of the 19th century and that marked non-standard varieties flourished from the 17th century onwards in particular.

Diatopic and diastratic varieties left only sporadic traces in writing, though texts of the period display diaphasic variation to some extent. Differences between the textual genres that had developed after the medieval period were reinforced by language elaboration in the 16th century. There was an increasing specialisation of knowledge and a growing emphasis on stylistic factors. Moreover, during this period, an early form of 'folk literature' emerged, which made use of dialectal varieties and imitated spoken language in writing (cf. 10.5.3 no. 7). The former continuum between exclusively spoken

dialects and elaborated regional varieties, whereby dialects had also become adapted to communication at a 'supralocal' level, was permanently broken.

Finally, the spread of literacy led to a phenomenon unknown in the Middle Ages: the emergence of texts whose authors had little knowledge of written culture. As we have seen (cf. 10.5.3 no. 2), the first of these were Spanish chronicle writers with minimal education who had travelled to the Americas and who had written of their everyday experiences for a European public eager for exotic tales; they were followed from the 17th and especially the 18th century onwards by lesser educated persons in Europe who left autobiographical texts in non-standard varieties. Thus, paradoxically, during the 'great age of normalisation', a new dimension of language appeared in writing which contrasted with the newly established norms. A wide array of heterogeneous factors contributed to the development of a diversified culture of writing.

10.6 The present-day period (1880 until today)

Owing to the abundance of written and oral sources available for the present-day era, attempts to describe this period encounter even greater problems than those associated with the modern age. Soaring population growth in the *România* (which today includes some 700 million speakers), the spread of literacy and the inclusion of oral language within the scope of linguistic research have caused the potential fields of observation to be multiplied a hundredfold or even a thousandfold. The observations made here will be limited to a synthetic and interpretative (and thus necessarily simplified) treatment of this period with regard to its position in present-day research. The present-day period is also treated in the context of the current geographical distribution of the Romance languages, variational linguistics (cf. ch. 3 and 4) and the history of Romance studies (cf. ch. 2).

→ *Histoire de la langue française* by G  rald Antoine *et al.* (3 vols., 1985–2000) and its precursor by De Mauro, *Storia linguistica dell'Italia unita* (1963)

10.6.1 Geographical space, geopolitical factors and varieties

The present-day period has witnessed the consolidation of the national boundaries of Romance-speaking countries in Europe, America and Africa. These frontiers have remained relatively unaltered, despite two World Wars, Fascist and totalitarian regimes in Europe and the various civil wars and dictatorial regimes that arose in Romance-speaking countries outside Europe. The triad 'national power, national territory and national language' became established as a geopolitical and geolinguistic principle.

Within the newly formed nations, interactions among the various communicative spaces of individual national languages formed a fine-meshed network. All vectors of the modern nation – such as increasingly centralised administration, compulsory schooling and military service – contributed to this development. Only from the last decades of the 20th century onwards did ‘centrifugal’ movements begin to challenge this reality.

Beginning in Europe, the 19th century witnessed a significant increase in the number of speakers of standard national languages. In the 20th century, these varieties asserted their presence within all communicative spaces: from the end of the First or Second World War (varying from country to country), standard varieties became the usual languages of communication, including oral communication, for a large percentage of the population. In the third quarter of the 20th century, the great majority of the population learned these languages in speech and in writing. In many dominant socio-cultural groups, standard varieties ultimately acquired the status of mother tongue languages, or at least second languages learned during early childhood.

Today’s standard languages only compete with other (colloquial or popular) varieties belonging to the same internal language continuum. Older varieties such as primary dialects or minority languages (e.g. Occitan, Greek or Friulian), which exhibit considerable internal divergence, have decreased in importance or have even disappeared. Only in Italy is there still a large number of dialect speakers, as a result of the highly developed regionalisation of the country, which began in the pre-modern period and is still continuing today. Everywhere else, the last mother tongue speakers of ‘natural’ dialects have reached retirement age: the primary dialects of French are on the verge of extinction, as are those spoken in the northern regions of the Iberian Peninsula (*cf.* 3.4).

Despite this general tendency, some of the old regional Romance languages have managed to maintain their vitality owing to political support. Naturally, centrifugal tendencies such as these do not favour dialects of the old, primary type, but rather new varieties in the process of standardisation:

- among the ‘distance languages’ which emerged during the modern period, Galician and Catalan have acquired the characteristics of elaborated languages, as well as a growing number of mother tongue speakers, supported by the political autonomy of the regions in which they are spoken;
- the other distance languages of the *România* have not benefitted from a comparable degree of elaboration, and have experienced different fates: Sardinian has maintained its vitality to some extent, while that of Francoprovençal and Occitan has gradually weakened; Dalmatian was probably already extinct by the end of the 19th century. Friulian, Ladin and Romansh occupy an intermediate position, having undergone elaboration during the 20th century, supported by language politics in Italy and (to a greater degree) in Switzerland; however, the number of native speakers is decreasing (except for speakers of Ladin in recent years);

- some varieties that are characterised by less significant internal distance with regard to the standard or neighbouring varieties have also received some political support and are today recognised as varieties of regional identity. This category includes Corsican, Asturian, Aragonese and Aranese; however, the number of native speakers of these varieties among younger generations is stagnating.

Geolinguistic changes in the *Romània nova* show yet further tendencies that run contrary to extreme language standardisation and unification: national languages tend increasingly towards pluricentric codification, which is accompanied by growing divergence between the individual national standards (examples within Europe include standard French in France compared with French in Switzerland or Belgium, Romanian in Romania compared with Romanian in Moldova, or German in Germany compared with German in Austria or Switzerland).

This phenomenon is the most pronounced in the extra-European *hispanidad*, the whole of which was politically detached from Spain by 1989 at the latest. The Spanish elite claims cultural and historical superiority over its former colonies, above all through a number of pan-Hispanic institutions. These claims are in part accepted (Peninsular Spanish still constitutes a ‘super-norm’ in the Americas), and in part rejected or negated by the concrete situations of today’s world. In nineteen countries of the Americas, Spanish is both the mother tongue of a large percentage of the population and the dominant language. Depending on the country and region, however, Spanish is also involved in situations marked by varying degrees of bilingualism (cf. 3.4.9).

In the United States, in contrast, Spanish is in the position of a socio-culturally dominated language and its speaker numbers are continuing to increase, be it in California, Texas, or New York. The presence of Spanish in the Philippines (where it is still barely alive) as well as in equatorial Africa (where it is disappearing), should also be acknowledged.

Standard Spanish as a written language of prestige maintains a certain unity across the entire Hispanic-speaking world. Modern means of communication allow multiple interactions among the countries concerned. This unity is nevertheless relative, since pluricentric codification (*codificación pluricéntrica*) is currently leading all Spanish-speaking countries to develop their own norms, each of which displays slight differences with regard to the other varieties of standard Spanish (cf. 4.2.4).

Familiar Spanish (*español coloquial*), and especially the ‘peripheral’ varieties of the language, display pronounced centrifugal tendencies characterised by an impressive number of lexical neologisms as well as their own specific phonetic, morphological, syntactic and onomastic innovations. Great geographical distance, the demographic dynamics of the metropolitan areas involved, as well as the extent of illiteracy and bilingualism have been the drivers of these ‘centrifugal’ tendencies.

The *lusitanidade* primarily encompasses Brazil, Angola and Mozambique. In East Asia, in contrast, Portuguese has almost completely disappeared; only a few Portuguese-

based creoles remain as witnesses to the colonial past of the language there. In Brazil in particular, the sociolinguistic situation and the internal dynamics of Portuguese resemble those described for Spanish in the Americas. The percentage of speakers is also comparable: there are between 340 and 360 million speakers of Spanish in Hispanic America as opposed to 42 million on the Peninsula; Portuguese speakers in Brazil number approximately 180 million, compared to 10 million in Portugal.

Despite its prominent role in French political discourse, the extent of the area known as '*la Francophonie*' is more limited, both in terms of geographical space and demographics. The term, which was coined in 1880 by Onésime Reclus to designate a geographical concept, acquired the status of a political emblem (largely due to the efforts of Leopold Senghor) in 1962, aimed at affirming the identity of the various French-speaking countries after their decolonisation. In 1970, the *Agence de Coopération Culturelle et Technique* (ACCT) was founded on the initiative of the president of Niger, Hamani Diori; in 1997, it became the *Agence Intergouvernementale de la Francophonie* (AIF), an organisation responsible for promoting the values of the *Francophonie*.

As is often the case, language is secondary to the specific political issues for which it serves as a pretext. In contrast to the vast geographical area covered by the Iberian languages, there are very few territories in which French is the native language of a large percentage of the population: these consist of Quebec (which gained independence in 1867), and the overseas departments and territories of France (DOM-TOM; cf. 3.4.1). Nevertheless, French also displays pluricentric codification. Outside Europe, French normally exists in a situation of intense language contact with English, creole languages or indigenous languages.

In all other contexts, French is a second language, used by varying numbers of speakers as a *lingua franca* or a prestige language, as in sub-Saharan Africa and the Maghreb – at one time it also held this status in Indo-China, beginning later than in the other countries mentioned and only for a limited period. Schooling and contact with emigrants ensure some degree of continuity with regard to language use, although a number of specific internal developments can be observed which affect lexis as well as morpho-syntax. Nevertheless, the future of French in these contexts is far from certain.

A discussion of the countless examples of language contact in the 20th and 21st centuries is beyond the scope of this section. In general, however, it may be said that major migratory movements are far from slowing down: one need only consider Hispanic immigration to the United States (with approximately forty million *hispanohablantes*), Arab immigration to various European countries (with an Arabic-speaking population possibly amounting to two million in France) or immigration from Asia to many countries of the Americas.

Among situations of language contact that do not involve migration, both the British and the American models play an important role in the *România*, as elsewhere. The various Romance languages often borrow the same terms from English, which thereby

transcend the boundaries of individual countries. Such Anglicisms contribute to a form of linguistic homogenisation somewhat comparable to that of Latin in previous centuries, English having replaced the latter as the language of international communication (cf. 9.6.2).

The geolinguistic configuration of the *România* today is thus characterised by two opposing tendencies: pluricentric codification (which is ultimately a centripetal process) on the one hand, and on the other, the numerous centrifugal tendencies induced by language contact as well as by the coexistence of multiple socio-cultural contexts.

10.6.2 Demographic, infrastructural and socio-cultural factors

The development of civilisation as we know it today has been influenced by multiple factors: industrialisation, scientific progress, geographical and social mobility, demographic growth on a global scale, the spread of literacy, the emancipation of women in Western culture and (at least partial) democratisation. A parameter that intersects with many others while simultaneously determining them is urbanisation, emblematic of and, at the same time, a reality inherent in modernity: Europe and North America are almost completely urbanised, with the phenomenon even affecting life in the countryside. This has led to the establishment of a permissive society, in some cases following dramatic periods dominated by Fascism, and has introduced new forms of mobility as well as the freedom of lifestyle choices into the Western world.

The second half of the 20th century appears to have seen the transition from a fundamentally rural world that followed a rhythm defined by the whims of nature (uncontrolled by man) to an urban world, with all its wealth but at the same time possessing extreme destructive power. This transformation was accelerated in the 1960s by the memorable events of the Second Vatican Council (which redefined the attitude of the Church towards modern society) and the events of May 1968, which consolidated the consequences of the Second World War.

These events constitute a veritable turning point in history – possibly the most significant since the invention of agriculture and man's adoption of a sedentary lifestyle some 8000 years ago. In a perspective oriented more towards language history, it is certain that persons brought up in the Western world since 1960 inhabit a communicative space that differs fundamentally from the world in its previous state. It is also clear that present-day *România* must face up to the contrasts between societies in Europe and North America and those in Latin America and Africa, which, although still in a critical stage of their development, possess creative potential. During the last quarter of the 20th century, French literature from sub-Saharan Africa, the Maghreb or the Antilles, Spanish literature from Latin America and Brazilian literature outshone European literature written in these languages.

Once again, writing is inextricably linked to these processes: the discarding of Latin and the rise of the Romance national languages, associated with the general spread of literacy, have led to a great increase in the number of potential players in the political and socio-cultural spheres. In recent decades, education has been extended to third level for a significant percentage of the population, while longer life expectancy has furthered longer-term processes of learning.

10.6.3 Written culture and linguistics

The most important break separating the present-day era from previous periods concerns written culture; the former is marked by widespread literacy, while during the latter, writing was a privilege reserved for the social elite. Textual genres increased and diversified in all imaginable areas of writing, including handwritten texts and texts of a private character. The media, powerful and omnipresent, have transformed contexts of communication: newspapers, the radio, cinema and television, as well as, more recently, the Internet, all play a predominant role in language evolution today. Of these, newspapers appear to be the most reliable upholder of the written standard.

Consequently, diaphasic variation, which is linked to the different discourse traditions, has increased: the development of the natural sciences and the humanities at the beginning of the 20th century led to the elaboration of technical language, which, in turn, provided new linguistic registers used by an ever-increasing number of speakers. Later, in the second half of the 20th century, other sectorial languages developed, which were associated in particular with the mass media, such as the language of sports or advertising.

The diversification of textual genres in the modern period has been amply covered by literary theory. Literary texts were exploited to the full as a testing ground for the expression of social ideologies and aesthetic values, until the avant-garde and more recent literary trends brought about the dissolution of the traditional concept of the text. Consequently, a systematic and coherent definition of textual genres and their characteristics is difficult, if not impossible. Still, the persistence of early literary traditions in the *România* ensured that 20th-century literature was constructed on the basis of – and through the restructuring of – the principles of previously established Romance genres.

It is important to note that literary texts make use of a wide range of linguistic varieties: dialectal literature has existed since the 16th century and constitutes a rich source for non- or partially-standardised languages in particular; as an example of the use of regional varieties, we have seen that the production of novels or even poetry in the *România nova* in the second half of the 20th century held its own against literature written on the European continent; diastratic and diaphasic variation generally feature heavily in the novels of many recent authors (e.g. Queneau; cf. the detailed overview in RSG 2, 172–178).

Language culture, language politics and linguistic thought are offshoots of written culture that are particularly important for language change, since they are situated at the interface between language and society. While language elaboration played a role in the shaping of pre-modern states, the 19th and 20th centuries witnessed a significant increase in the interactions between language and the state.

Language planning is a particularly representative example of a phenomenon that has undergone unprecedented development since the beginning of the present-day era. First of all, a state defines a national language, which thus becomes dominant with regard to all other varieties in use throughout the territory. In a second stage, the state also grants recognition to the languages spoken in specific regions displaying a degree of political autonomy – the consequences of this can be just as far-reaching. It is also the state that regulates the use of language in education by means of literacy campaigns, and ultimately makes decisions on orthographic reforms.

France constitutes the perfect example in this respect, beginning with the efforts of the Abbé Grégoire, who, at the dawn of the post-revolutionary regimes, undertook surveys investigating the dialects spoken in France, thus paving the way for their elimination. The *Loi Guizot* (1833) specified that compulsory schooling for children in all municipalities was to take place in French and in no other local or regional language; this decision achieved its full significance with the introduction of free and compulsory schooling by Jules Ferry in 1882 (cf. Gerner, *Éducation et histoire des langues: Gallo-Romania*, RSG 2, art. 107).

Half a century later and in a far less incisive manner, the *Loi Toubon* contested the use of English and imposed the obligatory use of French in all contexts, i.e. ‘dans la désignation, l’offre, la présentation, la description de l’étendue et des conditions de garantie d’un bien, d’un produit ou d’un service, ainsi que dans les factures et quittances [...]’ (‘when describing, offering or presenting goods or services for sale, in describing their use, defining their guarantee, as in bills and receipts [...]’, Article 2 of the law of August 4, 1994).

The same phenomenon can be observed in Italy, Spain, Portugal and Romania, where language use and, in part, the concrete forms of standard varieties, were subject to political decisions (cf. Lebsanft, *Spanische Sprachkultur*, 1997). Naturally, the influence of the state is not as immediate as some of its representatives would wish: the futile attempts to quash Anglicisms initiated by certain persons in France illustrate the limits of the decisions taken by authorities. Nowadays, the state nevertheless plays a paramount role in the acceptance of dominant linguistic norms as well as alternative norms by the population.

The development of universities from the 19th century onwards represents another state intervention that had far-reaching consequences on language use. The impact of universities is many-faceted, as may be inferred from the development of linguistics alone. Over the course of the 19th and 20th centuries, the shaping of this new discipline

was nurtured by concerns arising in the context of school education as well as by more scientific considerations.

Specifically, the various offshoots of linguistics play an active part in the generalisation of standard languages as national languages, in particular synchronic lexicography and grammatical description. These, in turn, draw from historical lexicography and grammar, which, together with the historiography of language *per se* ('the history of national languages') and the edition of old texts, contribute to the shaping of a national awareness, founded on the notion of linguistic unity (the '*unidad del idioma*'). The sub-discipline of dialectology deals with the most complex sociolinguistic realities related to the acquisition of national standard languages (which has become a necessity in today's society).

These different vectors of linguistics are thus both reflections of the spread of modern written culture and participants in the same process. Their interaction, which is often imperceptible but which nonetheless maintains an underlying omnipresence, has continued to intensify throughout the 20th century.

Chapter Two on the development of Romance linguistics therefore can and should be read as a section in its own right within the external history of the Romance languages. Linguistics as an academic discipline unquestionably contributes to the processes of language change (both external and internal) that may be observed during the present-day period, even acting as one of its principal catalysts, and it plays a constitutive role in modern society, alongside the other humanities. These developments are part of a long process, in which the spread of literacy and the 'intellectualisation' of society go hand in hand with an intensification of linguistic thought and culture. The two mutually nurture and condition one another.

The establishment of standard languages can be better explained from this perspective: the idea of a modern nation, or even a republic, is closely associated with that of the combined presence of a territory, a language – and a history (cf. 10.6.1). Even though history bears witness to anything but unity, and even though linguistic unity had not yet been achieved in the 19th century, political imagination gains the upper hand regarding this issue. In the context of national historiography, the past is interpreted through the lens of the present. Following this vein of thought, the language of medieval Romance texts is often intentionally normalised by editors so that it more closely resembles the modern standard (cf. 11.5.1). Linguistic unity is thus constructed in retrospect, modifying reality to suit the image.

The large literate population and the development of linguistics as a university discipline in the 20th century gave rise to a new form of reflection on language, termed 'folk-linguistics' (German *Laienlinguistik*). This facet of today's linguistic culture appears in popular literature and newspaper articles on language (such as the *chroniques de langue*, which appeared at an early date in France). For Spanish, there have also been lively discussions on linguistic norms in readers' letters in important newspapers (cf. Lebsanft, *Spanien und seine Sprachen in den Cartas al Director von El País*, 1990). Today,

such discussions have reached the Internet and crossed the Atlantic (*cf. id.*, “*Lingüística popular*” y *cultivo del idioma en Internet*, 2001).

Representatives of political power, linguists and ordinary members of the population all participated in the new wave of standardisation which occurred during the 20th century. Catalan has had its own lexicographical and grammatical reference works as well as its own tradition of historical linguistics since the early 1900s (Fabra, *Gramàtica de la llengua Catalana*, 1912; *Diccionari general de la llengua catalana*, 1932). Thanks to the efforts of Romance linguist Heinrich Schmid, Romansh has its own neutralised variety, which serves as a standard (*‘Rumantsch Grischun’*, 1932). The elaboration of Galician was brought about through translations and the development of an academic branch of linguistics, mainly in the last three decades of the 20th century.

These mechanisms are similar to those which drove the less fruitful attempts at standardisation, particularly that of Occitan, undertaken first by the (Rhodanian) Provençal movement named *Félibrige* (under the impetus of Frédéric Mistral, winner of the Nobel Prize in literature and author of a large descriptive dictionary of Provençal, *cf.* 3.4.2; 9.9.2), and then by an opposing Languedocian movement, the *Institut d’Estudis Occitans* (IEO), which established a standard orthography for Occitan. Due to the low number of native speakers of Occitan varieties, and as a result of language politics in France, which forged an inextricable link between social mobility and personal commitment to the single national language, these attempts were doomed to failure, despite Occitan’s glorious literary past and the territorial and economic importance of the southern provinces.

10.6.4 Evolutionary trends and present-day non-standard varieties

The popular varieties of national languages have witnessed unprecedented growth, radiating from two partially overlapping epicentres: youth language and urban language. The prolonging of the period of transition from adolescence to adulthood over the course of the 20th and the beginning of the 21st centuries, as well as the socio-cultural, if not economic, independence acquired by groups whose self-professed identity is associated with ‘youth’, also drives the linguistic phenomenon of the creation of identity-building language. At the same time, growing urbanisation has led to the development of disadvantaged urban areas throughout the *România*.

These two types of language variation typically display a great deal of internal creativity (particularly as regards vocabulary) and a fast rate of innovation, which is further accelerated by the extensive mediatisation of today’s society: neologisms are for the most part short-lived. Different aspects of these two peripheral spheres, as well as factors relating to electronic and social media, are, moreover, inextricably linked: urban language, more prone to evolution than more formal varieties and catalysed by a low degree of literacy, is most commonly used among the young (from disadvantaged social groups) and is partially imitated by middle-class youth.

From a language-internal perspective, the numerous phenomena of variation – diatopic, diastratic and diaphasic – operate within the framework of national languages. Standard languages and their representatives and vectors face intense variation within social or regional groups, reflecting different identity crises within nations and harbouring considerable potential for social tension. The current period is marked by language conflict – the immediate translation of and at the same time the catalyst for underlying social differences. Here, the power of language in the creation of identity unfolds to its fullest, alternatively favouring social integration and disintegration.

Linguistics assumes a new function in this context: that of identifying and describing the centrifugal forces at play within the architecture of a language, in order to discover the underlying contradictions they reflect, with the aim of defusing them.

10.7 Contributions of external history

The extralinguistic factors that determine language use help us to identify internal linguistic phenomena and to interpret changes – both in spoken language and in written culture. The presentation and hierarchical organisation of the different external factors that are pertinent for internal changes, together with the data presented in the chapters on internal history, provide an overall account of the history of the *Romània*. A combined perspective such as this allows an understanding, based on solid foundations, as to where and how, in a given period, Latin experienced variation to such a degree that it led to the formation of new, geographically differentiated Romance languages, as well as to how these languages evolved and were elaborated.

This perspective thus allows us to understand the process by which new diasystems are constituted: between the 8th and 11th centuries, in the absence of a written Romance language, only variation across geographical distance was able to develop within the Romance diasystems. Socio-cultural variation developed later, in parallel with the evolution of written varieties of Romance throughout the entire late medieval period (*cf.* the following chapter for more on this subject). Since World War I, the diffusion of standard languages throughout society has finally led to the development of pronounced diaphasic variation. Even though their influence cannot be determined in detail, it is clear that these macroscopic transformations had a considerable impact on internal change within the different Romance varieties.

A line of argumentation based on the interpretation of external history – reinforced by philological evidence (*cf.* 11) – is capable of explaining the marked differences in the degree of standardisation and the diasystems of specific languages: the internal physiognomy of French, which is as elaborated as it is ‘exotic’ within the *Romània* from a typological point of view, is determined by the high degree of centralisation, beginning with the late medieval royal court and intensified by the state under the *Ancien Régime*. The elaboration of French, in turn, explains the substantial differences that exist between its oral and its written code, as well as the strong presence of dias-

tratic variation characterising both, which is more highly developed than in Italian or Spanish.

In contrast, the premature standardisation of Spanish in the 13th and 14th centuries did not continue during the modern and present-day eras. The modern language has thus attained a less advanced stage of standardisation, but shows greater consistency between oral and written codes. Italian has witnessed further intense linguistic reflection since the modern period; the process of standardisation has been slowed down by the absence of a central and centralising state. On the other hand, the importance of diatopic variation in the history of Italian has resulted in its persistence.

Portuguese and Romanian witnessed a lesser degree of standardisation, due to the relative weakness of the tradition of linguistic thought on Portuguese and the limited nature of the written tradition in Romanian. Note that the absence of primary dialects in both Romanian and the Ibero-Romance languages (with the exception of varieties spoken along the northern fringe of the Iberian Peninsula) can be explained by the major migrations that occurred during the Middle Ages.

Finally, Catalan and Galician, which exhibit significant morphological variation, display the general characteristics of recently standardised languages. This is even more pronounced in Romansh, Ladin, Sardinian, Occitan or Gascon, not to mention Francoprovençal or Friulian.

In a given linguistic diasystem, the effects produced by external factors combine in various different and rather unpredictable ways: language history, like historiography in general, always inevitably explains developments in retrospect, without being able to provide clear predictions. Nevertheless, it allows us to gain a better understanding of given situations of language variation in a historical context and to evaluate internal contradictions and potential patterns of evolution.

Part 5: **Practical and ideological aspects of research
and teaching in Romance linguistics**

11 Philological approaches and corpus linguistics

As we have seen in the section on the history of linguistics, the formation of the modern discipline was based on the combination of a comparative and a genetic approach to languages, and on the intensive study and edition of texts from older periods, closely followed by the description and analysis of modern dialects (*cf.* 2.2.2). Early scholars such as Jacob and Wilhelm Grimm, August Wilhelm Schlegel, François Raynouard and Friedrich Diez edited medieval German and Romance texts alongside the linguistic syntheses that they elaborated. The study of the diachronic dimension of language was at that time predominant, and the only direct way to observe the older stages of a language was – and remains – the study of texts and manuscripts. This approach is anything but trivial, necessitating the development of a specific methodology that led to the elaboration of philology in a narrow sense, as a discipline dedicated to the theory, practice and criticism of textual edition (called ‘*ecdotics*’). Highly developed in Romance studies due to the great number of older texts and the excellent conditions of their transmission, this field also emphasises the nature and complexity of the relationship between raw textual data and their linguistic interpretation.

Philological practice ideally begins with an inventory of all texts and text types that have been transmitted to us (11.1). It should be noted that far more attention has been paid by linguistic research to literary texts alone than to all other textual genres together. The inventory process is complemented by the study of material aspects of the texts (11.2), and by an attempt to understand the particularities of their transmission which arise from the processes of copying and adaptation to which a given text has been subjected throughout the centuries (11.3). Only on this broad basis is it then possible to approach the question of how a text or a manuscript should be edited (11.4), and only after it has been edited can a text be studied and its linguistic implications understood (11.5). In recent times, the constitution of large corpora for historical, literary and linguistic aims has opened up fascinating new perspectives for analysis. However, the exploitation of such tools requires a depth of knowledge in the domain of information technology (IT) as well as rigorous attention to the underlying methodology (11.6).

The present section provides an overview of these complementary aspects of the discipline, which are fundamental to any diachronic approach.

11.1 Inventories of medieval Romance textual genres

The production of medieval texts was determined by textual genres which, in turn, determined the stylistic, lexical and even syntactic features of the individual texts (*cf.* 4.5). From a methodological point of view, the thorough consideration of textual genres is complementary to the two more abstract approaches – systematic and chronological – employed in the previous sections.

11.1.1 Reference catalogues offering a systematic description of medieval texts

Published around 1900, Gustav Gröber's *Grundriss der romanischen Philologie*, which reflects the sum of knowledge of the discipline of that time, already lists a large number of textual genres as sources for historical linguistics. However, the systematic study of different textual traditions and the critical evaluation of their sources were still in their beginnings. In the period that followed, philological and linguistic research focused mainly on secular literary genres, while little attention was paid to the wide variety of other discourse traditions.

Towards the end of the 20th century the canon of Romance sources was finally expanded to include a broader array of texts, also taking their Latin models into account (cf. the previous section). To date there have been several attempts to classify textual genres that exist in the *Romània* (cf. the specific articles in RSG, cited below). The categorisations focused on the Middle Ages are particularly interesting, not only because this period harbours the most complex (and thus the most intellectually stimulating) philological problems, but also because the diversity of genres is more limited and thus easier to observe (cf. *ibid.*).

The *Grundriss der romanischen Literaturen des Mittelalters* (GRLMA, 1968–, 13 vols.), founded by Hans-Robert Jauss and Erich Köhler, represents the first modern attempt to classify medieval textual genres. The goal of this substantial and reliable encyclopedia is to rewrite the section on medieval literature contained in Gröber's *Grundriss* (2 vols., 1886–1906), and it is organised according to the different textual genres. As it is based on the modern concept of 'literature', the annotated repertoire also includes religious literature and texts reflecting specialised knowledge. The GRLMA is based on a number of other repertoires concentrating on literary genres pertaining to specific periods and languages, such as Bossuat's *Manuel bibliographique de la littérature française du Moyen Âge* (1951, + 3 supplements, 1955–1986) or the *Bibliographie der Troubadours* by Pillet and Carstens (1933; now superseded by the *Bibliografia elettronica dei trovatori* / BEdT by Asperti).

With regard to documentary writing, the contents of the GRLMA are completed by the *Typologie des sources de l'Occident médiéval*, founded by Léopold Génicot (Typ-Sources, 88 fasc., 1972–2016). This historical project, which, like the GRMLA, remains incomplete, focuses in particular on Latin sources, from which a great many vernacular texts are derived.

More recently these two encyclopedic works have been completed by the above-mentioned linguistically-oriented *Inventaire systématique des premiers documents en langues romanes*, conceived by Wolfgang Raible and realised by Frank, Hartmann and Kürschner (InvSyst, 5 vols., 1997). The descriptive catalogue lists texts transmitted by original manuscripts dated before 1250. Its weaknesses with regard to details (cf. Viellard's review, CCM, 2000) do not detract from the more than adequate nature of the categories distinguished for classification, which are based on a pragmatic approach that considers the texts according to their context of use.

The earliest texts to appear in a vernacular language, which, for the most part carried strong pragmatic connotations, have already been presented (cf. 10.3.5 no. 1). Although the value of these sources for linguistic analysis is limited, their role as precursors gives them a special place in the history of language and writing. In the following four sections, we will present the major textual traditions present in the medieval *România* from approximately 1100 up to the 16th century.

11.1.2 Secular literature

Works of secular literature written in the Romance languages developed very early on. The first genres of lyric poetry appeared during the 11th to 12th centuries; these included Mozarabic poems, troubadour poetry, Gallego-Portuguese lyric, the *Scuola siciliana*, the *Dolce Stil Novo* and the first lyric texts in French, Spanish and Catalan (GRLMA, vol. 2; cf. InvSyst 4.6: “Poésie lyrique”). These texts were followed by epic poems, significant examples of which are the previously mentioned French *chansons de geste*, such as the *Chanson de Roland* and the epics of the *Cycle de Guillaume d’Orange*, GRLMA, vol. 3; InvSyst 4.1: “Chansons de gestes”, TypSources, fascicule 49, *id.*).

In the 12th and 13th centuries, romances – first in verse, then in prose – occupied a place of particular importance, with the earliest Arthurian verse romances by Chrétien de Troyes, followed by works of Arthurian prose, which, moreover, were imitated on the Iberian Peninsula (GRLMA, vol. 3; InvSyst 4.2 and 4.3; TypSources, fasc. 12; cf. also *ibid.*, fasc. 52: “sagas”). Short narrative forms also played an important role; these include the *lais* of Marie de France (GRLMA, vol. 5; InvSyst 4.4) or the works of allegorical and satirical literature, sometimes of religious inspiration (GRLMA, vol. 6/A).

Literary genres then diversified significantly in the 14th and 15th centuries, in French (GRLMA, vol. 8), as well as in Italian and other medieval Romance languages (vols. 9–10). The genres mentioned above were joined by new forms such as the rondeau, the ballad, the lament, the *dit* (a type of short verse narrative, often didactic in content, which is spoken rather than sung) and the debate (GRLMA, vol. 8; cf. TypSources, fasc. 9 “nouvelles et *novas*”, fasc. 13 “fabliaux”), political literature (GRLMA, vol. 10/B.2) as well as different forms of religious and secular drama (*ibid.*, vol. 12).

As for the modern and present-day periods, secular literature has received by far the most attention in studies on language. There are a number of reasons for this, beginning with the undeniable linguistic richness of these genres, which also represent important historical sources. TypSources devotes a significant amount of space to “Sources littéraires proprement dites” (part VII), although the majority of the planned topics have not yet been treated. Owing to the growing interest that has been shown in these texts throughout the 20th century, critical editions – often of high quality – exist for the majority of known texts, as well as a large number of studies and synthetic overviews that allow them to be understood in the light of their respective traditions. They

therefore constitute a reliable body of texts that have already been treated in depth by philologists (*cf.* 10.7) and that are consequently available for linguistic study.

11.1.3 Religious literature

Religious literature was of particular importance in medieval and pre-modern times. Bearing the strong imprint of Latin, this category of texts has much in common with secular literature, which it often precedes in terms of historical chronology. In many cases, religious texts also incorporate an aesthetic dimension; moreover, they have a rich textual tradition going back centuries, have been transmitted through numerous copies, and, in the past, they appealed to a broad audience that was not confined to literate individuals.

The most important religious text is, of course, the Bible, of which various translations into Romance languages exist, some complete, and many partial, dedicated to specific Gospels or books of the Old Testament (Kings, Judges, Maccabees), the Apocalypse, Psalms (Psalters) or canticles (*cf.* InvSyst 2.1.1; the GRLMA contains only brief chapters on 'religious literature' in vol. 6; vol. 12, on both religious and secular theatre, was never published). Other genres are paraphrases of the Bible, some of which are written in verse (*cf.* InvSyst. 2.1.2; TypSources, fasc. 64: "Bibles annotées"), and biblical Apocrypha, which enjoyed great popularity during the Middle Ages.

Apart from this key text, numerous other genres linked to cult and spirituality developed: sermons were the first genre of significance for the Romance languages, beginning with the Council of Tours in 813 (*cf.* InvSyst 2.5: one of the earliest known testimonies is represented by the 12th century *Sermoni Subalpini*, which were written in a mixed variety of Franco-Piedmontese). For a long time, this was the only part of the Divine Office that involved vernacular language, and it was not until the Second Vatican Council in 1962 that the Latin Office was abandoned (*cf.* TypSources, fasc. 56 "*ordines*", 70 "sacramentaires", 26 "martyrologes", 52 "livres de chant", 55 "prières rimées", 61 "*artes praedicandi*").

Gradually, translations and paraphrases of liturgical texts appeared (beginning with texts such as the *Pater noster* and the *Credo*, *cf.* InvSyst 2.2) as well as various paraliturgical texts (*cf.* InvSyst 2.3), including religious chants, songs recounting the Passion or the lives of saints (e.g. the Occitan *canzon* about the life of Saint Foy (Faith) of Agen), laudes (e.g. the *Laudes Creaturum*, also known as *Cantico di Frate Sole*, by Saint Francis of Assisi), heretical liturgies (such as the Cathar ritual in Occitan) and liturgical drama (for example, the Spanish *Auto de los reyes magos*).

Hagiography is a well-developed genre, very popular in Western Culture until recently. The majority of texts are written in Latin (there are, for instance, around a thousand manuscripts of the large collection of saints' lives known as the *Legenda Aurea* or "Golden Legend", which achieved great popularity in the medieval period). Nevertheless, there is also a long and rich tradition of Romance-language hagiography

(beginning with the 10th century Franco-Occitan *Life of Saint Léger*, and early versions of the *Life of Saint Alexis* or the *Voyage of Saint Brendan*; cf. InvSyst 2.4 and TypSources, fasc. 24–25 “sources hagiographiques”, 33 “vies, miracles, translation”).

Clerical texts for practical use occupied a less important position in the vernacular languages. Thus there are Old French versions of the *Rule of the Cistercians* and the *Rule of Saint Benedict* (cf. InvSyst 2.6), but these texts addressed an ecclesiastical rather than a lay audience and are therefore more closely associated with Latin (cf. TypSources, fasc. 44sq.: “disputes scolaires: faculté de théologie”). This phenomenon is even more marked for theological writing, which has remained dominated by Latin until today. There is nevertheless a Romance tradition of moral treatises and sermons in verse (cf. InvSyst 3.2.2; cf. also *ibid.* 3.3 for a number of other texts of a religious nature, as well as 2.6 for the *Formula di confessione umbra*, probably aimed at a lay audience; cf. TypSources, fasc. 68 “florilèges moraux”, 40 “*exempla*”, 30 “*complaintes funèbres*”).

Textual genres in medieval Romance were inspired above all by popular reform movements amongst minor orders (Franciscans and Dominicans), and diversified during the time of the Reformation and Counter-Reformation. Printed Romance-language Bibles were diffused to some extent; reform and counter-reform pamphlets appeared in the form of leaflets (cf. Saint Francis of Sales, who later became the patron saint of publishers and the press, and who instigated the Counter-Reformation in Savoy by means of such printed leaflets). Spiritual poetry and mystical literature developed in the 16th century with Spanish mystics including Teresa de Avila and San Juan de la Cruz, or French mystics such as Jean de Sponde (cf. TypSources, fasc. 57: “visions, révélations”). The Romance languages even made sporadic appearances in the domain of theological writing. During the modern period, parish registers were also numerous. These became established after the Council of Trent in particular (1545–1563; baptism, marriage and death registers were extremely stereotyped, but nevertheless contained unique information on personal names of common use in Catholic circles (cf. TypSources, fasc. 23 and 80: “visites pastorales”).

The role and impact of religious literature in Romance-speaking society has changed significantly over the centuries (cf. RSG 2, art. 179–182). In the early period of the Romance languages, these genres were of considerable importance for secular literature – this was true, for example, for liturgical drama and allegorical literature – as well as for other types of texts: a large number of glosses and glossaries emerged in conjunction with biblical and religious contexts. Theological discussion, moreover, contributed to the emergence of other disciplines such as modern philosophy and the humanities in general.

11.1.4 Texts reflecting specialised knowledge

The category of texts focusing on specialised knowledge concerns scientific and technical texts in particular, often subsumed under a category entitled ‘specialised prose’

(German *Fachprosa*). This category encompasses a wide array of different topics, ranging from medical or texts of a legal character to mathematical and linguistic treatises. They are generally heterogeneous in terms of both content and form, with as many subcategories as there are areas of specialisation in science, technology and the arts.

The sciences were already varied in the Middle Ages, and included theology (mentioned above), philosophy, rhetoric, law, medico-biological and mathematical-astronomical science. They were joined by other educational sources such as encyclopedias, chronicles, and works on technology or handicrafts. Furthermore, genres focusing on specialised knowledge have multiplied during the modern period.

Some of these texts display an aesthetic dimension comparable to that of literary texts and the majority of religious texts, whereas others are stylistically rather ordinary, such as medical prescriptions or cooking recipes. In present-day works in the hard sciences, texts in natural language have been replaced by formats that use numbers and symbols; the linear order has been doubled by a hierarchical order, achieved by the use of tables and illustrations (*cf.* RSG 2, art. 194 on terminologies).

Nevertheless, all works focusing on specialised knowledge represent the continuation of an older textual tradition: a medical treatise written in the 15th century may faithfully reproduce descriptions or prescriptions contained in other treatises from the Arabic-speaking Orient or from Greco-Latin Antiquity, making no mention of the knowledge possessed (or not) by its Western author. Similarly, an economics manual written in the 21st century contains terminology developed during the 19th and 20th centuries, with no acknowledgement of either the period or the circumstances in which it was created.

Medico-biological writing already formed a well-developed subcategory of texts in the Middle Ages. These texts included various defined genres: treatises on surgery, pharmacopoeia, writing on dentistry, treatises on cauterisation, herbals and works on medical treatment for horses and birds of prey. These genres witnessed a noticeable development in the 14th and 15th centuries in particular; they were based on Latin texts, many of which, in turn, derived from Arabic and Greek sources (*cf.* TypSources, fasc. 69, “consilia médicaux”; InvSyst 3.1.2; 3.3; RSG 2, art. 190–193; *cf.* also the short entries in the GRLMA).

Mathematico-astrological genres are less diversified, as are works on philosophical doctrine or rhetorical theory. Nonetheless, there are a substantial amount of texts pertaining to these categories, which must have played a role in medieval courtly and urban culture (*cf.* TypSources, fasc 39 “astrologie”, 32 “chimie”, 58 and 60 “traités de rhétorique et commentaires”, 59 “arts poétiques”).

Texts devoted to legal doctrine or theory proved particularly important for the elaboration of all Romance languages, as they presuppose decontextualised argumentation and thus necessitated the elaboration of complex syntax. The core tradition was that of the Latin *Corpus Iuris Civilis*, which was translated early on, but only partially, into Occitan (followed by Francoprovençal) in the mid-12th century, slightly later into Spanish, and in the middle of the 13th century into French (*cf.* Kabatek, *Die Bolognesische*

Renaissance, 2005 and Duval, *Les traductions galloromanes du Corpus iuris civilis*, 2016). Parallel to this learned tradition, a large number of customary law texts developed, beginning in England with the *Leis Guillelmi*, but more particularly in Gallo-Romance, while Italy mostly adhered to the Latin tradition throughout the Middle Ages (for legal sources, cf. TypSources, fasc. 6 “jurisprudence”; 22 and 63 “lois, traités”; 2, 10, 11, 43 and 46 “législation canonique, décrétale”; 41 “sources coutumières”; 48 “livres de droit, *artes notariae*”; InvSyst 6 “Lois et coutumes; chartes-loi”).

Finally, there were various (more or less developed) minor genres in the medieval Romance languages, such as bestiaries or treatises on agriculture, hunting, land surveying, cooking, painting and (at a later stage) architecture and maps (cf. TypSources, fasc. 77 “livres de cuisine”, 51 “portulans [= navigation maps] et cartes géographiques”, 75 “littérature cynégétique”). From a typological point of view, texts containing (meta-) linguistic commentary also constitute a minor genre, as such content was relatively rare in the Middle Ages (cf. no. 5 below). Encyclopedias were more advanced, bearing witness to fragments of knowledge available at the time, displaying diverse forms and transgressing the thematic boundaries of other genres. Collections of proverbs represented compilations of knowledge of a more popular nature. A further genre that is often considered to be minor but which was nonetheless well-developed is historiography, including chronicles in particular (cf. TypSources, fasc. 16 and 74 “chroniques”, 14 “Annales”, 37 “histoires et gestes”, 15 “généalogies”; these are excellently described by the GRLMA, vol. 11/1–11/3; InvSyst 5).

The diversification undergone by these textual genres in the 14th and 15th centuries continued throughout the Renaissance and the Enlightenment. Some of them adopted characteristics of literary texts, as in the case of travel literature, (auto-)biographical writing and even the earliest economic treatises (cf. TypSources, fasc. 38 “récits de voyage”; RSG 2, art. 187 “économie”). The quantity of individual works increased considerably, mainly due to the printing press. For the period between the 16th and the 18th centuries, even specialised bibliographies (such as those listing works concerning music or travel accounts) fall short in the attempt to account for the vast amount of texts produced in the areas of knowledge they aim to cover.

Whilst there are sporadic studies on specialised prose genres or even on Bible translations, there are as yet no synthetic overviews of the textual characteristics displayed by religious, scientific and technical genres. As previously indicated, there are far fewer linguistic studies focusing on these genres than on literary genres, even though the number of texts reflecting specialised knowledge far exceeds that of literary texts and – contrary to common belief – they display more lexical and syntactic diversity. The relationship between specialised and literary texts with regard to their linguistic physiognomy or to the conditions in which they were composed is another subject which has never been examined. As a consequence, most studies of a synthetic nature focusing on language history entirely disregard the areas of specialised or religious prose. The present-day period is the only segment of history for which intensive research into

specialised languages has been carried out; even this body of research, however, has not yet been completely integrated into variational linguistics, and there is a lack of diachronic research in this area.

Therefore, little is known of the role played by genres focusing on specialised knowledge in the elaboration of standard Romance languages, nor of their possible impact on spoken language. Historical linguistics would benefit greatly from the addition of these texts to the canon of sources studied.

11.1.5 Documentary writing

The last main category is that of documentary texts, which bear witness to medieval and modern administration, commerce and legal practice and were produced by secular and religious institutions, as well as by those involved in commerce. Although the diversity of genres displayed by documentary texts is limited compared to that of literary writing and texts reflecting specialised knowledge, the various types have existed since the medieval period. These include:

- charters and letters patent, referring to all types of judiciary acts: sales, donations, contracts, enfeoffment, agreements, arbitration and judicial sentences; as far as legal vocabulary is concerned, these sources display similarities with those of legal doctrine, although they are more open to the various aspects of daily (rural) life and contain a specific type of syntax that combines elements belonging to spoken language with hypotactic elaboration; *cf.* the following large sections of the InvSyst 7 “chartes”, 9.1 “cartulaires”, TypSources, fasc. 3 “actes publics”;
- in the domain of property management, financial statements relating to individuals, households and land; *cf.* InvSyst 9.2 “tarifs”, 9.3 “relevés”, 9.4 “notices”; TypSources, fasc. 18 “relevés de feux”, 19 “tarifs de tonlieux”, 28 “polyptyques ou relevés des droits dus à une institution ou à un individu”, 65 “matricules universitaires”;
- in commerce and administration, a large number of accounting registers and letters; *cf.* InvSyst 8 “lettres”; TypSources, fasc. 17 “lettres personnelles”; *cf.* also *ibid.* 4 “nécrologes”, 31 “catalogues de bibliothèques”.

Documentary texts differ from literary, religious and scientific texts in a number of ways: firstly, their purpose is to provide a testimony of a specific event in order to enable better management of social relations or property and in order to guarantee the maintenance of a political, commercial and social infrastructure. In historiographical terms, they are ‘unintentional’ sources, as opposed to literary texts, which are ‘intentional’ sources. Unintentional sources are the secondary result of a practical objective that transcends the original purpose of the text, whereas intentional sources are destined for posterity from the outset and are intended to be rewritten during the course of their transmission. This is an important difference, though it is relative, as, contrary

to what one might believe, social realities were invented in medieval charters as much as they were in *chansons de geste*; both categories of texts originated from and served the interests of the same individuals and social groups, and were written by the same scribes (cf. 11.2.2).

Another difference resides in the quantity of documents, which is significantly greater for documentary writing than for the remaining categories of texts, at least from the 14th century onwards. Thus the archives of a single Tuscan merchant, Francesco di Marco Datini, which, by chance, are conserved in their entirety, encompass 125,000 commercial and private letters written in 14th-century Italian (cf. Melis, *Aspetti della vita economica medievale: studi nell'archivio Datini di Prato*, 1962; Origo, *The Merchant of Prato*, 1960). The Vatican archives contain entries for no less than two million patronyms for the medieval period, and even those from a single region such as Lorraine include some 300,000 to 400,000 original pages written in French and dating between 1300 and 1500, as opposed to roughly 150 literary, religious and scientific works for the same period. Romance documentary texts thus encompass hundreds of millions of pages for the medieval period alone, and have continued to increase in number throughout the modern and present-day period, throughout which printed, typed and handwritten documents for practical use exist in immeasurable quantities. Today, documentary writing has also reached the sphere of the new media, which has resulted in the generation of a large quantity of digital records. At the same time, documentary texts represent sources of a serial nature, characterised by a high rate of repetition, a fact which has led to the widespread but erroneous assumption that they constitute a tradition almost completely devoid of linguistic interest.

Finally, there are differences in the manner of transmission: documentary sources survive either as originals or in the form of first-hand copies, whereas texts pertaining to the other genres are often handed down through multiple copies, due to their character as 'intentional' sources. Documentary texts continued to be handwritten throughout the modern period, before moving directly to digitised form at the end of the 20th century. For reasons related to their practical purpose, they were not printed (cf. 11.2). Moreover, they are typically conserved in archives, whereas literary works and texts relating to specialised knowledge are conserved in libraries.

Like the other textual traditions, documentary genres do not constitute an isolated group of texts. Epistolary genres, for example, were influenced by rhetorical thought and thus displayed affinities with literary texts from the medieval period onwards. The texts relating to legal practice were in constant interaction with learned and customary law, although they developed their own lexical and syntactic features. Parish registers and other sources relating to ecclesiastical administration reflect specialised knowledge but are also modelled on documentary writing. During the modern period, the dividing lines between administrative writing and scientific or technical writing have become somewhat blurred, as the latter have evolved considerably since entering into general use.

The oldest documentary texts from the 12th and 13th centuries were studied by historical linguists relatively early on; the material produced in the six centuries that followed, however, has been almost entirely neglected. This is due to the particular challenges posed by the study of sources of a serial nature as opposed to individual sources (such as literary texts or even the oldest documentary sources). Nevertheless, documentary texts contributed as much to the elaboration of the Romance languages in writing as the other three main textual genres, if not more. They represent a category in their own right, which follows its own patterns of evolution and displays a specific type of interaction with texts belonging to other domains.

The establishment of an inventory of the major categories of medieval Romance texts – consisting of religious and secular literary writing, texts focusing on specialised knowledge, and documentary sources – provides a solid foundation for the systematic description of the textual patrimony and for its use in linguistic analysis.

11.2 ‘Material encoding’ and manuscript culture

11.2.1 Material encoding of texts in the history of the *Romània*

In the study of old texts and the consideration of genres and discourse traditions, it is essential to understand the historical context which led to their appearance but also the characteristics of the ‘material encoding’ of texts, which depends on the conditions of the period regarding mediality. Entire branches of philology have developed around the dating and localisation of old manuscripts and texts, considering both physical – external – aspects (material, form, place of conservation) as well as the internal dimensions of content and linguistic particularities.

Generally speaking, the support material for texts – which may be manuscript, printed, digitised or oral – is a determining factor for the contents and linguistic form of a given text and also for its diffusion and transmission. Before one can edit a manuscript or a text or interpret the language of a text or textual genre, the characteristics of its material transmission should be identified. The practical work of the language historian thus begins with codicological analyses.

Among the different methods of textual encoding used in the *Romània*, writing in manuscript form has the longest tradition. Manuscript texts on portable bases such as sheets of papyrus appeared shortly after cuneiform writing etched on clay tablets and inscriptions in stone, which were widespread in Antiquity, above all for hieroglyphs (cf. no. 3 below). Manuscript writing on papyrus, parchment and – later – paper, represents the dominant form of linguistic encoding from the beginnings of Latin writing up until the 16th century. Other methods, although marginal, existed from Late Antiquity, including inscriptions on stone, wax tablets, metal (such as coins) and, more rarely, woodblock printing.

Manuscripts remained dominant in the 16th century when the printing press became established, as they still fulfilled numerous functions in administrative as well as personal writing up until the 19th century. Thereafter, typewriters would restrict the use of manuscripts to private, educational and spontaneous writing – all spheres in which handwriting has been in competition with personal computers since the end of the 20th century.

The printing press with movable type revolutionised the world of writing. Between 1440 and 1600, this new technique was elaborated and quickly spread throughout Europe. Forged on the model of luxury manuscripts, printing ultimately developed its own characteristics. In parallel, the production of texts using this new technique increased rapidly: whereas the books printed in the 15th century (so-called *incunabula*) could still be listed in a single, though large, catalogue (*Gesamtkatalog der Wiegendrucke*, 1925–) and there are various catalogues that aim to provide exhaustive descriptions for printed works of the 16th century, such a task is no longer possible for the 17th and 18th centuries, given the explosive increase in production.

The development of printing played an integral role in the formation of standard languages in Europe (cf. 10.5.3 no. 2). This encoding technique enabled a wide diffusion of writing and thus of written language among an audience that was geographically disparate and socially heterogeneous. This promoted the development of a homogeneous language with little variation, reduced diatopic markers and of high prestige. Printers were involved in this process as they corrected the language of the texts that they produced. Figures such as Aldo Manuzio and Gabriele Giolito played a key role in this domain (cf. Trovato, *Con ogni diligenza corretto*, 1991). Some authors even directly implemented the new requirements: the French poet Pierre de Ronsard, for example, reworked his *Amours* five times, the successive editions (which appeared between 1552 and 1589, the last one posthumously) reflecting the author's search for an appropriate linguistic form in minute detail (cf. Terreaux, *Ronsard correcteur de ses œuvres*, 1968).

Works prepared for a large audience in this way experienced a growing diffusion and, thereby, favoured language standardisation. Their adherence to the new norms, in turn, contributed to their success. Without the printing press, neither Bembo's *Prose della volgar lingua* (1525), nor Luther's Bible (1534) nor Cervantes' *Don Quixote* (1605–1615) would have had the impact they had on the elaboration of written language.

Early printed books mainly contained the same texts as luxury manuscripts; these were, above all, the Bible and other religious works, secular literature and specialised prose. Soon, however, the printing press furthered the development of new, less prestigious genres, including pamphlets and leaflets (the forerunners of newspapers), as well as popular literature, such as farces and Nativity plays (for example, the French traditional '*noëls*', cf. Rézeau, *Les Noëls en France aux XV^e et XVI^e siècles*, 2013).

The 19th and 20th centuries brought about new changes to the culture of printing. Newspapers spread, becoming a dominant medium for written language and a rich source for linguistic study. A vast diversification of printed material took place; at the

end of the 20th century, new possibilities of printing brought about by computer technology further increased the diffusion of information.

This diversification of printed sources was contemporary with the development of new means of textual production: typewriters and, later, personal computers, replaced manuscripts in administrative, professional and even private writing. In recent years, the Internet has catalysed forms of semi-oral writing. Audiovisual media (radio, cinema, television) and the multimedia-oriented use of computers continue to transform the context of spoken language. The impact of the media revolution caused by information technology is no less significant than the consequences of printing five hundred years earlier, and its effects will only become apparent in the course of the next few decades.

The computerised management of data has already changed the conditions in which communication takes place, and it has altered the nature of writing, which no longer has the advantage of longevity that is intrinsic to stone inscriptions, manuscripts and printed texts. The rapid changes that take place in the field of information sciences render the majority of computerised texts inaccessible within the course of a few years, thus creating new problems relating to knowledge management and new challenges for specialists of written culture.

11.2.2 Manuscript culture

1 The support material of manuscripts

Philology developed on the basis of the study of classical and medieval manuscripts. This causes great difficulties for linguistic analysis, of both a material and a theoretical nature. The material encoding, and thus the medium-related aspect of language, generates constraints which are independent of its nature but which have a strong impact on its physiognomy. The support materials and writing that encode and transmit the language of ancient textual sources therefore require particular attention.

The earliest material used as a support for classical manuscripts was papyrus, which appeared alongside clay, stone or marble and wax tablets during Antiquity. Papyrus was used in the form of single sheets (so-called *chartae*) or scrolls consisting of several sheets glued together. It was used throughout almost the entire first millennium, in particular in Merovingian and Carolingian chanceries (until the 9th century).

Whereas Greek papyri preserved in deserts in the Eastern part of the Roman Empire survive in considerable quantities, Latin papyri from the Western Empire and later Germanic kingdoms have almost all been lost. This is one of the reasons for the relative scarcity of documentary sources for the transitional period between Latin and Romance, another being the low level of literacy which characterised that era.

The only sources that survive from this period are stone inscriptions and coins. Inscriptions reflect a limited number of textual genres, which are rather stereotypical in nature (these include funerary or commemorative inscriptions as votive offerings; *cf.*

the epigraphic collection of the CIL, *Corpus Inscriptionum Latinarum*, 3.3.4 no. 2); coins bear very little actual text and are thus more interesting for the history of economy and religion than for the history of language (cf. nevertheless the studies by Chambon and Greub, *Les Légendes monétaires mérovingiennes*, 2000, and Eufe, *Merovingian coins and their inscriptions*, 2012).

Parchment is of much greater importance in Romance studies. This resistant material appeared in the 2nd century BC and accounts for nearly all textual sources for the 7th to 12th centuries that are known to us today (cf. the collections indicated above, 10.3.5 no. 2, which contain numerous facsimiles). Charters were typically written on the flesh side of the hide; for longer texts, several pieces of parchment were sewn or glued together and then rolled in a similar manner to antique papyri. From the 14th century onwards, a new technique began to spread, which involved four sheets (*folios*) of parchment that were folded into a booklet (consisting of 4 x 4 = 16 pages), modelled on antique codices which consisted of several wax tablets bound together. This technique, which used the less valued hair side of the hide, forms the basis of modern book production, and in the medieval period it was used for many literary texts written on parchment⁴⁶.

The high cost of this material explains why pieces of parchment that had already been written on were reused. Sometimes the original text was erased by scraping, enabling a new text to be written in its place (as in the *Appendix Probi*, for instance [cf. 6.4.4], which was written in the 8th century in northern Italy on such a 'palimpsest' (= a scraped and reused manuscript) containing fragments of a *Vetus Latina* from the 5th century). This technique of erasure was more commonly used for individual corrections; moreover, pieces of parchment were often used in the binding of paper books, usually without having had their contents erased first. Such bindings may contain data of great interest for the history of texts: thus, the first witnesses of the *Scuola siciliana* were discovered on the binding of a register known as the *Memoriale Bolognese* (1292).

From the 12th and 13th centuries, parchment was in competition with paper, which was imported into Europe from China and the Near East via Sicily and Spain. In the 14th century, the production of this versatile and supple material developed rapidly to the detriment of parchment, becoming predominant by the 15th century. Parchment was

⁴⁶ Medieval manuscripts often do not contain original page numbers; these are thus introduced by modern editors. Numbering takes into account the front (*recto*) and back (*verso*) sides of individual *folios*. Medieval books and books from the modern period are thus cited as follows: "fol. 1r[ecto] – 1v[erso], 2r – 2v", etc. If the text is written in columns, these are also indicated: the *recto* of a *folio* thus may contain two columns *a* and *b*, and the *verso*, columns *c* and *d*: "fol. 1a – 1b [= 1r] – 1c – 1d [= 1v]". In printed books, numbering within quires may be introduced: A – A₂ – A₃ – etc. (= fol. 1 [recto and verso], 2 [id.], 3 [id.] of quire A), B1 – B2 – B3 – etc. (= fol. 1, 2, 3 or quire B); or Ai, Aii, Aiii (= id.). Only in the 16th and especially in the 17th centuries did numbering by pages, which was rare before, become commonplace.

still used at this time for luxury books and legal documents of great value, and it continued to be used sporadically until the 18th century.

The first major advantage of paper was the low cost of the textile fragments which constituted its primary material. This factor was complemented by the relatively high resistance of paper to deterioration and the ease with which it could be bound into booklets. For example, the 150 copies of the Latin Gutenberg Bible printed on paper were much less expensive than the thirty copies on parchment, which required the hides of approximately one thousand cows. The production of texts that involved such considerable financial investment necessarily remained a rare exception. Had paper not enabled the industrial production of texts, neither the considerable development of Latin and then Romance writing between the 13th and 15th centuries, nor the rise of the printing press in the 16th and 17th centuries would have been imaginable.

The history of support materials for writing is thus intimately linked to the elaboration of written languages. It also provides useful information for determining the period and context in which texts were written. For example, the date of production of paper can be determined through the study of watermarks: for the period between 1282 and 1600, over 16,000 watermarks have been collected (*cf.* Briquet, *Les Filigranes*, 1907; Stussi 1994). This method can lead to fairly accurate results, and it has been established that the production of paper usually occurred in the period immediately preceding its use.

2 Writing

Writing underwent significant changes from Late Antiquity onwards, determined in part by the type of support material and writing implements used. In early periods, the latter were quills made from bird feathers and reed stems that were sharpened to a point (= *calamus*). In Roman Antiquity, four types of writing were used:

- square capitals, also known as ‘book hand’, the oldest type of writing that has survived in the form of today’s capital letters;
- cursive, the newer, minuscule (i.e. lowercase) form of which was used in documentary texts;
- uncial, an intermediate form, close to square capitals;
- semi-uncial, closer to the cursive form.

From these, several cursive, minuscule scripts developed in 7th- and 8th-century Latin Europe. These included the Beneventan script in Italy, the Merovingian cursive, Visigothic script and, finally, the Carolingian (or Caroline) minuscule, which would achieve the greatest success. Chiefly developed in French monasteries, it was inseparable from the Carolingian reform, which had an impact not only on writing but also on the language used in written texts (*cf.* 10.3.5 no. 3, as well as the figure below, which reproduces

the manuscript containing the *Sequence of Saint Eulalia* (dating from the end of the 9th century).

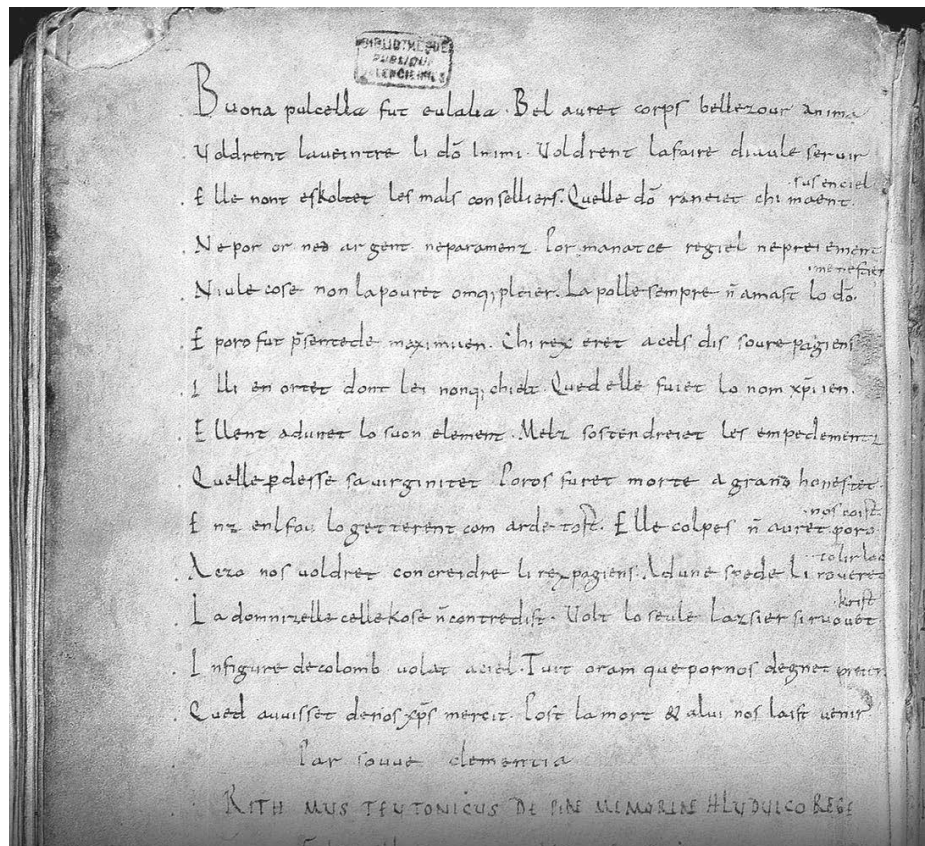


Fig. 32: The *Sequence of Saint Eulalia* (Valenciennes Municipal Library, MS 150, fol. 141v)

Carolingian minuscule was a forerunner of the Gothic scripts that developed in the 12th and 13th centuries and whose variants were still dominant in the centuries that followed (cf. the example of a charter in fig. 34)⁴⁷.

The intensification of written production in the 12th and 13th centuries favoured the differentiation between cursive writing, used for practical and documentary purposes ('chancery cursive'), and book-hand (*libraria*), more elaborate and aesthetic and reserved for literary and scientific texts. The two types of writing each had defining characteristics involving the marks used for the structuring of texts (i.e. punctuation)

⁴⁷ Today, there are excellent Internet sites that illustrate manuscripts across the centuries; cf. for instance, the exemplary site <<http://www.codices.ch>> for manuscripts conserved in Switzerland.

as well as different systems of abbreviations, some of which are very complex (cf. the dictionary of abbreviations by Cappelli, 1899, excellent despite its age).

When the manuscript as a prestigious document gave way to the printed book, book-hand script served as a model for printed characters, thus becoming the predecessor of today's print types. In contrast, the cursive used in documentary texts and, a little later, for writing of a private nature, continued to develop the character as a hand used for the purpose of speed: whereas some degree of familiarisation is required in order to be able to read medieval cursive texts, 16th- and 17th-century chancery cursive is indecipherable to the untrained eye. The fact that documentary sources from this period have been almost completely neglected by research can be explained in part by the effort required to decode them.

The cursive styles used in present-day handwriting were formed later, during the 19th and 20th centuries. They display differences according to country and schooling tradition, as well as a great deal of individuality and idiosyncrasy, which was an exception in the writing of earlier centuries.

Latin script was largely dominant in the *România*, though there existed various other traditions that used other alphabets (cf. 6.4.1). These include the Old Spanish *kharjas* written in Arabic or Hebraic characters, the Greek alphabet in various Old Italian texts (e.g. in a *Chirurgia* by Guglielmo da Saliceto, 1275), the Hebraic letters in Old Occitan medical glossaries (cf. Mensching/Savelsberg *Manual of Judaeo-Romance linguistics and philology*, MRL 31, 2023) or the Cyrillic writing habitually used in Romanian texts from the 16th to the mid-19th century, and until 1991 in Moldova (cf. Baglioni and Dahmen, *Editionsprobleme bei Balkanica*, 1997). The transposition of Romance texts into other writing systems poses problems of phonetic interpretation; conversely, such transcriptions facilitate the identification of particular pronunciations.

3 Text and manuscript production

The production of written texts in the Middle Ages was inextricably linked to the way in which manuscripts were produced and diffused. To persist in space and time – which is one of the intrinsic purposes of writing – texts had to be not only written but also copied by hand. This double process of production and reproduction was accomplished by professionals, who were educated in ecclesiastical *scriptoria* and noble chanceries. It was there that the manuscripts also received decorative elements, such as rubrics (letters written in red < Lat. *rubrica* “red earth, ochre”), illuminations and miniatures (cf. 9.3.1). Aside from these professionals, few men and even fewer women learned to write.

Reading was more widespread, since numerous merchants and nobles were literate; for the most part, however, they produced few texts and therefore remained relatively uninvolved in the history of language (see, however, the counter-example of Datini, 11.1.4). In the Late Middle Ages, the establishment of schools and universities promoted a socio-cultural diversification of the type of individuals who were capable

of writing. This development intensified in the 16th century, though it was not until the 18th and 19th centuries that non-professional writing became unremarkable (cf. 10.5.4).

In the Middle Ages, established writing centres determined to a large extent the language used by all scribes, especially those working in the *scriptoria* of abbeys, monasteries and bishops' palaces, and in the chanceries of important princely courts. Public scribes or notaries, as well as individuals associated with the administration or universities, had less influence on the development of standards (cf. Glessgen, *Les lieux d'écriture dans les chartes lorraines du XIII^e siècle*, 2008). The dominant *scriptoria* and chanceries produced texts belonging to all genres: the main discourse traditions (i.e. literary, religious, texts reflecting specialised knowledge, and documentary texts) were all produced in the same places by the same authors and scribes (cf. 11.1). These institutions developed their own methods of production and reproduction and were thus responsible for the emergence of exemplary linguistic varieties.

As we have seen, documentary writing linked to administration, accounting and jurisdiction was by far the most voluminous in all historical periods, accounting for 95, if not 99 % of all written texts (cf. 11.1.5). Producing a written statement of a sale, lease or court decision had the purpose of preserving a lasting witness of the act concerned and of ensuring its authenticity. It was therefore sufficient to produce one or two copies of such a text for each of the two parties involved. In particular cases, the text was written twice on two halves of a single piece of parchment, which was then cut in half; the authenticity of such documents (known as 'chirographs') could be proven by matching the cut edges of the parchment. In some cases, there was also an abbreviated version of an act, produced either before or after the fully elaborated version, and documents were sometimes reproduced later for specific aims. Most frequently, however, documents belonging to an institution (e.g. an abbey, a noble house) were copied and stored collectively in so-called cartularies. Such collections of charters may present significant chronological discrepancies: during the 16th and 17th centuries, it was common to assemble all the charters produced by an institution and to reproduce in a single volume all original documents dating from between 1100 and 1450 (cf. 10.3.5 no. 2; 11.1.4).

11.3 The problem of copies and textual adaptation ('mouvance')

11.3.1 Copying and adaptation

As we have seen, a characteristic of documentary writing is that it is often transmitted as the original, in contrast to all other genres. As numerous reproductions of documentary texts also exist, it is possible to study in detail and to gain insight into the relationship between the language of an original text and its copy. Every copy alters the linguistic form of its model to some degree, with the single exception of chirographs. While contemporary copies may only differ slightly, divergence increases over time as well as over space. Regardless of chronological discrepancies, copies display a natural

tendency to neutralise the diatopic marking present in their models, by the alteration of their graphemic and morphological features in particular. Copies thus contribute to a process of linguistic homogenisation and (pre-)standardisation.

The process of copying, then, may be likened to a second phase in the written elaboration of a document. The compilation of cartularies at a later stage, for instance, even involved the unacknowledged translation of texts from Latin into a Romance language. This trend is particularly prevalent in literary texts and writing focusing on specialised knowledge. Whereas charters and written accounts are produced following existing models provided by defined textual genres, literary and scientific texts reflect their respective discourse traditions in a more indirect manner. They appear as the echo of earlier texts, which they transform while at the same time maintaining certain elements of content and form. Some medieval works arose from the creativity of an individual author who chose a subject and put it into writing; the resulting text was then actively diffused with the help of copyists. More often, however, literary, religious or scientific texts were adaptations of earlier texts (Classical or medieval, Latin or Romance). They were the product of processes of translation or rewriting, and sometimes even adaptations of old oral traditions. Thus, a romance such as Chrétien de Troyes' *Cligès* reworks material from the British legend of *Tristan and Iseult*; Boccaccio's *Decameron* on the other hand is founded on the earlier traditions of popular novellas and French *fabliaux*, as well as texts from Greek and Latin Antiquity. When adapting and composing works, professional writers coloured them with their own aesthetic ideals and linguistic characteristics.

Through the process of copying, the form and content of such works were subject to change – as the result of scribal errors and misunderstandings⁴⁸, but much more so through the intentional, often generalised reinterpretation of written content, as well as graphemic and morphological neutralisation and further adaptation of the text to the linguistic model used by individual scribes. In this way, many Anglo-Norman texts were Picardised in the 13th and 14th centuries, while still maintaining specific elements

⁴⁸ The characteristic phenomenon of copyists' errors in manuscript transmission is due in part to specific features of medieval writing styles. The letters *i* and *j*, *u* and *v* were not systematically distinguished, and the formal representation of *n* and *u* and, often *ni*, *in* and *m* often coincided. This frequently led to formal identity between lexemes: the French words *vivre*, *vinre*, *juure* ("jure") and *nuire* could all be written in exactly the same way, comprising four down-strokes followed by the sequence *re*. Moreover, the letters *f* and *s* have a similar *ductus* (i.e. the order and properties of strokes involved in tracing a letter), as do *t* and *c* and, sometimes, *r*.

In addition to errors resulting from misreading, a whole range of accidental scribal errors can be observed (*lapsus calami*, literally "slips of the pen"), including phenomena such as haplography (letters, or sequences of letters that should be reduplicated are written only once: e.g. It. *stale* instead of *statale*) and dittography (the opposite phenomenon: e.g. Lat. *manenet* instead of *manet*), as well as the unintentional omission of abbreviation symbols, letters, lines or even entire paragraphs.

of their original language, and many Old French texts from the 12th century took on the character of a Middle French text when copied in the 14th or 15th century.

11.3.2 Textual 'mouvance'

It is often difficult for present-day readers, accustomed to modern conceptions of authorship and originality, to understand the ease with which texts, when transplanted into different times and places, were adapted in terms of both content and linguistic form. The following example illustrates the fundamental instability and changeability of medieval texts, which modern critics have referred to as 'mouvance':

- The tale of Griselda, which is present in Boccaccio's *Decameron* under the title *Il marchese di Sanluzzo* (ca. 1370) was taken up shortly thereafter by Petrarch and adapted in Latin (*De insigni obœdientia et fide uxoris*, 1373), omitting the name of the original author.
- The text then made its way into French, where it was adapted, and at least five versions were produced during the following thirty years: the tale was integrated into the *Miroir des dames mariées* of Philippe de Mézières (1384–1389), the *Ménagier de Paris* (ca. 1393) and the *Livre de la cité des Dames* by Christine de Pizan (1404–1405).
- Between the 15th and 18th centuries, dozens of adaptations appeared, in both manuscript and printed form, most of which mention neither Boccaccio nor Petrarch (with the exception of, for instance, Perrault's *La Marquise de Salusses, ou La Patience de Griselidis*, 1691).
- The different adaptations took the form of various textual genres, from the morality tale to the mystery play, the verse tale, narrative or poem. This variation was accompanied by a reinterpretation of the original message of the text (the initial emphasis on women's superiority shifted to the suffering of women; cf. Berger, 'Mouvance, variance' und die Folgen: *Griselda und ihre 'Nachkommen'*, 1997).

The transformation of Griselda's tale illustrates an absence of the concept of authenticity of the original – in both form and content – among the writers who reinterpreted it. The same mindset prevailed in painting and sculpture, where biblical figures and personalities from Antiquity were represented according to the dress customs of the artists' times.

The flexibility of the pre-modern principles of textual transmission also explains the wealth of false documents produced in the medieval period. They can be illustrated by the famous case of the donation of the emperor Constantine († 337): papal scribes forged a document between 750 and 850, testifying a fictive donation supposedly made by the Western Roman Emperor to Pope Sylvester I in 335. By means of the written text, they thus re-established what they believed to be the true order, since according to their

world view, God and His representative on Earth were entitled to reign over the whole world.

The ‘mouvance’ displayed by scientific, technical and other specialised texts is just as pronounced as that observed in literary works, even though the way in which it manifests itself varies due to differences in the restrictions imposed by content. As a consequence, the source material for the history of language as well as for that of literature and the sciences consists of a multitude of copied manuscripts, strongly divergent despite their common descent, rather than a clearly defined set of original texts. We have already mentioned the existence of almost one thousand manuscripts of the Latin *Legenda aurea*, translated several times into Occitan, French, Catalan and Italian (cf. 11.1.2). Another illustrative case is Dante’s *Divine Comedy*, which exists in the form of approximately six hundred Italian manuscripts.

Even in a case such as the latter, where an individual author can be identified, each copy can be of interest to the history of the language and the text in question. Copies provide valuable information on the diffusion of a work, and reflect the different historical stages of a language, which, in turn, bear witness to language change. Finally, these copies represent the form of the text as it was perceived by the people of the medieval period, for whom the concept of the ‘original’ simply did not exist.

Manuscript production continued into the modern period, not only in documentary writing, but also in literature, scientific and technical texts. For many modern authors, important collections of manuscripts are available, some of which remain unedited. In contrast to documents from the medieval period, the majority are autographs (i.e. in the author’s own handwriting); this makes it possible to follow the process of literary creation. The scope of philological research has broadened to include present-day sources, with the study of handwritten manuscripts by authors such as Manzoni, Balzac, Proust, Kafka, Musil or Joyce.

In addition to these major genres, writing of a private and semi-private nature emerged, including correspondence (with a tradition that extends back to medieval times), memoirs and autobiographies. The use of writing by the partially educated opened up new horizons for linguistic analysis (cf. 10.5.3 no. 2 for the 16th/18th centuries). The most striking example of this comes from the period of the First World War, during which millions of letters were exchanged between the French, German and Italian soldiers at the front and their families (cf. Rézeau, *Les mots des Poilus*, 2018 and Carles/Glessgen, *Les écrits des Poilus*, 2020).

11.3.3 Aspects of textual transmission

The textual heritage that has been handed down from early periods represents only a small percentage of what was originally produced in writing. Depending on the period and place in question, up to 90% of medieval and modern texts were destroyed, sometimes by fires, or lost by other means over the course of the centuries.

Thanks to the large number of copies, the loss of works from the medieval period is far less considerable than the loss of material witnesses (i.e. single manuscripts). The situation can be likened to that of Classical Latin literature. These texts would be as little known to us as paintings from the same period if it were not for the numerous copies produced during late Antiquity and as a result of the Carolingian reforms, not to mention the revival they experienced during the Renaissance. For the medieval period, the well-studied example of troubadour poetry neatly illustrates this phenomenon: the textual genre developed for the most part during the period between the beginning of the 12th and the end of the 13th century, and is known to us today through approximately one hundred *chansonniers* – collections of poems that were compiled in the 13th and 14th centuries (cf. 10.4.3 no. 1; Glessgen, *La langue des premiers troubadours*, 2022b). The fact that no earlier written witnesses survive is suggestive of significant losses at first glance. There are, however, many mentions of the names of troubadours (both in the poems themselves and in other sources), which allow us to place the issue into perspective: in total, some 460 troubadours are known by name, including as many as forty women, called *trobairitz*. In most cases, these names refer to the presumed authors of the poems, and texts by about 400 of these authors have been identified; only about 10% of the authors have no known surviving works (Rieger, *Les troubadours fantômes en Italie*, 1993).

Comparable rates of textual transmission are confirmed by other sources. Thus, manuscript or printed witnesses survive for at least part of the works of 80–90% of medieval authors, and for 90–95% of modern authors.

These figures serve to highlight the importance of studying the material transmission of old texts before proceeding to linguistic and historical analysis. The existing manuscripts of a given group of texts must be classified and their reciprocal relationships determined. Since nine out of ten manuscripts are presumably lost, a painstaking process of investigation and reconstruction necessarily precedes and guides the edition and study of texts.

11.4 The theory and practice of editing

Editorial philology must comply with two contradictory requirements: authenticity in the reproduction of medieval witnesses and comprehensibility of the edited texts. Maximum authenticity of a medieval text is ensured by the means of accurate transcription or even a photographic reproduction of all known manuscript (and printed) witnesses. However, even the results of a maximalist solution of this sort necessarily differ from historical reality, since photographs distort the material aspect of manuscripts and since it must be assumed that the majority of witnesses no longer exist. Furthermore, this form of reproduction is only feasible in the case of texts with a limited number of versions; it becomes burdensome and difficult to manage in cases where there are twenty or even as many as two hundred parallel manuscripts. The

modern reader confronted with so many material witnesses, themselves difficult to read, would find it impossible to reconstruct the text.

At the other end of the scale is the reconstitution of an ‘ideal’ original text and its translation into a modern standard language. The objective of comprehensibility is thus completely fulfilled and the text becomes immediately accessible. However, this solution results in the reconstruction of a text by the modern editor according to his or her own intuition and, if translated, it loses all interest for linguistic study.

All forms of editing – be they based on handwritten or printed texts, early or more recent – are situated somewhere on a continuum between these two extremes. Opting for one method or another depends in part on the objectives of the editor, who may be more interested in the content of a text (e.g. literary scholars and historians) or in its form (e.g. linguists). The state of the transmission of a text with regard to its content and linguistic characteristics also influences editorial choices. If there is only a single witness, the editor will necessarily choose to reproduce the text contained in this manuscript or printed book. Nevertheless, there is a large range of possible approaches, ranging from diplomatic editions, which reproduce minute details (i.e. punctuation, capital letters, the separation of words, the graphic form of letters or the spatial layout of the text on a page) to critical editions, which introduce modern punctuation, restructure the text into paragraphs, correct scribal mistakes and explain obscure passages.

When approaching works with a complex textual tradition, the editor will first attempt to identify and classify the manuscripts in question, before making decisions on the best way to reproduce them. External elements of a manuscript (i.e. its support material, handwriting, place of conservation) and its internal elements (i.e. linguistic aspects and content) always allow an approximate dating and can also provide clues as to its localisation. Copyists’ errors and textual variation then make it possible to determine relationships between manuscripts, although they often also highlight the existence of numerous missing links that cannot be resolved. The graphic representation of a textual genealogy of this type is called a *stemma*, as described in Bédier’s pioneering study, *La tradition manuscrite du Lai de l’Ombre* (1929; cf. Baker et al., *L’Ombre de Joseph Bédier*, 2018; cf. 2.2.3).

The following figure shows the relatively simple *stemma* established for the Latin manuscripts of the *Moamin* treatise on falcon medicine (of Arabic descent):

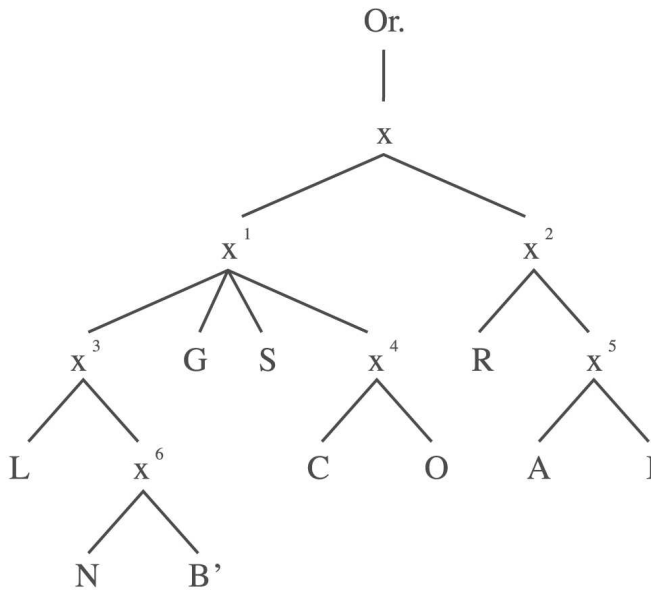


Fig. 33: *Stemma of the Latin manuscripts of Moamin*

Source: Glessgen, *Die Falkenheilkunde des 'Moamin'*, 1996.

Remarks:

- In the stemma, 'Or.' stands for the lost original, which is presumed – rightly or wrongly – to have existed. In this specific case, the hypothetical original is the Latin translation of an earlier Arabic text; x is the archetype, directly descended from the original and forming the presumed basis on which the existing witnesses depend; x¹ and x² are – also presumed – 'sub-archetypes', upon which certain known manuscripts (G, S and R) depend, as well as other presumed manuscripts which are lost (x³–x⁵).
- G, S, R, L etc. are the actual manuscript witnesses known to us today, and whose position in the stemma does not necessarily reflect their actual age. In the given example, MS. A (Paris B.N. lat. 7019, dating from the 14th century) is probably older than MS. G (Vatican, Bibl. Ap. Vat., Reg. lat. 1111, from the 16th century), although G is situated on a genetically higher branch of the *stemma* and could thus (mistakenly) be assumed to be older.

Stemmata may be considerably complicated owing to the quantity of witnesses and the nature of the relationships among them. In cases of 'open traditions' with multiple rewritings and phenomena of 'contamination' between texts from different branches, the accurate reconstruction of a *stemma* is sometimes impossible.

Determining the relationships that exist among the available witnesses enables the editor to make choices on an objective basis. A first possibility consists in the exhaustive (more or less diplomatic or interpretative) transcription of all manuscripts. Such 'synoptic editions' demand a considerable effort on the part of both editor and reader and are rather rare (*cf.* e.g. Noomen's edition of French *fabliaux* (1983–1998) with the parallel reproduction of eight manuscripts). For more complex textual traditions, this solution is not feasible; at most, it is possible to choose manuscripts reflecting different levels of the *stemma* in order to account for textual and linguistic transformations that took place throughout the textual tradition.

A second possibility consists in the choice of a single manuscript as a representative which is well-preserved and presumed to be close to the original. This manuscript, known in French as the 'manuscrit de référence' (reference manuscript), is then edited using a diplomatic approach. The interesting variants found in other manuscripts are given as alternative forms in the footnotes or can even be inserted into the text to correct obvious errors or to fill any gaps in the chosen manuscript. This method was explicitly laid out for the first time by Joseph Bédier in his editions of the *Lai de l'Ombre* (1890, expanded in 1913) and of the *Chanson de Roland* (1921/29), and has since been known as 'Bédier's method'. This method is a logical choice when a good quality manuscript has survived and when the textual tradition displays minimal variation.

A third method consists in the attempt to reconstruct the lost original on the basis of existing manuscripts. Rather than adhering to the specific forms present in a single manuscript, variants taken from different manuscripts are combined, with a certain amount of flexibility, to create what the editor supposes to have been the intention of the medieval author. This is the oldest method, which was scientifically established by Karl Lachmann and explained in the *Commentarius* section of his edition of Lucretius (1850). Since Classical Latin possessed orthographic and morphological norms, the 'Lachmann method' poses fewer problems for Latin texts than it does for the medieval Romance languages, for which it is necessary to reconstruct both the text and linguistic forms that are not based on norms and where the manuscripts are highly variant.

This third method is useful when only late witnesses of a literary work have survived, as in the case of Troubadour poetry. In a case such as the *Cronica de l'Anonimo Romano*, written in the 14th century but only transmitted by witnesses from the 16th, basing an edition on one of the surviving witnesses would necessarily result in the reproduction of a text far removed from the original state of the work; this could speak in favour of an attempt at reconstruction, even though such an attempt would inevitably remain imperfect (in this respect, *cf.* the debate between editor Giuseppe Porta and lexicographer Max Pfister, 1983–1985).

From a theoretical point of view, the three methods of textual edition – i) synoptic, ii) based on a single base manuscript and iii) relying on reconstruction, complemented by a more diplomatic or more interpretative perspective – provide the framework for every editorial project. In practice, many intermediate approaches representing compromises between these three basic methods have been developed. Editions may differ

greatly, depending on the editor's faithfulness with regard to manuscript witnesses. Editors often make substantial changes to the linguistic form of texts, even though, in their introductions, they profess to have adhered rigidly to the original readings of the text. Sometimes, a single edition is impossible, as in the case of the novella of *Griseldis*, which displays such marked *mouvance* that the reproduction of the different textual strata is necessary.

As might be expected, the philological and linguistic study of printed texts does not pose the same problems as the study of manuscripts. Normally, photographic reproductions of these texts are sufficiently accessible to modern readers, who are able to understand the syntactic and lexical features of printed texts dated after the end of the 16th century, since their language is sufficiently close to that of today. Nevertheless, the layout and punctuation, as well as the orthographic choices found in such texts, display characteristics that deserve closer study.

Difficulties relating to the content and resulting from lexical, syntactic and textual choices, which are often of a very specific nature, only come to light during an attentive reading of the texts. Whereas '*stemmata*' of printed texts are easier to establish than those of manuscripts, owing to the initial existence of a large number of identical copies, a text which has been subjected to a long series of re-editions presents almost insurmountable problems, since it is impossible to decide which version should serve as a point of reference.

11.5 Editing texts for linguistic analysis

11.5.1 Methods of textual edition

The methodology for the edition of old Romance texts has evolved considerably since the beginning of the 19th century. At first, interpretative methods were dominant: editors tended to normalise the linguistic form of medieval manuscripts by minimising graphic and morphological variation and by going so far as to alter words or parts of sentences in order to make the text more comprehensible. Medieval works were reconstructed according to the ideal of a unique text, composed by (or written under the direction of) a single author. Linguistically speaking, texts were adapted in order to comply more closely with the norms of a modern standard language – an operation that corresponded to the ideological endeavour of reconstructing a national past (cf. 10.6.3). This same mindset led to the rebuilding of medieval churches in a historicising style, the construction of Neo-Romance and Neo-Gothic buildings and the restoration of castles in the style of the revolutionary French architect Viollet-le-Duc (1814–1879).

This 'rewriting of the Middle Ages' represented a first step towards a new awareness of history, but it nevertheless carried weighty implications for literary and linguistic studies, which only later distanced themselves from the ideological image promoted by early interpretative editions (cf. Grimm, *Mittelalter-Rezeption*, 1991). Still today, we

are often obliged to use unreliable editions of texts for the early periods of all Romance languages, since most of the great editorial endeavours undertaken between 1800 and 1930–1950 have not been revised on the basis of modern principles.

By the end of the 20th century, philologists had developed a strong awareness of the authenticity of old textual witnesses, and editions increasingly tended towards the principles of diplomatic edition. One need only consider the recommended editions of Boccaccio (Branca 1999) and Dante (Petrocchi 1966–1967, still considered to be reliable) or other editions that have been published in the *La Pléiade* collection, such as Montaigne's *Essais* or Ronsard's *Amours*, all of which illustrate the importance placed on the concept of proximity to the original in recent times. Repercussions of the earlier ideological orientation nevertheless still make themselves felt in editions aimed at a broad, non-specialist audience, in which old texts are frequently altered and normalised unhesitatingly (a process that remains unacknowledged in the introductions).

Philological tradition and textual criticism are currently most actively pursued in Italy, where numerous high-quality editions have been produced in recent decades. For Spanish, Portuguese and Romanian, in contrast, language historians often encounter shortcomings that prevent an adequate treatment of fundamental subjects, in lexis as well as in grapho-phonetics and syntax.

Historical linguistics relies entirely on the editions produced by philologists over the course of the past two centuries. Following the transcription of the text, a critical apparatus is provided (in which errors of transmission are corrected and obscure passages are explained); the date and provenance of the text are determined (with varying precision), and finally, its position within the textual tradition to which it belongs and the historical context of its composition are discussed.

The study of the linguistic characteristics of specific texts and manuscripts within the context of editions is generally more problematic (cf. Duval/Guillot/Zinelli, *Les introductions linguistiques aux éditions de textes*, 2019). Romance studies have a long tradition of analysing grapho-phonetic, morphological and lexical features, but the requirements and methods of diachronic linguistics have evolved considerably over the last thirty years. The editors of (literary) texts are usually philologists with a literary rather than a linguistic background and they apply methods which are often considered by linguists to be outdated (cf. Chambon, *Lexicographie et philologie: réflexions sur les glossaires d'éditions de textes (français médiéval et préclassique, ancien occitan)*, 2006). Modern morphosyntax and syntactic approaches are almost completely absent from the linguistic descriptions that accompany text editions. This is no trivial observation, as a linguistic analysis founded on modern principles contributes to a better understanding of texts and only a text that has been properly understood can be edited correctly.

11.5.2 The contribution of textual genres to linguistic analysis

The patterns and specific characteristics displayed by the major textual genres and discourse traditions transcend the boundaries of individual languages. Of course, they are not restricted to the *Romània* and parallels are to be found in English as well as in Germanic, Slavonic, Semitic or Indo-Arian languages. It is nevertheless useful to consider Romance traditions as an entity in their own right, owing to the numerous phenomena of interdependence and their privileged relationship with Latin, both consequences of their genealogy.

When studying 12th- and 13th-century Occitan charters, for example, it is useful to compare these texts to charters written in Spanish, Italian and French. Similarly, administrative writing in the Ancien Régime in France could be compared to texts relating to administration in Spain or Hispanic America. The study of Italian poetic schools such as the *Scuola siciliana* and the *Dolce Stil Novo* requires a consideration of the influence of troubadour poetry and French *trouvère* lyric poetry. A scientific text such as the *Moamin* treatise on falcon medicine exists in Arabic and Latin, as well as in Italian, French and Spanish (*cf.* Glessgen, *Die Falkenheilkunde des 'Moamin'*, 1996); all these versions display similarities with regard to their structure, semantic content, and frequently even their lexical and syntactic features.

The suitability of the different genres for linguistic analysis depends on the domain of language to be observed: for studies focusing on grapho-phonetics and phonology, documentary sources are particularly useful as they are generally transmitted in original form – copies always result in significant and abrupt changes to the grapho-phonetic and morphological physiognomy of a text (*cf.* no. 4 below). Documentary texts also span a much larger geographical area and a larger chronological period than all other discourse traditions. The same applies to morphology, morphosyntax and basic sentence structure. For these domains, documentary texts allow an optimal description of the evolution of linguistic features as well as diatopic variation. In the case of lexis and of complex syntax, however, individual textual genres develop their own preferences. The syntax is particularly complex in texts of a legal character: charters combine elements characteristic of spoken language with a great deal of subordination. Literary texts, both secular and religious, mostly adhere to the unmarked syntax typical of narration. As for vocabulary, although a certain number of lexemes – generally high-frequency words – are found in most genres, every genre develops a vocabulary specific to a particular field of knowledge. This results in varying degrees of diaphasic variation. Generally speaking, therefore, the elaboration of the different textual genres contributed to the process of the elaboration of written languages. The Late Middle Ages thus saw the emergence of extensive lexical repertoires, which foreshadowed the specialised terminology that would appear in the 18th and 19th centuries. The emergence of literary, religious, legal, medico-biological, rhetorical, political and mathematic-astrological vocabularies enriched the traditions of regional writing, and, at the same time, lent them a pan-Romance dimension.

The above observations highlight the existence of a promising basis for research on the major processes of language evolution which occurred throughout the second millennium. Linguistic analysis should attempt to distinguish features that are characteristic of individual genres from those which pertain to all genres and therefore to language in general. Differences among textual genres can then be interpreted. In many cases the motivation behind the production of a text is to be sought in pragmatic objectives and thus in its communicative intent. The fact that a medical prescription and a popular tale serve completely different purposes has immediate consequences for the choice of vocabulary and the degree of syntactic complexity, as well as for the structure of the text. When such features differ significantly from a norm, they acquire a communicative value that surpasses their initial pragmatic objective. This is the way in which diaphasic and diastratic markers develop (*cf.* RSG 2, art. 170: *Prinzipien der Funktionalstilistik*).

11.5.3 The role of micro- and macroscopic description in historical linguistics

The crucial importance of the philological tradition lies in the fact that it provides the foundation for all interpretative work on language history, from Diez's etymological dictionary and comparative grammar of the Romance languages to the latest works and studies in the domain of diachronic linguistics (*cf.* 2.2.2 no. 3). Major dictionaries such as Godefroy, the FEW, Corominas, the LEI or the DERom used and continue to use the growing stock of textual editions for the identification and interpretation of the vocabulary of Romance languages in the context of their historical evolution, with increasing precision. The same applies to the major grammars by Renzi, Demonte and Buridant, as well as to all synthetic overviews of aspects of phonetics or morphosyntax.

Since the mid-19th century, however, a dynamic antagonism has developed between two opposing views on linguistic history: a synthetic view, which can draw from an ever-increasing amount of knowledge on the major evolutions of language, in opposition to a detailed view on individual sources and linguistic elements. The latter perspective concentrates on single texts, extracting data which enable an understanding of the role they play in the historical development of the language, as well as new information that helps expand general knowledge on language evolution. As the quality of synthetic works increases, on the other hand, research on texts becomes more precise. Thus, broad interpretations – even those founded on an imperfect basis – are necessary for detailed studies, which, in turn, improve general interpretations.

The antagonism between these two perspectives will never completely disappear since it is as impossible to rewrite a dictionary such as Corominas or the FEW every ten years as it is to revise a considerable quantity of editions and individual studies. Major differences between the two types of studies are inevitable, each of which is upheld by its own (editorial, lexicographical and grammatical) tradition. Their interdependence nevertheless determines all interpretations of language history.

We have shown that considerable work remains to be done in the area of linguistic historiography with regard to the study of older textual genres, with the exception of literary genres. In the same manner, the wealth of printed texts from the period between the 15th and the 19th centuries remains largely underexploited – again, with the exception of literary works. The vast quantity of texts and their apparent proximity to present-day language – in theory positive aspects which nevertheless contrast with the considerable difficulties encountered during their detailed study – have no doubt slowed the pace of research in this area. Only very recently did the historiography of language begin to include textual genres such as pamphlets, newspapers, travel accounts and treatises on specific areas of knowledge such as architecture, medicine or musicology. The lexical wealth found in travel literature, for instance, has been exploited by the *Deonomasticon Italicum* (cf. 9.7.1), which relies heavily on this type of source material. Finally, the compilation of extensive electronic textual databases has opened up new perspectives in this area of research (cf. 11.6), which, if continued, will result in the establishment of historical textual typologies founded on a linguistic approach.

→ Trotter, *Philologie éditoriale*, MRL

11.6 Corpus linguistics and philology in Romance studies

11.6.1 The general interest of corpus linguistics

In recent years, information technology has created new perspectives for the study of modern textual genres as well as for those from earlier periods. Computerised tools allow the management of large amounts of data and the quantification of linguistic characteristics. It is an area of research in the process of expansion and which is likely to undergo further significant development in the years to come.

Digital philology relies on the fundamental concepts underlying the practice of conventional philology, which consists of the compilation of corpora of defined texts as a basis for analysis. Decisions regarding the selection of texts to be included in a corpus are based on parameters of the diasystem (time, geographical space, textual genres), within which any sector can theoretically be considered as a legitimate object of study, be it an individual text, the work of a single author (e.g. the works of Rabelais, available in CD-ROM format, edited by Brunet, 1995), a large collection of texts reflecting a specific textual genre within a defined period (e.g. *Le Monde sur cd-rom depuis 1987*, which includes vast quantities of texts representative of the genre of present-day newspaper articles) or a selection of texts from different periods (e.g. the CORDE database of the Spanish Academy for Spanish texts from the Middle Ages until today).

The compilation of corpora – computerised as well as traditional – raises the problem of representativeness, which depends on several factors:

- the type of linguistic inquiry pursued: as an example, in order to study the functions of noun determination during a given period, it may be sufficient to consider relatively short excerpts of a dozen texts. Research on subordination would require longer excerpts from a more considerable diversity of textual genres. For lexicological research, finally, very large and diversified corpora are necessary in order to be able to identify low-frequency words;
- the textual genres available for a given period, their degree of elaboration and quantitative importance;
- the historical time frame chosen (it is possible to compare texts that are chronologically distant from each other or to focus on sources from adjacent periods).

The implementation of computerised tools allows a more effective use of quantification in corpus studies. It has proven particularly useful with regard to earlier historical periods, since written sources are more suited to electronic analysis than oral sources or sources that involve interaction between gesture, mimicry and language.

From a theoretical perspective, the attractiveness of IT-based corpus linguistics resides in the possibility of linking editorial philology, variational and text linguistics as well as pragmatics. A philological component is necessary in order to ensure the reliability of the sources with regard to transcription, dating and localisation, as well as the identification of their place within the textual tradition to which they belong. IT allows the analysis of a potentially unlimited number of manuscripts which exist for an individual work, as well as transcribed oral sources.

Distinguishing textual genres enables their comparison with respect to their role in language change. The possibilities for quantifying data offered by IT allow one to evaluate and measure existing linguistic differences among textual genres, which are first delimited with the aid of extralinguistic parameters. It can be assumed that large corpora, structured by genre and thus according to different contexts of language use, represent a series of socio-cultural and cognitive ensembles that play a part in the evolution of language. Consequently, it should be possible to determine on a non-intuitive basis whether, at a given time in history, specific linguistic forms are associated with defined contexts, if not genres, or whether their use is more general.

From an empirical point of view, we are nevertheless still far from having exploited such potential. Working with computer tools requires prior training and preliminary reasoning in order to avoid the waste of physical and intellectual energy. In practice, the gulf between the disciplines of information science and linguistics further adds to these difficulties.

Computerised management of linguistic data involves various elements: the principles and tools of encoding (*cf.* no. 2 below), programming and tools for linguistic analysis (no. 3) and the use of existing databases incorporating various tools (no. 4), this last being more widespread than the first two.

11.6.2 The processing of medieval textual data

In digital word processing, it is necessary to distinguish the definition of characters (i.e. linguistic signs, above all alphabetical) from that of the layout, which represents meta-structures (i.e. page structure, titles, paragraphs, characters in italics or bold, font and size of characters, etc.): the most elaborate form for the definition of characters is provided by Unicode, which takes into account approximately 65,000 (2¹⁶) different symbols⁴⁹. Thanks to this universal standard, it is possible to ensure the recognition of these symbols by different computer systems. Information concerning the layout can be entered using a type of neutral SGML (Standard General Markup Language), especially in the form of HTML (used for webpages) or XML, the most elegant and the most established standard markup language in use today.

The following illustrates the different stages of elaboration and processing of the data contained in old texts within the theoretical framework explained above (cf. 11.3). The editorial work consists of five stages:

1. the obtention of the photographic reproduction of a medieval manuscript (an episcopal charter from Lorraine, dated 1239, will serve as an example);
2. the transcription and digital processing of the text (the example shows a faithful diplomatic transcription and encoding using XML);
3. the displaying of the coded version of the text in a browser or in the form of a PDF file for export (= the actual edition: this is generally the only version visible to the user);
4. as explained above (11.3), a Modern French translation of the text in order to ensure the comprehension of the document;
5. a historical commentary on the context in which the document was produced.

These five stages must be completed before a linguistic analysis of the grapho-phonetic, grammatical or lexical aspects of the language used in the document can be undertaken.

For the benefit of the reader, these stages will be presented in a more traditional order (photograph, historical commentary, user-friendly version, translation, XML encoding):

⁴⁹ In its most basic form the definition of characters uses an ASCII/ANSI-type pure text encoding. Specifically, the accepted encoding of characters corresponds to a series of eight bits; eight bits equate to a byte or octet (i.e. eight times the opposing pair 1–0; e.g. the computer encoding for letter ‘a’ corresponds to a sequence of numbers such as “01100001”; a capital ‘A’, to “01000001”, and so on). With eight bits, it is possible to differentiate 256 characters (which is equivalent to 2⁸); initially, this was the capacity of North American typewriters, which was perpetuated in the form of the ASCII/ANSI repertoire.

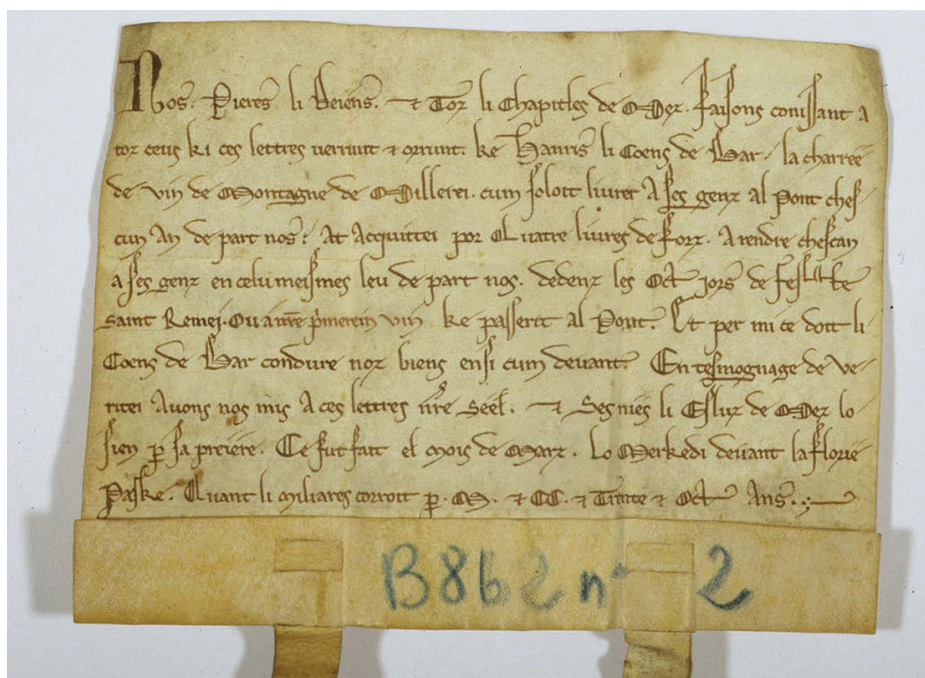


Fig. 34: Facsimile of a charter from Lorraine, dated 1239

Source: The episcopal charter in question is conserved in the *Archives Départementales* (A.D.) of the French department Meurthe-et-Moselle.

HISTORICAL COMMENTARY: By means of this charter, Count Henri of Bar has granted the right of way to the chapter of the Cathedral of Metz for the delivery of wine, in exchange for a payment of four *livres*, which equates to the contents of a cart of wine. The somewhat convoluted description of the facts emphasises that the count has accepted this exchange.

(SEMI-)DIPLOMATIC EDITION:

- 1 1 Nos · Pieres · li deiens · et toz li chapitles de Mez · 2 faisons conissant à
- 2 toz ceus ki ces lettres verrunt et orrunt · 3 ke Hanris li coens de Bar · la charree
- 3 de vin de montagne de Millerai · c'um soloit livrer à ses genz al Pont ches-
- 4 cun an de part nos · 4 at acquittei por quatre livres de forz · à rendre chesc'an
- 5 à ses genz en celu meismes leu de part nos · 5 dedenz les oct jors de feste
- 6 saint Remei · ou à notre primerein vin ke passerit al Pont 6 Et per mi ce doit li
- 7 coens de Bar condure noz biens ensi cum devant · 7 En tesmognage de ve-
- 8 ritei avons nos mis à ces lettres notre seel · et ses niés, li esluz de Mez, lo
- 9 sien, par sa preiere · 8 Ce fu fait el mois de marz · lo merkedi devant la florie
- 10 Paske · quant li miliares corroit par · m · et .cc · et trente et oct ans ·

MODERN FRENCH TRANSLATION: 1 Nous, Pierre, doyen du chapitre cathédral de Metz ainsi que tout le chapitre, 2 faisons savoir à tous ceux qui verront et entendront cette charte 3 qu'Henri comte de Bar [nous a libéré] de la redevance d'une charge de chariot de vin, provenant de la montagne de Milleray, que nous avons l'habitude de fournir tous les ans à ses sujets à Pont-à-Mousson 4 en échange d'un paiement de quatre livres de monnaie forte que nous donnerons chaque année à ses sujets en ce même lieu, 5 dans les huit jours après la fête de saint Rémi ou lors de notre premier passage de vin à travers Pont-à-Mousson [= the village and the bridge]. 6 Et en échange, le comte de Bar doit nous garantir le passage comme auparavant. 7 En témoignage de vérité nous avons apposé notre sceau à cette charte, et le neveu du comte, évêque de Metz, a apposé le sien à la demande du comte. 8 Ce fut fait au mois de mars, le mercredi avant le dimanche des Rameaux, dans l'année du seigneur 1238 [= 1239 new style, as the year changed on the 25th of March].

This example illustrates the semi-diplomatic editing principles elaborated for the *Documents linguistiques galloromans*, the corpus to which the charter depicted above belongs (cf. <https://gallrom/linguistik.uzh.ch>). Concerning the structural elements:

- Words have been separated according to modern orthography, but all editorial interventions are clearly visible: based on the transcription *c'um* [3], the layout programme can either reconstitute the form present in the manuscript (*cum*) or suggest an interpretative reading (*c'um*).
- The insertion of apostrophes and accents follows the same model: these symbols were added by the editor and, if desired, can be suppressed in the electronic edition (resulting in a diplomatic version).
- Bold type is used to indicate majuscules (i.e. capital letters, in cases where these can be identified) in the manuscript; abbreviations are expanded in italics. These features are coded explicitly in the XML version (cf. below for the tags <maj> </maj> [Nos, 1] and <abr> </abr> [et, 1]). All remaining (non-bold) upper-case letters have been introduced by the editor. Here too, the computerised version enables the identification both of forms present in the manuscript and interventions by the editor.
- Medieval punctuation (so-called *interpuncts*, situated at mid-line level) as well as conventional modern punctuation (as in *ses niés, li esluz de Mez, ...* [7]) are differentiated; this double encoding (cf. below: ,/.) gives the editor the choice of suppressing either the medieval or the modern form (as in the example provided here).
- The text is segmented according to semantic and syntactic parameters and paragraphs in order to facilitate comprehension and to provide an unequivocal means of referring to different sections of the text (cf. the example <div n="1"> ... </div>).

The XML encoding is established using tag pairs that structure the pure text using ASCII symbols alone. In the example provided, the edited text is entered between the tags <txt> and </txt> and structured according to semantic units (<div n= "1", etc.) and para-

graphs (<par/>). This abstract form of encoding can then be interpreted by layout programmes in different ways, for either an Internet-based representation or a printed version. The beginning of the text exemplified above is given in its XML form here:

```
<txt>
  <div n="1">
    <maj>N</maj>os,//. <maj>P</maj>ieres,//. li <maj>d</maj>eiens /. <abr>et</abr>
    <maj>t</maj>oz li <maj>c</maj>hapitles de <maj>M</maj>ez,//.
  </div>
  <div n="2">
    faisons conissant a <zw/> toz ceus ki ces lettres verrunt <abr>et</abr> orrunt /.
  </div>
  <div n="3">
    ke <maj>H</maj>anris, li <maj>c</maj>oens de <maj>B</maj>ar,//. la charree <zw/> de
    vin de <maj>m</maj>ontagne de <maj>M</maj>illerei,//. c'um soloit livrer a ses genz al
    <maj>P</maj>ont ches<zwt/>cun an de part nos,//.
  </div>
  <div n="4"> (...)
  <par/>
  (...)
</txt>
```

This encoding allows computerised analysis of the text, either individually or as part of a large electronic corpus. Due to the explicit structure of the encoding, the edition meets the standard of precision that is indispensable for rigorous linguistic analysis. The XML encoding also has the essential advantage of ensuring the longevity of the data. Before the development of these standards, hundreds of millions of hours of work were lost, as coded linguistic data became illegible after only a few years.

→ Glessgen, *L'écrit documentaire et le projet des Plus anciens documents linguistiques de la France*, in MRL 4, art. 11

11.6.3 Programming tools and tools for linguistic analysis

Once data have been entered, the linguist's work has only just begun. Information technology allows texts to be easily converted into lists of individual words, which can then be organised alphabetically in order to establish concordances, and quantified in order to generate frequency lists. Concordance lists with words in their immediate context (*cf.* the common format KWIC = key word in context) are useful for lexical analysis and also enable the lemmatisation of forms (i.e. the grouping together of the inflected forms of lexemes and grapho-phonetic variants) and the identification of collocations

and phraseologisms. The comparison of words in different contexts allows statistical procedures to be applied in order to determine the linguistic differences between the various uses. With the aid of reference dictionaries, word classes can be identified automatically (and tagged), though the distinction of constituents (by means of *parsing*) is more difficult.

These tasks can be conducted by implementing tools and techniques from a large pool of available software. Unix commands, the precursor in this area, allow the user to organise the forms of a text into lists and to compare them. Programming languages such as *Python* or *AWK* allow more complex queries. In recent decades, a large number of programmes for computing concordances, automatic lemmatisation and part-of-speech tagging, as well as other sophisticated textual analysis tools have been developed. Many of these are freely available and can be applied instantly to a broad range of textual data.

Regrettably, this domain suffers from a marked absence of critical evaluation, making orientation difficult. We are still a long way from having established standards and reference tools that enable far-reaching investigations to be performed on historical corpora. The next few years will undoubtedly bring further clarification to this domain, thanks above all to the principles of *Open Source*, which has introduced the scientific criterion of transparency into the field of information science.

→ cf. Pusch/Kabatek/Raible, *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft*, 2005

RK 34, *Digitale romanistische Sprachwissenschaft*, 2023

11.6.4 Principal databases

Most linguists are familiar with corpus linguistics in the form of databases that have already been compiled by other researchers and that consequently offer ready-to-use tools for linguistic investigation. At present, countless collections of texts can be accessed via the Internet. They are often created for the purpose of lexicological analysis, but an increasing number of research tools for graphemic, morphological and even syntactic and lexical-semantic features are being developed. This, too, is a field that suffers from a lack of systematic critical evaluation with regard to the quality of textual sources and tools for analysis. The contribution made by this discipline to linguistic study is nevertheless apparent.

For major Romance national languages such as French, Spanish and Italian, there are reference databases that include a large number of texts of various provenance (in terms of historical period and geographical location) and representing different textual genres. The following brief presentation of a selection of major projects also illustrates the constraints that these databases necessarily encounter:

- *Frantext* is the oldest reference database for Romance texts, located in a research centre in Nancy (formerly known as INaLF; ATILF since 2001). It contains French texts, mostly dating from between 1789 and 1970, although the whole period from the 16th to the 21st century is represented to some extent. Literary texts are dominant in this corpus, which nevertheless also contains scientific works, above all from the fields of the hard sciences.

Frantext served as a basis for the *Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècle*, and in its wake, a database of Middle French texts (1359–1500) was compiled as a basis for the *Dictionnaire du moyen français* (cf. 9.9.1).

The textual database *Frantext* was first stored in punched tape format, then on punched cards, before being transferred to digitised forms of support. The change in technology thus meant that many years' worth of typing went to waste (cf. Radermacher, *Le Trésor de la langue française: une analyse lexicographique*, 2004). During the compilation of the database, little thought was given to philological precision, once again for reasons linked to the history of the discipline. The software *stella*, employed for queries in *Frantext*, offers a wide range of functions, although effort is required in order to learn how to use it effectively.

- For the medieval period, and for documentary genres in particular, the database *GallRom* (<https://gallrom/linguistik.uzh.ch>) provides a model founded on up-to-date methodology. The database contains the earliest, for the most part original documentary texts from almost the whole of the Gallo-Romance territory, edited according to (semi-)diplomatic principles. The corpus is linked with the lexicological and bibliographical data contained in the major dictionaries for the medieval period: the DEAF, the DAG and the DAO (cf. ch. 9.9.1).
- The abandoned textual database of the *Hispanic Seminary of Medieval Studies* in Madison: a large number of diplomatic transcriptions of manuscripts in Old Spanish were initiated (accessible on microfiche) with the aim of preparing a *Dictionary of the Old Spanish Language*, which, however, was never published (cf. the incomplete *Diccionario español de textos médicos antiguos*, DETEMA, published in Spain in 1996). This is another example of the possible pitfalls of computer linguistics.
- More recently, the *Corpus diacrónico del español* (CORDE) and the *Corpus de referencia del español actual* (CREA) have been compiled by the Real Academia Española and are now accessible via the webpage of this institution. The two databases are integrated within the *Corpus del Diccionario histórico del español* (CDH). They are continually being expanded and have thus become a new point of reference for the analysis (above all historical) of Spanish (cf. 9.9.2).
- The Italian counterpart, the *Opera del Vocabolario Italiano* (OVI, Pisa) groups texts dating from before Boccaccio's death (in 1375), and is accompanied by a lexicographical project, the *Tesoro della lingua italiana delle origini* (TLIO, cf. 9.9.2). The OVI database covers only a relatively short historical period, since Italian texts earlier than 1300 are rare. It also concentrates on texts from the region of Tuscany

and – to a lesser extent – neighbouring regions (including Venice, but excluding the whole of Southern Italy). The query software *Gatto* is very efficient for the lexicological treatment of data, but, as in the case of *stella*, it is not available via open access.

The OVI/TLIO project is nevertheless the most successful of the historical corpora available for the Romance languages. High philological standards govern the choice of the editions upon which it is based, as well as the dating and localisation provided for the texts and manuscripts.

11.6.5 Perspectives

Quantification methods are on the verge of transforming linguistic research and linguistic history, just as they transformed the fields of sociological and psychological research several decades ago. This methodological revolution will certainly require a considerable amount of effort on the part of linguists, as it is in their best interests to familiarise themselves with these methods.

When advising students on possible avenues of research, one should emphasise the importance of understanding the different levels of textual data management (encoding, query tools and software for interpreting data). Furthermore, Open Source and freeware tools not only have the advantage of being freely available but also offer transparency as regards their internal architecture.

Once the many challenges inherent in the computerised management of linguistic data have been mastered, this field will open truly promising perspectives. It will be capable of bridging the gap that still exists between editorial philology and the systematic analysis of language. Quantificational analysis of the specific contexts in which linguistic forms appear, as well as the consideration of textual genres and diasystematic varieties, will lead to a better understanding of language configuration. It will be possible to found historical linguistics more directly on utterances and on the creative activity of *parole*.

In a perspective such as this, it should be possible to combine the traditional methodology of editorial philology with the recent techniques of corpus linguistics. Although this combination of old and new methods is extremely time-consuming, it produces excellent results and opens up many new and unsuspected horizons.

12 From theory to practical work

When first encountering linguistics, students are typically confronted with abstract rules: grammars, dictionaries and introductory textbooks such as the present one describe and analyse linguistic systems; works focusing on language learning describe phonological systems and grammatical rules, and provide lists of vocabulary. Nevertheless, these are all forms of interpretation that are ultimately based on the study of actual manifestations of language, which constitute the daily bread of the linguist. Linguists study concrete realisations of language in order to deduce abstract, and occasionally general, rules. They are thus obliged to compile corpora consisting of examples, in order to investigate their significance or to attempt to associate them with a specific linguistic variety. The work of the linguist therefore consists in the detailed study of source material, both written and oral.

Research methodology in (Romance) linguistics, then, presumes specific competences and familiarity with a variety of techniques, which have been partially described in the previous chapters:

1. A solid knowledge of linguistic theory, the internal and external features of the languages studied and the history of the language family is a fundamental prerequisite and point of orientation. This knowledge has been provided to a large extent in the preceding chapters of this *Companion*.
2. The ability to compile a bibliography of the most significant studies relating to the subject matter (*cf.* 2.4) and to interpret previous studies in the context of the research tradition to which they belong (*cf.* ch. 2).
3. Data must be collected from oral or written sources. The study of oral sources requires access to recordings or phonetic transcriptions of different varieties (dialectal, colloquial and popular, but also standard), which are then exploited in combination with data from other existing corpora, which, however, must first be subjected to critical evaluation. For written sources, especially early texts, previous editions must be consulted and the quality of these must be assessed; it may even be necessary to transcribe and edit manuscripts or printed texts (*cf.* 10.7).

In both cases, textual data must be prepared according to established methodology in order to ensure that accurate linguistic data can be extracted (such as the forms of the French subjunctive that appear in a particular poem or the interjections used in an oral corpus).

4. Depending on the area of research, specialised methodological knowledge is essential: in historical phonetics, familiarity with the concepts of relative chronology is indispensable; the study of etymology requires the knowledge of the underlying theoretical principles, a good grasp of the methods of comparative grammar and reconstruction as well as knowledge of historical lexicography; for the pursuit of

dialectology, it is essential to have a sound knowledge of etymology and historical phonetics; it can also be useful to understand some of the highly mathematical techniques of dialectometry, while for corpus linguistics, knowledge in the field of digital humanities is required.

5. Finally, studies must be conducted according to the basic principles of scientific research (reliability, transparency and, depending on the case, reproducibility of the results), and of logical presentation (clarity of argumentation, absence of redundancy). Furthermore, it is necessary to consider source criticism (i.e. the critical evaluation of source material) as well as the work of other linguists with regard to their reliability and coherence.

It is essential to keep these aspects in mind during practical work, particularly as regards the preparation of linguistic data and the criticism of sources, which are often complex. At the same time, it is essential to be aware of possible limitations which may affect the formation of a coherent theory, as well as of major differences in linguistic theory and practice in different countries or from different ‘schools’.

12.1 Romance studies in theory, practice and teaching

12.1.1 The development of linguistic theory

The brief historical overview of major trends in linguistic research in the first part of this book and the great variety of approaches presented in the ensuing chapters illustrate the profound changes that took place within this field of specialised knowledge. A secondary effect of the ongoing theoretical development of linguistics is that its external delimitations have become as blurred as the boundaries between its internal domains. Although different approaches show consensus in some aspects, individual researchers tend to put forward their own ideas on external and internal delimitations, and specialists in one particular area may be unaware of the work of others, despite their potential significance. Linguistics thus currently resembles an intellectual culture, rather than a discipline founded on well-established and generally admitted theoretical principles.

Linguistics and, more importantly, Romance studies have not developed a tradition of epistemological thought that enables a structuring of the theories and methods to which they have recourse. Linguistics consequently differs from other academic disciplines such as mathematics and physics, law and philosophy or even history, all of which are based on well-founded epistemological traditions. It is difficult to say why linguistics has not taken this step – especially since it is among the most abstract disciplines in the field of the humanities. The combined influence of several factors is undoubtedly responsible for this situation: one of these is the great diversity of theories and methods and their fluctuations, which have increased considerably since the 1960s; another is the importance of empiricism, which mobilises a great deal of resources in

terms of field work and observation, data collection, the study of large quantities of old texts and even foreign language learning, which is sometimes inevitable. In addition to these aspects, there are two even more fundamental reasons for the relatively weak epistemological dimension of linguistics:

1. One can assume that, until the very end of the 20th century, linguistics did not yet possess the adequate means for the development of coherent theoretical principles. The development of creole studies in the 1970s and the cognitive turn in the 1980s were certainly a prerequisite in this respect, as they placed previous knowledge from diverse traditions of research under the auspices of an anthropological interpretation.
2. The low applicability that distinguishes linguistics from other disciplines such as law, physics or medicine has no doubt limited the interest in the development of general theoretical principles. The most direct inputs of linguistics are its contributions to various state-associated institutions, including the education of future teachers, mostly at secondary school level. Close links with the educational systems of individual countries (and thus distinct civilisations), in turn, work against the formation of internationally shared methodologies.

It is likely that linguistics would be able to contribute more to modern-day society than it does currently, in terms of its interpretative as well as its practical applications, if researchers could agree on a shared theoretical basis. It goes without saying that the author's own attempt at structuring the subject matter in question does not claim to be founded on satisfactory epistemological rigour. The present *Companion* attempts to reach a balance between the actual data constituting the different Romance varieties as such and the specific phonetic, grammatical and lexical phenomena which characterise their evolution from Latin on the one hand, and on the other, the methods used to interpret these phenomena.

12.1.2 Romance studies in academia (1): non-Romance-speaking countries

Until now, our presentation has followed the assumption that there is an international community of researchers focusing their efforts in a coordinated and rational manner in order to elaborate a diversified methodology for linguistics. Such an assumption is necessary in order to build a coherent interpretation of the matter; however, it does not reflect reality. Rather, in so far as linguistics complies with the organisational needs of political institutions in different countries, it takes the form of individual scientific microcosms that vary greatly.

This observation holds true for developments from the time of the major linguistic innovations at the beginning of the 19th century, which led to the professionalisation of scholars in this field. Thereafter, academic structures played a determining role,

since the concrete practice of a discipline is closely linked to the establishment and the number of both professorial chairs and other university positions in individual universities and countries, as well as to the specific organisation of academic research. It also depends on job prospects for students, the curriculum and the degree programmes offered (e.g. BA, MA and PhD according to the current system).

In all these areas, differences among countries – or, more specifically, among countries of the Western World – are very marked. The specific case of Romance studies additionally involves differences between Romance-speaking countries and German-, English-, Scandinavian- or Slavonic-speaking countries. Romance studies is an ‘invention’ associated with German-speaking countries, and, objectively speaking, it enjoys a higher profile in non-Romance-speaking societies than in Romance-speaking countries – the latter typically being more interested in their own respective languages.

But even in terms of the teaching practices characteristic of Romance studies in German-speaking universities, some surprises await the present-day observer. The 19th century witnessed two major successive splits: the first separated the teaching of Classical languages (Latin and Greek) from that of modern languages, while the second organised modern language teaching into three major groups – German (and other Germanic languages), English and the Romance languages. This third group included language teaching, French linguistics and literature (especially Medieval French), as well as Old Occitan, some Italian and, occasionally, Spanish, with all other Romance languages being only marginally represented.

The number of universities offering a programme of Romance studies that extended beyond French and Occitan increased gradually, beginning with the universities of Bonn (with Friedrich Diez in 1854 and, more particularly, his successor Wendelin Foerster in 1876), Berlin (1876) and Strasbourg (1872; cf. Christmann, *Romanistik und Anglistik an der deutschen Universität im 19. Jahrhundert*, 1985). This limited programme was maintained until the mid-20th century. Moreover, its implementation faced further limitations: full professors were responsible for teaching both literature and linguistics, even if their principal research interests lay in only one of the two domains.

Furthermore, their grasp of the spoken modern forms of the languages they taught was often limited, even though in some cases they achieved extraordinary levels of competence in the study of the early written stages of these languages. Students of ‘Romance’ learned Latin, Greek and French (mostly Old French); in addition, they studied texts in the *oïl* and *oc* dialects. If required, they familiarised themselves with Old Occitan, Italian, Spanish and Portuguese, in the interests of general knowledge and in order to better understand phenomena of sound change, which occupied a place of undisputed importance in their studies. This curriculum had little in common with Romance studies today.

The professional goal of students of ‘Romance studies’ between 1850 and 1950 was almost exclusively to become teachers at secondary level (middle or high school); this was correlated with the fact that the university system was established throughout

Europe in the 19th century with the objective of equipping the newly formed nations with officials who had benefitted from a specialised education.

It was not until after World War II that (i) in Germany, Austria and German-speaking Switzerland, the chairs of Romance studies were separated into linguistics and literary studies, (ii) high-quality teaching of modern languages began, and (iii) Italian and Spanish as well as, to a lesser extent, Portuguese, Catalan, Romanian, and Romansh (in Switzerland) began to be more systematically included. The diversification of career prospects for students, for whom teaching in high schools was no longer the only goal, contributed to the diversification of Romance studies as taught at university level.

These main tendencies in the evolution of Romance studies in non-Romance-speaking countries are similar in Britain, North America, Scandinavian countries (*cf. Table ronde, Actes du XXIX^e Congrès international de linguistique et de philologie romanes*, vol. 1, 2021, pp. 103–134) and in central-eastern Europe (especially Poland, Hungary and the Czech Republic). The Iron Curtain nevertheless imposed serious restrictions on Romance studies and on linguistic research in general. In each of these countries, the number of chairs, as well as the number of students of one or more Romance languages remained limited compared with those in German-speaking countries.

Towards the beginning of the 21st century, non-Romance traditions of linguistic studies still retain a character of their own, even though increasingly widespread Internet communication has undoubtedly allowed a closer glimpse into the traditions of the Romance-speaking countries.

12.1.3 Romance studies in academia (2): Romance-speaking countries

In Romance-speaking countries, the respective mother tongue has always been the main centre of interest. In the third decade of the 21st century, the subject matter taught in France, Belgium and French-speaking Switzerland focuses on Modern French (from the 16th century onwards) and its diasystematic varieties. Medieval French is still taught, while *oïl* dialects and even Modern or Old Occitan, Gascon or Francoprovençal have become peripheral and only tend to be taught at a regional level. A better balance can be observed in Italy, where a triad consisting of the modern, medieval and dialectal varieties of Italo-Romance are taught, and where a strong presence of Latin has been maintained. In Catalonia, Valencia and Galicia, Catalan, Valencian and Gallego are studied alongside Spanish, while in the other parts of Spain the latter is dominant, as is Portuguese in Portugal. In Romania, in contrast, the study of Romanian and its history and folklore takes place within a more general Romance context – perhaps to emphasise its distance from the genetically unrelated Southern Slavonic languages.

Quebec follows the same course of tradition as the United States, though it has incorporated European traditions to a higher degree with regard to research on both French and the Ibero-Romance languages. The situation is more modest in the His-

panic countries of the Americas and in Brazil: here, developments long remained tied up with those taking place on the Iberian Peninsula, the latter under the deleterious influence of Francoism. Furthermore, political developments over recent decades have not favoured academic institutions in these countries. In the areas of geo- and socio-linguistic variation and in the history of Ibero-American varieties, contributions from Hispanic America are nevertheless noteworthy.

In the other countries of the world, the discipline of Romance studies has no basis for existence. The number of university chairs that can be closely or distantly associated with Romance studies are limited in Africa, Asia and Australia, although there are a number of universities in unexpected places that participate in Romance academic discourse at an international level (e.g. in Japan, Algeria or New Zealand).

Our observations will therefore be restricted to France and Belgium, Italy, Spain and Switzerland, which – among the Romance-speaking countries – have made the most significant contributions to linguistic research. In France, Italy and Spain, the teaching of their respective national languages was established at the end of the 19th and the beginning of the 20th century, within the framework of the study of the classical languages. Still today, Latin is often an integral part of the curriculum of Romance studies, and, in turn, courses on the various Romance languages enrich the curriculum of classical philology. Moreover, linguistics and literary studies have always been more closely linked in these countries than in non-Romance-speaking countries – in some cases to the detriment of linguistics, which has often been considered to be an auxiliary science of literary studies.

The concept of ‘Romance’ studies has never been particularly explicit in these Romance-speaking countries; however, as we have seen, major advances in linguistics up until the 1930s took place under the overarching influence of historical-comparative – and thus genealogical – approaches to language. Reasoning on an evolutionary and etymological basis, sometimes including dialectal data, long provided an equilibrium in studies on French in France or Germany, on Spanish in Spain and on Italian in Italy. More particularly, universities with an interest in minority languages have always asserted their ties to Romance studies; this is the case for Occitan or Gascon in southern France (Bordeaux, Toulouse, Montpellier), for Catalan and Valencian in Catalonia and Valencia (Barcelona, Valencia), for Francoprovençal and Romansh in Switzerland (Zurich, Neuchâtel, Fribourg), for Ladin in South Tyrol (Bolzano) and for the Walloon dialect of French in Belgium (Liège, Namur, Brussels, Louvain).

In Italy, a strong philological tradition led to the development of a vigorous *Filologia romanza* from the 1880s onwards, concentrating on the philological study of medieval texts, above all Italian and Occitan. More generally, Italy is without doubt the country that produces the greatest number of exemplary achievements within the realm of Romance studies today; however, these studies typically concentrate on Italo-Romance, although work is also carried out in the fields of Old French and Old Occitan.

In Spain, interest in Romance studies was catalysed by the rise of Catalan and Galician in the second half of the 20th century; however, *filología románica* remained rather

poorly represented compared with the three national philologies which had by then been established (Spanish, Catalan and Galician). From a methodological point of view, moreover, Francoist and post-Francoist Spain were strongly influenced by research in France, where Romance studies never achieved particular importance.

The case of France is most difficult to interpret: between 1870 and the 1930s, it was intensively involved in the development of modern linguistics and the Romance paradigm, to which it made valuable contributions. By the 1980s at the latest, however, the country had broken away from the mainstream of Romance studies and other countries followed suit: methodological developments in international Romance studies subsequently became almost unknown to specialised linguists, and thus even less so within the general framework of university studies in France. This has been further reinforced by the fact that the examinations in linguistics which form part of the State entrance exams for teachers (CAPES, *Agrégation*) represent an extremely limited part of the discipline and rely on long outdated methodology.

Belgium only partially follows the academic trends of France, as it has maintained a stronger tradition of Romance studies as well as an interest in the Walloon dialect. Switzerland is a special case in so far as it combines the traditions of German-speaking countries with those of Romance-speaking countries to an even greater extent, thanks to its four national languages. Despite the small size of the country, its contributions to Romance studies throughout the 20th century were exceptional: among the pioneers of the discipline were Swiss scholars Meyer-Lübke, Gilliéron, Saussure and Wartburg. After the disastrous years of the Nazi regime in Germany were over, a number of excellent Swiss Romance scholars went to Germany to teach Romance studies, including Johannes Hubschmid, Kurt Baldinger, Kurt Ringger, Max Pfister and Peter Wunderli, who were for the most part students of Walther von Wartburg. Still today, Swiss universities offer students a balanced linguistic curriculum, which covers most of the Romance paradigm.

12.1.4 Factors contributing to cohesion in Romance studies

Despite a great deal of divergence as regards academic practice in different countries, Romance linguistics does exhibit some degree of cohesion, which relies on instances of scientific organisation that surpass national boundaries and the limits of educational and research institutions. These include ongoing bibliographies listing scientific publications, scientific journals, introductory manuals and encyclopedias as well as international conferences, beginning with the triennial conference in Romance linguistics and philology organised by the *Société de linguistique romane* (SLR).

Of the most important traditional journals of Romance studies, the international *Revue de linguistique romane* (founded in 1925 by the SLR and based in France) is to be found in the great majority of universities that offer studies in one of the Romance languages. The distribution of the *Zeitschrift für romanische Philologie* (Germany), *Vox*

Romanica (Switzerland), *Estudis Romànics* (Catalonia), *Medioevo Romanzo* (Italy) the *Romanistisches Jahrbuch* (Germany), the *Revue roumaine de linguistique* (Romania) or *Romance Philology* (U.S.) is more disparate, but these periodicals are nonetheless found in the university libraries of a large part of the Western world (cf. 2.4). Recent years have also seen the development of Open Access publishing in the field, as well as pure OA journals such as *Isogloss. Open Journal of Romance Linguistics*, founded in 2015 (Barcelona).

Introductory manuals have always enjoyed a wide distribution, including among students; this is less true for the various encyclopedias, due to their high cost.

All these international entities reduce the centrifugal tendencies that operate at the national level, and they allow university teachers to remain up to date on methodological choices made by colleagues in other countries. Therefore, despite the great diversity inherent in teaching curricula, it is thus legitimate to assume the current existence of a Romance paradigm, even if it is not always particularly apparent.

- Glessgen, *Les Manuels de linguistique romane*, 2000
- Glessgen/Giolitto, *Los vectores de la romanística a través del tiempo*, 2007
- RSG 1, art. 38–40 (= journals, bibliographies and congresses in Romance linguistics)

12.2 Linguistics, society and politics

12.2.1 The intrinsic dangers of linguistics

The limited degree of immediate applicability of linguistic research leaves a great deal of freedom in the choice of scientific objectives. In comparison, the activity of researchers in medical science tends to be incomparably more focused owing to the frequency of certain diseases and their distribution across the world. The main risk in linguistic work is thus that of losing track of its defining objectives. As a brief reminder, the ultimate aim of linguistics is to contribute to mankind and to society, if only in terms of gaining a better understanding of our identity and past and of furthering an understanding of and respect for diversity.

Due to the complexity of the subject matter and the discipline, however, a near-complete detachment from these objectives is often necessary in order to arrive at a more thorough understanding of the specific features of language. Language thus becomes an object of study in its own right. Since their achievements have only a low immediate impact on society, it is not easy for linguists to recognise the point at which they have left the realm of the reasonable. At the same time, this also hints at the considerable effort required to convince non-specialists of the usefulness of specialised linguistic arguments.

Linguistics can therefore be defined as a fundamental discipline in the true meaning of the word. Its powerful interpretative force is coupled with a low potential

for immediate application. Consequently, there are very few professionals working in linguistics outside academia. Its study, on the other hand, is of potential value to many professions and fields of activity, including teaching and the media in its entirety, as well as politics, economy, sociology, computer sciences and psychology.

As is true of all other scientific fields, linguists must navigate the tension between the usefulness of the discipline to society and the necessity for freedom of action.

12.2.2 The influence of external factors on linguistics

Among other aspects, the chapter on external history emphasised the fact that the development of linguistic thought has always gone hand-in-hand with other socio-cultural and, above all, political developments. We have mentioned the extent to which nations are involved in shaping academic organisation, and thus also in determining the focus of university disciplines. Like history, sociology or psychology, linguistics is involved in the elaboration and the establishment of political, national or regional identities.

The study of medieval texts during the 19th century served to highlight ‘national antiquities’, and studies in dialectology emphasised ‘national folklore’; both tendencies participated in the shaping of a heritage in terms of identity that was indispensable for modern nations during the course of their establishment. The writing of major ‘histories of the French, Spanish or Italian language’ as well as historical dictionaries and grammars created references for a historiography of the respective languages and, in turn, bolstered the countries that funded the universities in which they were taught.

In German-speaking countries, studies on the *România* at first served as a methodological term of comparison and as a pretext for studies on the national language. It was no coincidence that Romance studies adhered to the framework of Indo-European studies, which at the time were referred to as ‘Indo-Germanic’ studies.

Established scientific trends are less prone to political influence, since they have partially developed a powerful dynamic of their own, responding to neutral objectives and aiming at contributing to a better understanding of the natural world. Nonetheless, it would be a mistake to underestimate the significance of incidental external factors in determining the establishment of scientific trends; in the end, the development of science depends on individuals whose role in academia has strong sociological implications (relating to questions of hierarchy, the focus of research, or choices regarding curricula).

→ Metzeltin, *Sprachgeschichtsschreibung: Möglichkeiten und Grenzen*, RSG 1, art. 2

Seidl, *Les langues romanes dans l'historiographie des langues indo-européennes*, RSG 1, art. 42

12.2.3 Ideological dependence and scientific (in)dependence

More concrete dangers for work in linguistics emerge where it is exploited for objectives that carry strong implications for issues of identity and norms. The negative effects of political motives underlying linguistic research are most immediate in repressive political contexts. This was apparent at the end of the 19th century in the desire to construct homogeneous and standardised national languages that were shared by all citizens of a nation, and in the attempt to situate the emergence of these languages as far back in history as possible (cf. 10.4.5).

In a more aggressive form, Fascist and totalitarian regimes made use of linguistics for their own benefit, rewriting the history of language and society. Another compelling and archetypal example is that of research on the Germanic *superstrata* (cf. 10.3.3) – which includes assertions that are as baffling as the overall quality of the work of researchers such as Wartburg or Gamillscheg is exceptional (cf. the nuanced studies by Hausmann on the role and the conduct of philologists in Fascist Germany, especially “*Vom Strudel der Ereignisse verschlungen*”: *deutsche Romanistik im ‘Dritten Reich’*, 2000). Examples of linguistic research and teachings influenced by ideologies are numerous, not only under the Nazi regime, but also in Francoist Spain or in Romania under the dictatorship of Ceaușescu.

Even in the most democratic societies of the second half of the 20th century, however, significant trends of an ideological nature can be observed: the expansion of certain areas of applied linguistics (the teaching of foreign languages, for instance) goes some way towards meeting the requirements of an era characterised by migration. However, these requirements have no lasting repercussions on the (often low) quality of language teaching programmes, which only serves to highlight the fact that such efforts are not always rationally motivated. Thus, curiously, it is in its most challenging aspects that linguistics becomes both most dangerous and most endangered.

The dialectic between society and scientific research is an inherent part of science and is therefore not a cause for concern. Nevertheless, scientists should be aware both of their dependence on a given society at a given time, as well as of their task of looking beyond the opinions and ideologies of their era. Linguistics has frequently failed to maintain a neutral stance against the multiple pressures to which it has been subject during the two centuries of its existence. It is legitimate to suppose that once the discipline has developed a set of theoretical principles which take into account its objectives as well as the nature of its dependence on socio-political factors, it will be better able to maintain its independence.

12.2.4 Towards a new view of the discipline

The analysis of the characteristics of Romance studies involves many disparate elements: the development of a theoretical base and the elaboration of empirical knowl-

edge, the organisation of research within national academic systems and international institutions, university courses offered and job prospects for persons with a background in linguistics and, finally, the interaction between all these factors in the socio-cultural and political world that surrounds them.

Studies aiming at gaining a deeper insight into the nature of these interactions are only in their beginnings. The sources for such a novel view on the discipline are to be found not only on the shelves of libraries but also in the correspondence, the (auto-) biographical writing and the teaching materials of scholars, in university archives, in studies on the education and role of literati in society, and in personal accounts (*cf.* e.g. Gauger/Pöckl, *Wege in der Sprachwissenschaft*, 1991).

The development of an interpretative view of the discipline such as the above essentially arises from reflective thought on its current methodology.

Epilogue: On the usefulness of (Romance) linguistics

Throughout this manual, we have raised – both implicitly and explicitly – the subject of the importance of linguistics and, in particular, that of the contribution of Romance linguistics to the general understanding of language.

1 The place of Romance linguistics among the language sciences

Romance studies occupy a special position amongst the language sciences with good reason, as they lend themselves not only to the study of specific languages but also to comparative and typological research. The close relations between the languages that make up the discipline, the great variety of concrete linguistic situations they exemplify, the richness of the available historical documentation as well as the international nature of the field of research – all these are factors that combine to make Romance studies easily accessible while highlighting its highly diversified nature.

As mentioned at the beginning of this book, the particular strengths of the discipline are represented by the historical, comparative and variational approaches it involves (*cf.* 1.1), and which are the result of the unique opportunity it affords for the observation of 2,500 years of documented language history and variation (*cf.* ch. 10). The main contributions of Romance studies to general linguistics have thus traditionally been made in the fields of historical phonetics and phonology (*cf.* 6.4), semantic theory and word-formation, as well as etymology and word history (*cf.* ch. 9). More recently, contributions in the areas of morphology, morphosyntax and (micro)syntax have experienced a remarkable increase, with a particular emphasis on research which takes into account evidence from dialectal varieties (*cf.* ch. 7 and 8). In addition to the above-mentioned areas, researchers in Romance studies have always been and continue to be forerunners in dialect research methodology, in the establishment of theoretical frameworks for variational linguistics (*cf.* ch. 4) and in the elaboration of research methods concerning the intersection of philology and internal language history (*cf.* ch. 11).

2 Linguistics and society

Were one to enumerate every tangible effect of the study of language and the teaching of linguistics on the development of thought over the centuries, and thus on the development of the world as we know it, their significance for modern society would immediately become apparent. The usefulness of linguistics ultimately derives from the unquestionable usefulness of language, in so far as linguistic thought contributes to a better understanding of language, and consequently to its effectiveness as a tool. Although the contributions made by Romance and non-Romance linguistics to society are limited, they are nonetheless very real, and are frequently significant. Like writing,

linguistics can magnify the effects of the essential functions of language (*cf.* 1.2.5), particularly its cognitive and social aspects.

The cognitive functions of language are intensified in that linguistics increases our knowledge of the process of conceptualisation. More effective powers of conceptualisation enable a better understanding of the world and lead to the implementation of complex abstract systems such as modern democracy. Linguistics is indispensable for all forms of language elaboration, which rely above all on accurate grammatical analysis, and – more particularly – on solid lexicography. Within the education system, both native and non-native language teaching make use of the tools provided by linguistics. Other examples of the application of linguistics to everyday life involving the cognitive functions of language include the development of scientific terminology, translation, the development and use of computerised tools for textual analysis such as search engines, and inter-cultural contacts cultivated for the purposes of global commerce.

As far as the social function is concerned, linguistics allows a better appreciation of the role of language in shaping individual and collective identities. Once it has been acknowledged that language constitutes one of the factors that are instrumental in either uniting or dividing humankind, its dangers can more easily be avoided: linguistic awareness, for instance, may prevent the labelling of certain people as hostile merely because they speak differently. The field of linguistics remains the primary point of reference for discussions on current language change, for issues arising from language conflict, and, more generally speaking, for questions revolving around language politics or language culture. As an example of an application involving both the cognitive and the social functions of language, linguistics has played an essential role in the recent standardisation of Catalan, Galician, Romansh and Ladin.

The main advantages of interpretative historical linguistics lie in the opportunities it provides for reflection and self-reflection. In addition, linguistics in its historical dimension provides the tools for the production and analysis of texts and allows elaborate discourse. Finally, it contributes to preserving, understanding and breathing new life into a diversified written and oral heritage, through the study of textual genres and the editing of texts from historical periods or recordings of spoken dialectal texts.

In the absence of linguistic thought, written culture and the media would stagnate, the quality of higher education would suffer and the education of professional writers such as journalists and editors would lack a vital cornerstone. Eventually, the abandonment of linguistic thought could spell the end for the elaboration of complex data, thereby leading to the weakening of democratic societies owing to the ensuing reduction in administrative and other organisational capacities.

It can thus be concluded that linguistics acts as a catalyst in the same way that writing does, and as such, it further increases the potential inherent in writing. Its capacity to intensify the uses that can be made of the three functions of language is particularly apparent in languages with a significant historical heritage, amongst which the Romance languages may be counted.

In the author's opinion, therefore, linguistics should be taught at secondary level wherever possible. A democratic society requires individuals who are not only literate but who possess solid conceptual abilities, who are secure in their identity but who are aware of the dangers which arise from exclusion based on language. The capacity of linguistics to magnify the functions of language makes it a classic 'auxiliary science': while its immediate applicability to daily life may be low, it is an essential tool for the development of strategies which aim to provide a foundation for human existence and co-existence built on conscious reflection.

References

The bibliography provides complete references for the works that are cited in abbreviated form (i.e. author or acronym + (in some cases) *title*, date) throughout the *Companion*. References for the individual articles from the reference encyclopedias LRL, RSG, HSK, CambrHist, OxGuide, CambrHandb, OxEnc and MRL (cf. 5.1 and 2.4.1 above) cited either in the text itself or as suggestions for further reading at the end of a chapter are not provided separately here. Likewise, only the most important dictionaries presented in section 9.9 (*Historical and etymological lexicography*) are included.

The first section lists acronyms and abbreviations, including those of journals and series that appear in individual references in the second section (e.g. RLiR for *Revue de Linguistique Romane*).

1 Acronyms and abbreviations

(Journals, bibliographies, collections, dictionaries, databases)

ADDU = Thun, Harald / Adolfo, Elizaincín (2000–). *Atlas lingüístico diatópico y diastrático del Uruguay*. Kiel: Westensee.

AIS = Jaberg, Karl / Jud, Jakob (1928–1940). *Sprach- und Sachatlas Italiens und der Südschweiz / Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale*, 8 vols. Zofingen: Ringier. [cf. Jaberg and Jud 1928]

ALCat = Veny Joan / Pons i Griera, Lúdia (2001–2018). *Atles lingüístic del domini català*, 9 vols. Barcelona: Institut d'estudis catalans.

ALD-I/ALD-II = Goebel, Hans et al. (1998, 2012). *Atlant linguistisch dl ladin dolomitich y di dialec vejins*. Wiesbaden: Reichert; Strasbourg: SLiR/ELiPhi.

ALF = Gilliéron, Jules / Edmont, Edmond (1902–1910). *Atlas linguistique de la France*. Paris: Champion.

ALI = Matteo, Giulio Bartoli et al. (1995–2018). *Atlante linguistico italiano*, 9 vols. Rome: Istituto poligrafico e Zecca dello Stato. <<https://www.atlantelinguistico.it>>.

ALiR = Tuaillon, Gaston (dir.) (1996–). *Atlas linguistique roman*. Rome: Istituto poligrafico e Zecca dello Stato.

ALPI = Navarro Tomás, Tomás (dir.) (1962). *Atlante lingüístico de la Península ibérica*. Madrid: Consejo Superior de Investigaciones Científicas.

AND = Stone, Louise / Rothwell, William (dirs) (1977–1992). *Anglo-Norman Dictionary*. London: Modern Humanities Research Association. Second edition: Rothwell, William / Gregory, Steward / Trotter, David (2005). *Anglo-Norman Dictionary: Revised Edition, A–C; D–E*, 2 vols. London: MHRA. Online version: <www.anglo-norman.net>.

BEC = *Bibliothèque de l'École des Chartes*.

BEdT = Asperti, Stefano (2014). *Bibliografia elettronica dei trovatori (BEdT)*, <www.bedt.it>.

BLL = *Bibliography of linguistic literature/Bibliographie linguistischer Literatur* (1975–). Frankfurt am Main: Klostermann.

BIWbg = Bloch, Oscar / Wartburg, Walther von (^c1975). *Dictionnaire étymologique de la langue française*, Paris: PUF.

BrCor = Corominas, Joan (^c1973). *Breve diccionario etimológico de la lengua castellana*. Madrid: Gredos.

BSLP = *Bulletin de la Société de Linguistique de Paris*.

- CambrHandb = Ledgeway, Adam / Maiden, Martin (eds) (2022). *The Cambridge Handbook of Romance Linguistics*. Cambridge: CUP.
- CambrHist = Maiden, Martin / Smith, John C. / Ledgeway, Adam (eds) (2011–2013). *The Cambridge History of the Romance Languages*, 2 vols. Cambridge: CUP.
- CCM = *Cahiers de Civilisation Médiévale*.
- CDH = Real Academia Española (2008). *Corpus del Diccionario histórico del español*, online access: <<https://www.rae.es/banco-de-datos/cdh>>.
- CIL = Concilio et auctoritate Academiae Scientiarum Berolinensis et Brandenburgensis (ed.) (1863–). *Corpus Inscriptionum Latinarum*, [currently 17 vols.]. Berlin: de Gruyter.
- CLA = V.v.A.a. (1954–1998). *Chartae Latinae antiquiores: facsimile-edition of the Latin charters prior to the ninth century*, 49 vols. Dietikon: Graf.
- CODOLCAT = V.v.A.a. (2008). *Corpus documentale latinum Cataloniae* <<http://gmlc.imf.csic.es/codolcat>>.
- CORDE = Real Academia Española (2008). *Corpus diacrónico del español*, online access: <<https://www.rae.es/banco-de-datos/corde>>.
- CREA = Real Academia Española (2008). *Corpus de referencia del español actual*, online access: <<https://www.rae.es/banco-de-datos/crea>>.
- DAG = Baldinger, Kurt et al. (dirs) (1975–2021). *Dictionnaire onomasiologique de l'ancien gascon*. Tübingen: Niemeyer.
- DAO = Baldinger, Kurt (dir.) (1975–2007). *Dictionnaire onomasiologique de l'ancien occitan*. Tübingen: Niemeyer.
- DCELCH = Corominas, Joan / Pascual, José Antonio (²1980–1991). *Diccionario crítico etimológico de la lengua castellana y hispánica*, 6 vols. Madrid: Gredos.
- DCR = Cuervo, Rufino José (1886–1998). *Diccionario de construcción y régimen de la lengua castellana*, 8 vols. Paris: Roger; Bogotá: Instituto Caro y Cuervo.
- DCVB = Alcover, Antoni M. / Moll, Francesc de B., *Diccionari Català-Valencià-Balear*, 10 vols. Palma de Mallorca, 1926–1969; vols. 1+2, ²1964/1969.
- DEAF = Baldinger, Kurt / Möhren Frankwalt / Städtler, Thomas (dirs) (1974–2022). *Dictionnaire étymologique de l'ancien français*. Tübingen: Niemeyer.
- DEAFCompl = Möhren, Frankwalt (²1993, ²2007). *Dictionnaire étymologique de l'ancien français. Complément bibliographique*, Tübingen: Niemeyer; Quebec: PU Laval.
- DEAFél = Möhren, Frankwalt / Städtler, Thomas (dirs) (2010–2023). *Dictionnaire étymologique de l'ancien français électronique* (including printed DEAF and DEAFpré). Heidelberg: Heidelberg Academy of Sciences, <<http://deaf-server.adw.uni-heidelberg.de>>.
- DEAO = Glessgen, Martin (dir.) (2023–), *Dictionnaire étymologique d'ancien gascon*, online access: <<https://gallrom/linguistik.uzh.ch>>.
- DECLC = Corominas, Joan (1980–1989). *Diccionari etimològic i complementari de la llengua catalana*, 9 vols. Barcelona: Curial.
- DEEH = García de Diego, Vicente (²1985). *Diccionario etimológico español e hispánico*. Madrid: Espasa-Calpe.
- DEI = Battisti, Carlo / Alessio, Giovanni (1950–1957). *Dizionario etimologico italiano*, 5 vols. Florence: Barbèra.
- DELIN = Cortelazzo, Manlio / Zolli, Paolo (²1999). *Il nuovo etimologico: DELI–Dizionario etimologico della lingua italiana*. Bologna: Zanichelli.
- DELR = Ciorănescu, Alexandru (2001). *Dicționarul etimologic al limbii române*. Bucharest: Editura Saeculum I. O.
- DEM = Müller, Bodo (1987–). *Diccionario del español medieval*. Heidelberg: Winter.
- DÉRom = Buchi, Éva / Schweickard, Wolfgang (dirs) (2008–). *Dictionnaire Étymologique Roman*. Nancy: ATILF, <<http://www.atilf.fr/DERom>> [3 printed volumes: Iid. (eds) (2014–2020). 1. *Genèse, méthodes et résultats*. – 2. *Pratique lexicographique et réflexions théoriques*. – *Entre idioroman et protoroman*. Berlin; Boston: de Gruyter].
- DES = Wagner, Max Leopold (1960–1964). *Dizionario etimologico sardo*, 3 vols. Heidelberg: Winter.
- DETEMA = Herrera, María Teresa (ed.) (1996). *Diccionario español de textos médicos antiguos*, 2 vols. (+ CD-ROM edition). Madrid: Arco Libros.

- DH = Real Academia Española (1960–). *Diccionario histórico de la lengua española*, Madrid: Real Academia, online access: <<https://www.rae.es/dhle>>.
- DHFQ = Poirier, Claude (ed.) (1998). *Dictionnaire historique du français québécois*. Quebec: PU Laval.
- DI = Schweickard, Wolfgang (1997–2013). *Deonomasticon Italicum. Dizionario storico dei derivati da nomi geografici e da nomi di persona. I: Derivati da nomi geografici*, 4 vols. Tübingen: Niemeyer.
- DiccAut = Real Academia Española (1726–1739). *Diccionario de Autoridades*, 3 vols. Madrid [Facsimile edition, Madrid: Gredos, 1969].
- DictEtFrLux = Bender-Berland, Geneviève / Kramer, Johannes / Reisdorfer, Joseph (2003–). *Dictionnaire étymologique des éléments français du Luxembourgeois*. Tübingen: Narr.
- DictHist = Rey, Alain (ed.) (1992). *Le Robert. Dictionnaire historique de la langue française*, 2 vols. Paris: Robert. New expanded edition, 2010.
- DiccIdeol = Casares, Julio (1942, ²1959). *Diccionario ideológico de la lengua española*. Barcelona: Gili.
- DLR = *Diționarul limbii române (DLR)*. Bucharest, 1913; Seria noua, EA, 1986–.
- DME = Alonso Pedraz, Martín (1986). *Diccionario medieval español. Desde las Glosas emilianenses y Silenses (s. x) hasta el s. xv*, 2 vols. Salamanca: Universidad Pontificia.
- DMF = Robert, Martin (dir.) (2021). *Dictionnaire du moyen français*, online access: <zeus.atilf.fr/dmf>.
- DocLing = Glessgen, Martin (dir.) (⁴2023). *Documents linguistiques galloromans, Édition électronique*, en collaboration avec Hélène Carles, Frédéric Duval et Paul Videsott, online access: <<https://gallrom.linguistik.uzh.ch>>.
- DOM = Stimm, Helmut / Stempel, Wolf-Dieter (dirs) (1996–2022). *Dictionnaire étymologique de l'occitan médiéval*. Tübingen: Niemeyer.
- DRAE = A.a.V.v. (²¹1992). *Diccionario de la lengua española*, 2 vols. Madrid: Real Academia Española.
- DRF = Rézeau, Pierre (ed.) (2001). *Dictionnaire des régionalismes de France*. Brussels: de Boeck.
- DRG = De Planta, Robert (ed.) (1939–). *Dicziunari Rumantsch Grischun*. Cuoir: Bischofberger; Stamparia: Winterthur S.A.
- DRM = Carles, Hélène / Glessgen, Martin / Robecchi, Marco / Bossone, Alessandra (2024). *Dictionnaire des régionalismes médiévaux: la Galloromania orientale (DRM). Une analyse des Documents linguistiques galloromans (XII^e–XV^e s.)*, Strasbourg, ÉLiPhi.
- DSEF = Pellegrini, Giovan Battista (ed.) (1984–). *Dizionario Storico Etimologico Friulano*. Udine: Casamassima.
- DSR = Thibault, André (1997, ²2004). *Dictionnaire suisse romand. Particularités lexicales du français contemporain*. Carouge; Geneva: Éditions Zoé.
- DTS = Kristol, Andres (ed.) (2005). *Dictionnaire toponymique des communes suisses*. Frauenfeld: Huber.
- EA = *Español Actual*.
- ER = *Estudis Romànics*.
- EWD = Kramer, Johannes (1988–1999). *Etymologisches Wörterbuch des Dolomitenladinischen*, 7 vols. Hamburg: Buske.
- FEW = Wartburg, Walther von (continued by Theodor Gossen, Jean-Pierre Chambon and Jean-Paul Chauveau) (1922–2002). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprach-schatzes*, 25 vols. Bonn: Schroeder; Basel: Zbinden.
- FEWCompl = Chauveau, Jean-Paul / Greub, Yan / Seidl, Christian, *Französisches Etymologisches Wörterbuch. Complément* (³2010). Strasbourg: SLR/ELiPhi.
- FEWGuide = Carles, Hélène / Dallas, Marguerite / Martin, Glessgen / Thibault, André (2019). *Französisches Etymologisches Wörterbuch: Guide d'utilisation*. Strasbourg: SLR/ELiPhi.
- FEWInd = Buchi Éva (ed.) (2003). *Französisches etymologisches Wörterbuch. Index*, 2 vols. Paris: Champion.
- Frantext = cf. TLF.
- GallRom = Glessgen, Martin (dir.) (2023–). *Documents et analyses linguistiques de la Galloromania médiévale*, online access: <<https://gallrom.linguistik.uzh.ch>>.
- Gdf = Godefroy, Frédéric (1880–1902). *Dictionnaire de l'ancienne langue française et de tous ses dialectes du ix^e au xv^e siècle*, 10 vols. Paris: Bouillon, free online access: <<http://micmap.org/dicfro/search/dictionnaire-godefroy/>>.

- GDLI = Battaglia, Salvatore (1961–2002). *Grande Dizionario della lingua italiana*, 21 vols. Turin: UTET.
- GMF = Riegel, Martin / Pellat, Jean-Christophe / Rioul, René (°2016). *Grammaire méthodique du français*. Paris: PUF [°1994].
- GPSR = Gauchat, Louis *et al.* (eds) (1924). *Glossaire des Patois de la Suisse Romande*. Geneva: Droz.
- GRMLA = Jauss, Hans Robert *et al.* (eds) (1968–). *Grundriß der romanischen Literaturen des Mittelalters*. Heidelberg: Winter.
- HistLing = *Historiographia linguistica*.
- HSK = *Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of linguistics and communication science/Manuels de linguistique et des sciences de communication*. Berlin; New York: de Gruyter.
- HSK 5 = Hausmann, Franz Josef *et al.* (eds) (1989–1991). *Wörterbücher. Ein internationales Handbuch zur Lexikographie*, 3 vols. Berlin; New York: de Gruyter.
- HSK 10 = Günther, Hartmut / Ludwig, Otto (eds) (1994–1996). *Schrift und Schriftlichkeit/Writing and its use* (HSK 10. 1/2), 2 vols. Berlin; New York: de Gruyter.
- HSK 11 = Eichler, Ernst *et al.* (eds) (1995–1996). *Namenforschung/Names studies/Les noms propres. Ein internationales Handbuch zur Onomastik*, 3 vols. Berlin; New York: de Gruyter.
- HSK 12 = Goebel, Hans *et al.* (eds) (1996–1997). *Kontaktlinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, 2 vols. Berlin; New York: de Gruyter.
- HSK 20 *cf.* LangTyp.
- HSK 23 *cf.* RSG.
- HSK 40 = Müller, Peter O. *et al.* (eds) (2015–2016), *Word-Formation. An International Handbook of the Languages of Europe*, 5 vols. Berlin; Boston: de Gruyter.
- HTOED = *Historical Thesaurus of the OED*. Second edition (2020–). <<https://www.oed.com/thesaurus>> [= *Historical Thesaurus of English*, University of Glasgow: <<https://ht.ac.uk>>].
- Hu = Huguet, Edmond (1925–1967). *Dictionnaire de la langue française du seizième siècle*. Paris: Champion and Didier.
- InvSyst = Franck, Barbara / Hartmann, Jörg (eds) (1997). *Inventaire systématique des premiers documents des langues romanes*, 5 vols. Tübingen: Narr.
- KVK = *Karlsruher Virtueller Katalog*: <www.ubka.uni-karlsruhe.de/kvk.html>.
- LangTyp = Haspelmath, Martin / König, Ekkehard / Oesterreicher, Wulf / Raible, Wolfgang (eds) (2001). *Language typology and language universals/Sprachtypologie und sprachliche Universalien* (HSK 20. 1/2), 2 vols. Berlin; New York: de Gruyter.
- LEA = *Lingüística española actual*.
- LEI = Pfister, Max / Schweickard, Wolfgang / Prifti, Elton (eds) (1979–). *Lessico etimologico italiano*. Wiesbaden: Reichert.
- LEISuppl = Pfister, Max / Schweickard, Wolfgang (°2012). *LEI. Supplemento bibliografico*. Wiesbaden: Reichert.
- Le Monde = A.a.V.v. (1989–). *Le Monde sur cd-rom: disque cumulatif* (1987–). Paris: Le Monde SARL/Research Publications International.
- LRL = Holtus, Günter / Metzeltin, Michael / Schmitt, Christian (eds) (1988–2005). *Lexikon der Romanistischen Linguistik*, 8 t. (12 vols.). Tübingen: Niemeyer.
- Lv = Levy, Emil (1894–1924). *Provenzalisches Supplement-Wörterbuch. Berichtigungen und Ergänzungen zu Raynouards Lexique roman*, 8 vols. Leipzig: Reisland.
- LvP = Levy, Emil (°1966). *Petit Dictionnaire provençal-français*. Heidelberg: Winter.
- M = Mistral, Frédéric (1878). *Lou tresor dou Felibrige, ou dictionnaire provençal français embrassant les divers dialectes de la langue d'oc moderne*, 2 vols. Aix-en-Provence: Éd. Ramoun Berenguié.
- MLA-IB = A.a.V.v. (1921–). *Modern Language Association International Bibliography*. New York: MLA.
- MRL 4 = Trotter, David (ed.) (2015). *Manuel de la philologie de l'édition*. Berlin; Boston: de Gruyter.
- MRL 31 = Mensching, Guido / Savelsberg, Frank (2023). *Manual of Judaeo-Romance linguistics and philology*. Berlin; Boston: de Gruyter.
- MRL 35 = Esher, Louise / Sibille, Jean (eds) (forthcoming, July 2024). *Manuel de linguistique occitane*. Berlin; Boston: de Gruyter.

MSL = *Mémoires de la Société linguistique de Paris*.

OxfEnc = Loporcaro, Michele, [and Francesco Gardani] (eds) (2019–). *The Oxford Encyclopedia of Romance Linguistics (ORE)*. Oxford: OUP, advance online publication: <<https://oxfordre.com/linguistics>>.

OxfGuide = Maiden, Martin / Ledgeway, Adam (eds) (2016). *The Oxford guide to the Romance languages*, 1 vol. Oxford: Oxford UP.

PatRom = Cano González, Ana Maria / Germain, Jean / Kremer, Dieter (eds) (2004–). *Dictionnaire historique de l'anthroponymie romane. Patronymica Romanica*, vol. II/1– (vol. I = introductory volume, 1997). Tübingen: Niemeyer.

PRob = Rey-Debove Josette / Rey, Alain (eds) (2022). *Le Nouveau Petit Robert*. Paris: Dictionnaires le Robert.

PU = *Presses Universitaires de (...)*.

RB = A.a.V.v. *Romanische Bibliographie/Bibliographie romane/Romance Bibliography* (1878–). Tübingen: Niemeyer.

RevFilEsp = *Revista de filología española*.

REW = Meyer-Lübke, Wilhelm (1935). *Romanisches etymologisches Wörterbuch*. Heidelberg: Winter.

RGG = *Romanistik in Geschichte und Gegenwart*.

RION = *Rivista Italiana di Onomastica*.

RJb = *Romanistisches Jahrbuch*.

RK = *Romanistisches Kolloquium* (a series with various editors, focusing on key topics in Romance linguistics, cf. the following five RK volumes).

RK III = Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael (eds) (1990). *Die romanischen Sprachen und die Kirchen*, RK III. Tübingen: Narr.

RK VIII = Dahmen, Wolfgang / Holtus, Günter / Kramer, Johannes / Metzeltin, Michael / Schweickard, Wolfgang / Winkelmann, Otto (eds) (1995). *Konvergenz und Divergenz in den romanischen Sprachen*, RK VIII.

RK IX = Dahmen, Wolfgang et al. (eds) (1996). *Die Bedeutung der romanischen Sprachen im Europa der Zukunft*, RK IX.

RK XIII = Dahmen, Wolfgang et al. (eds) (2000). *Schreiben in einer anderen Sprache*, RK XIII.

RK XIV = Dahmen, Wolfgang et al. (eds) (2000). *Kanonbildung in der Romanistik und in den Nachbardisziplinen*, RK XIV.

RK XXXIV = Becker, Lidia et al. (eds) (2023). *Digitale romanistische Sprachwissenschaft. Stand und Perspektiven*.

RLiR = *Revue de linguistique romane*.

Rn = Raynouard, François (1838–1844). *Lexique roman ou dictionnaire de la langue des troubadours comparée avec les autres langues de l'Europe latine*, vols. 2–6, reprint. Heidelberg: Winter.

RSG = Ernst, Gerhard / Glessgen, Martin / Schmitt, Christian / Schweickard, Wolfgang (eds) (2003–2008). *Romanische Sprachgeschichte/Histoire linguistique de la Romania* (HSK 23.1–3). 3 vols. Berlin; New York: de Gruyter.

StMed = *Studi Medievali*.

TB = Tommaseo, Nicolò / Bellini, Bernardo (1865–1879). *Dizionario della lingua italiana*, 7 vols. Rome: Unione tipografico-editrice.

TesLex = Gili Gaya, Samuel (1960). *Tesoro lexicográfico. 1492–1726*. Madrid: Aguirre Torre.

ThesDipl = CETDOC et al. (1997). *Thesaurus Diplomaticus* (TD-1). Turnhout: Brepols [CD-ROM].

ThesLL = A.a.V.v. (1900–). *Thesaurus linguae latinae*, [currently 10 vols.]. Leipzig: Teubner.

TL = Tobler, Adolf / Lommatzsch, Erhard (continued by Hans H. Christmann and Richard Baum) (1925–2002). *Altfranzösisches Wörterbuch*, 11 vols. Berlin: Weidmann; Wiesbaden: Steiner.

TLF = Paul Imbs (ed.) (1971–1994). *Trésor de la langue française. Dictionnaire de la langue du ^{xx}e et du ^{xx}e siècle (1789–1960)*. 16 vols. Paris: CNRS; Klincksieck; Gallimard.

TLF-Étym = mise à jour des notices étymologiques du TLF informatisé (1983–). <<http://stella.atilf.fr/scripts/TLFETym.exe>>.

TraLiPhi = *Travaux de Linguistique et de Philologie*.

TypSources = Genicot Léopold (ed.) (1972–). *Typologie des sources du Moyen Âge occidental*. Turnhout: Brepols.

VR = *Vox Romanica*.

WALS = Dryer, Matthew S. / Haspelmath, Martin (2011). *The World Atlas of Language Structures Online*, online access: <<https://wals.info>>

ZfSL = *Zeitschrift für französische Sprache und Literatur*.

ZrP = *Zeitschrift für romanische Philologie*.

2 General reference works, monographs and articles

Abeillé, Anne / Godard Danièle (eds) (2021). *La grande grammaire du français*, 2 vols. Arles: Actes Sud/Paris: Imprimerie nationale éditions.

Adams, James N. (2003). *Bilingualism and the Latin Language*. Cambridge: CUP.

Adams, James N. (2007). *The Regional Diversification of Latin 200 BC – AD 600*. Cambridge: CUP.

Adams, James N. (2013). *Social Variation and the Latin Language*. Cambridge: CUP.

Allières, Jacques (2001). *Manuel de linguistique romane*. Paris: Champion.

Antoine, Gérard et al. (1985–2000). *Histoire de la langue française*, vols. 1 + 2: Antoine, Gérard / Martin, Robert (eds), 1880–1914 (ed. 1985); 1914–1945 (1995); vol. 3: Antoine, Gérard / Cerquiglini, Bernard (eds), 1945–2000 (2000). Paris: CNRS.

Aprile, Marcello (2004). *Le strutture del Lessico Etimologico Italiano*. Galatina: Congedo.

Arnauld, Antoine / Lancelot, Claude (1660). *Grammaire générale et raisonnée (de Port-Royal)*. Geneva: Slatkine reprints (1972) [1754 edition accessible on Gallica].

Aronoff, Mark (1994). *Morphology by Itself: stems and inflectional classes*. Cambridge, MA: MIT Press.

Arveiller, Raymond (1999). *Addenda au FEW XIX (Orientalia)*, ed. Max Pfister. Tübingen: Niemeyer.

Austin, John L. (1962). *How to do things with words*. New York: Oxford UP.

Baglioni, Daniele (2010). *L'italiano delle cancellerie tunisine (1590-1703). Edizione e commento linguistico delle "carte Cremona"*. Roma: Scienze e Lettere.

Baglioni, #forthcoming#.

Baker, Craig et al. (2018). *L'Ombre de Joseph Bédier. Théorie et pratique éditoriales au XX^e siècle*. Strasbourg: ELiPhi.

Bal, Willy / Germain, Jean / Klein, Jean / Swiggers, Pierre (²1997). *Bibliographie sélective de linguistique française et romane*. Louvain-la-Neuve: Duculot.

Baldinger, Kurt (²1972). *La formación de los dominios lingüísticos en la Península Ibérica*. Madrid: Gredos (German edition: 1958).

Baldinger, Kurt (1988–2003). *Etymologien. Untersuchungen zu FEW 21–23*, 3 vols. Tübingen: Niemeyer.

Bally, Charles (²1944). *Linguistique générale et linguistique française*. Paris: Leroux; Bern: Francke [¹1932].

Banniard, Michel (1992). *Viva voce: communication écrite et communication orale du IV^e au IX^e siècle en Occident latin*. Paris: Institut des Études Augustiniennes.

Barbato, Marcello (2015). 'Il francoitaliano: storia e teoria', *Medioevo romanzo* 39, 22–51.

Barbato, Marcello (2017). *Le lingue romanze. Profilo storico-comparativo*. Rome: Laterza.

Bartoli, Matteo (1925). *Introduzione alla neolinguistica*. Geneva: Olschki.

Bec, Pierre (1970–1971). *Manuel pratique de philologie romane*, 2 vols. Paris: Picard.

Bédier, Joseph (ed.) (1913). *Le Lai de l'Ombre par Jean Renart*, critical edition. Paris: Firmin-Didot (revision of the earlier edition, Fribourg: Imprimerie et Librairie de l'Œuvre de Saint-Paul, 1890).

Bédier, Joseph (ed.) (1921–1927). *La Chanson de Roland*. Paris: Piazza.

Bédier, Joseph (1929 [1928]). *La tradition manuscrite du Lai de l'Ombre. Réflexions sur l'art d'éditer les anciens textes*. Paris: Champion.

Beltrami, Pietro (2017). *La filologia romanza. Profilo linguistico e letterario*. Bologna: il Mulino.

Berkenbusch, Gabriele (1990). *Klassiker der spanischen Sprachwissenschaft: eine Einführung in die Sprachwissenschaft des 16. und 17. Jahrhunderts: Quellentexte, Kommentare, Bibliographie*. Bonn: Romanistischer Verlag.

- Berger, Günter (1997). '«Mouvance, variance» und die Folgen: Griselda und ihre «Nachkommen», in Glessgen, Martin / Lebsanft, Franz (eds), *Alte und neue Philologie*. Tübingen: Niemeyer, 255–65.
- Bernard, Carmen / Gruzinski, Serge (1991–1993). *Histoire du nouveau monde*, 2 vols. Paris: Fayard.
- Bernissan, Fabrice (2012). 'Combien de locuteurs compte l'occitan en 2012?', *RLiR* 76: 467–512.
- Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: CUP.
- Bickerton, Derek (1975). *Dynamics of a creole system*. London: CUP.
- Bierbach, Mechtild (1997). *Grundzüge humanistischer Lexikographie in Frankreich: ideengeschichtliche und rhetorische Rezeption der Antike als Didaktik*. Tübingen; Basel: Francke.
- Blanche-Benveniste, Claire (1990). *Le français parlé: études grammaticales*. Paris: CNRS.
- Blanche-Benveniste, Claire (2010). *Approches de la langue parlée en français*. Paris: Ophrys [1997].
- Blank, Andreas (1997). *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: Niemeyer.
- Blank, Andreas (2001). *Einführung in die lexikalische Semantik für Romanisten*. Tübingen: Niemeyer.
- Bloomfield, Leonard (1933). *Language*. New York: Henry Holt.
- Blumenthal, Peter / Rovere, Giovanni (1998). *PONS-Wörterbuch der italienischen Verben: Konstruktionen, Bedeutungen, Übersetzungen*. Stuttgart et al.: Klett.
- Bolocan, Gheorghe (1981). *Dicționarul elementelor românești din documentele slavo-române (1374–1600)*, Bucharest, Editura Academiei Republicii Socialiste România.
- Bosque, Ignacio / Demonte, Violeta (eds) (1999). *Gramática descriptiva de la lengua española*, R.A.E., 3 vols. Madrid: Espasa Calpe.
- Bossong, Georg (1990). *Sprachwissenschaft und Sprachphilosophie in der Romania: von den Anfängen bis August Wilhelm Schlegel*. Tübingen: Narr.
- Bossong, Georg (1998). 'Vers une typologie des indices actantiels. Les clitics romans dans une perspective comparative', in Ramat, Paolo / Roma, Elisa (eds), *Sintassi storica. Atti del XXX Congresso Internazionale della SLI*. Rome: Bulzoni, 9–43.
- Bossong, Georg (2006). 'La sintaxis de las Glosas Emilianenses en una perspectiva tipológica', in de Bustos Tovar, José J. / Alconchel, José L. Girón (eds), *Actas del VI Congreso Internacional de Historia de la Lengua española*. Madrid: Arco Libros, 529–44.
- Bossong, Georg (2008). *Die romanischen Sprachen. Eine vergleichende Einführung*. Hamburg: Buske.
- Bossuat, Robert (1951–1986). *Manuel bibliographique de la littérature française du Moyen Age*, Melun: Librairie d'Argences. 1955, *Supplément* [1949–1953]; 1961, *Second supplément* [1954–1960], in collaboration with Jacques Monfrin; 1986, *Troisième supplément* [1960–1980], Françoise Vieliard and Jacques Monfrin.
- Bréal, Michel (1897). *Essai de sémantique*. Paris: Hachette.
- Briquet, Claude Moïse (1907). *Les filigranes. Dictionnaire historique des marques du papier*, 4 vols. Paris: Picard.
- Brunel, Clovis (1926–1952). *Les plus anciennes chartes en langue provençale: recueil des pièces originales antérieures au xiii^e siècle*, 2 vols. Paris: Picard.
- Brunet, Étienne (1981). *Le Vocabulaire français de 1789 à nos jours d'après les données du Trésor de la Langue Française*, 3 vols. Geneva: Slatkine; Paris: Champion.
- Brunet, Étienne (1995). *CD-Rom Rabelais*. Paris: Les Temps qui courent.
- Brunet, Étienne (2004). *Rabelais sur Internet*, online: <<http://ancilla.unice.fr/rabelais.html>>.
- Brunot, Ferdinand (1905–1939). *Histoire de la langue française des origines à 1900* [since 1948: [...] *des origines à nos jours*], 13 vols. (vols. 12–13 ed. Charles Bruneau). Paris: Armand Colin.
- Brunot, Ferdinand (1911–1914). *Archives de la parole. Enregistrements*. <<https://gallica.bnf.fr/html/und/enregistrements-sonores/archives-de-la-parole-ferdinand-brunot-1911-1914>>.
- Buchi, Éva (1994). 'Les noms de genres dans la *Flore française* de Lamarck (1778): genèse et réception d'une terminologie', *RLiR* 58: 97–141.
- Buchi, Éva (1996). *Les Structures du 'Französisches Etymologisches Wörterbuch'*. Tübingen: Niemeyer.
- Buchi, Éva (2001). 'La méthodologie de l'étymologie des noms de famille (domaine français et gallo-roman)', *RION* 7: 105–27.
- Bühler, Karl (1934). *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena: Fischer.

- Buridant, Claude (2022). *Grammaire du français médiéval (GFM). XII^e–XIV^e siècles*. Strasbourg: SLR/ELiPhi [2019].
- Busse, Winfried / Dubost, Jean-Pierre (1977). *Französisches Verblexikon. Die Konstruktion der Verben im Französischen*. Stuttgart: Klett [repr. 1983].
- Busse, Winfried (1994). *Dicionário sintático de verbos portugueses*. Coimbra: Almedina.
- Camproux, Charles (2019). *Les Langues romanes*. Paris: PUF [1974].
- Cappelli, Adriano (1899). *Dizionario di abbreviature latine ed italiane*. Milan: Hoepli.
- Carles, Hélène (2011). *L'émergence de l'occitan pré-textuel. Analyse linguistique d'un corpus auvergnat (IX^e–XI^e siècles)*. Strasbourg: SLR/ELiPhi.
- Carles, Hélène (2013). 'L'innovation lexicale chez Chrétien de Troyes', *Romania* 131: 281–337.
- Carles, Hélène (2017). *Trésor galloroman des origines (TGO). Les trajectoires étymologiques et géolinguistiques du lexique galloroman en contexte latin (ca 800–1120)*. Strasbourg: ELiPhi.
- Carles, Hélène / Glessgen, Martin (2020). *Les écrits des Poilus*. Strasbourg: ELiPhi.
- Castro, Amerigo (2017). *De la edad conflictiva. Crisis de la cultura española en el siglo XVII*. Madrid: Taurus [1961].
- Chambon, Jean-Pierre (1995). 'Remarques sur la patrie de l'auteur du ms. Brit. Mus. Add. 17920 (Brunel= Ms 13)', *RLiR* 59: 5–24.
- Chambon, Jean-Pierre (1996–1997). 'Aspects de l'œuvre linguistique de Georges Straka: chronologie relative', *Orbis* 39: 97–126.
- Chambon, Jean-Pierre (1998). 'L'identité langagière des élites cultivées d'Arvernie autour de l'an Mil et la *scripta latina rustica*: réflexions à propos du *Breve de libros* du chapitre cathédral de Clermont (984–1010)', *RLiR* 62: 381–408.
- Chambon, Jean-Pierre (1999). 'L'agencement spatial et fonctionnel des vicairies carolingiennes dans le midi de la Gaule: une approche linguistique', *RLiR* 63: 55–174.
- Chambon, Jean-Pierre (2002). 'Sur le système latin de dénomination des localités (toponymie de la Gaule)', *RLiR* 66: 119–29.
- Chambon, Jean-Pierre (2004). 'L'onomastique du censier interpolé (ca 946) dans la charte de fondation du monastère auvergnat de Sauxillanges', *RLiR* 68: 105–80.
- Chambon, Jean-Pierre (2005). 'Toponymie et grammaire historique: les noms de lieux issus de 'cappella' et 'forestis' et la diffusion spatiale de l'article défini dans la Galloromania', in James-Raoul, Danièle / Soutet, Olivier (eds), *Par les mots et les textes. Mélanges (...) offerts à Claude Thomasset*. Paris: PU Sorbonne: 143–55.
- Chambon, Jean-Pierre (2006). 'Lexicographie et philologie: réflexions sur les *glossaires* d'éditions de textes (français médiéval et préclassique, ancien occitan)', *RLiR* 70: 123–41.
- Chambon, Jean-Pierre (2007). 'Remarques sur la grammaire comparée-reconstruction en linguistique romane (situation, perspectives)', *MSL* 15: 57–72.
- Chambon, Jean-Pierre (2017). *Méthodes de recherche en linguistique romane*, 2 vols. Strasbourg: ELiPhi.
- Chambon, Jean-Pierre / Greub, Yan (2000). 'Données nouvelles pour la linguistique gallo-romane: les légendes monétaires mérovingiennes', *BSLP* 95: 147–82.
- Chambon, Jean-Pierre / Greub, Yan (2002). 'Note sur l'âge du (proto)gascon', *RLiR* 66: 473–95.
- Chambon, Jean-Pierre (2014). 'Sur la date de la fixation de l'article défini devant les noms de lieux: sondages en domaine occitan', *Estudis Romànics* 36, 2014: 373–81.
- Chambon, Jean-Pierre (2023). *Recherches sur la toponymie de l'arrondissement de Lure (Haute-Saône): linguistique historique, dialectologie, traces d'histoire*. Strasbourg: ELiPhi.
- Chauveau, Jean-Paul / Lavoie, Thomas (1993). 'À propos des origines dialectales du lexique québécois', *RLiR* 57: 373–420.
- Chomsky, Noam (1957). *Syntactic structures*. The Hague: Mouton & Co.
- Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Christmann, Hans H. (1958). *Lateinisch 'calere' in den romanischen Sprachen: mit besonderer Berücksichtigung des Französischen*. Wiesbaden: Steiner.

- Christmann, Hans H. (1985). *Romanistik und Anglistik an der deutschen Universität im 19. Jahrhundert. Ihre Herausbildung als Fächer und ihr Verhältnis zu Germanistik und klassischer Philologie*. Stuttgart: Steiner (Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse, Akademie der Wissenschaften und der Literatur, Mainz, Jg. 1985/1).
- Comrie, Bernard / Haspelmath, Martin / Bickel, Balthasar (2008). *Leipzig Glossing Rules*. Dept. of Linguistics, MPI for Evolutionary Anthropology, <www.eva.mpg.de/lingua/resources>.
- Confais, Jean-Paul (1990). *Temps, mode, aspect: les approches des morphèmes verbaux et leurs problèmes à l'exemple du français et de l'allemand*. Toulouse: PU Mirail.
- Corder, Pit / Roulet, Eddy (eds) (1977). *The notions of simplification, interlanguages and pidgins and their relation to second language pedagogy. Actes du 5^e colloque de linguistique appliquée de Neuchâtel, 20 au 20 mai 1976*. Geneva: Droz.
- Coseriu, Eugenio (1956). *La creación metafórica en el lenguaje*. Montevideo: Publicaciones de la Universidad.
- Coseriu, Eugenio (1966). 'Structure lexicale et enseignement du vocabulaire', in *Actes du Premier Colloque International de Linguistique Appliquée*, Nancy: [s.n.], 175–252 [= *Einführung in die strukturelle Betrachtung des Wortschatzes*, Tübingen: Narr, 1970].
- Coseriu, Eugenio (1981). 'Los conceptos de 'dialecto', 'nivel' y 'estilo de lengua' y el sentido propio de la dialectología', *LEA* 3: 1–32.
- Coseriu, Eugenio (1988). *Sprachkompetenz: Grundzüge der Theorie des Sprechens*. Tübingen: Francke (Spanish trans.: *Competencia lingüística. Elementos de la teoría del hablar*. Madrid, Gredos, 1992).
- Courtois, Michèle / Grandjean, Marie-José Gasse- (2001). *La Diplomatie française du Haut Moyen Âge. Inventaire des chartes originales antérieures à 1121 conservées en France*. Nancy: Artem.
- Cresti, Emanuela / Moneglia, Massimo (2005). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.
- Croft, William (2001). *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford UP.
- Cunha, Antônio Geraldo da (1986–1988). *Índice do vocabulário do português medieval*. Rio de Janeiro: Fundação Casa de Rui Barbosa.
- Curell, Aguila Clara (2009). *Diccionario de galicismos del español peninsular contemporáneo*. Strasbourg: SLiR/ ELiPhi.
- D'Achille, Paolo (1990). *Sintassi del parlato e tradizione scritta della lingua italiana: analisi di testi dalle origini al secolo XVIII*. Rome: Bonacci.
- Dahmen, Wolfgang (1997). 'Editionsprobleme bei Balkanica', in Glessgen, Martin / Lebsanft, Franz (eds), *Alte und neue Philologie*. Tübingen: Niemeyer, 371–84.
- Damourette, Jacques / Pichon, Édouard (1911–1949). *Des mots à la pensée. Essai de grammaire de la langue française*. Paris: Editions d'Artrey.
- Dardel, Robert de (1996). *À la recherche du protoroman*, Tübingen: Niemeyer.
- Dauzat, Albert / Rostaing, Charles (²1978). *Dictionnaire étymologique des noms de lieux en France*. Paris: Larousse [¹1963].
- Dauzat, Albert (³1975). *Dictionnaire étymologique des noms de famille et prénoms de France*. Paris: Larousse (ed. Marie-Thérèse Morlet) [¹1951].
- Dauzat, Albert (³1977). *Traité d'anthroponymie française. Les noms de famille en France*. Paris: Payot [¹1945].
- De Mauro, Tullio (²1991). *Storia linguistica dell'Italia unita*. Bari: Laterza [¹1963].
- Diamond, Jared (1997). *Guns, germs / steel. The fate of human societies*. New York: Norton & Co.
- Dietrich, Wolf / Geckeler, Horst (1990). *Einführung in die spanische Sprachwissenschaft: ein Lehr- und Arbeitsbuch*. Berlin: Schmidt.
- Diez, Friedrich (1854). *Etymologisches Wörterbuch der romanischen Sprachen*. Bonn: Marcus.
- Dik, Simon C. (³1981). *Functional Grammar*. Dordrecht: Foris [¹1978].
- Dufter, Andreas / Stark, Elisabeth (2002). 'La variété des variétés: combien de dimensions pour la description? Quelques réflexions à partir du français', *Rjb* 53: 81–108.

- Dülmen, Richard van (²1988). *Theater des Schreckens: Gerichtspraxis und Strafroutine in der frühen Neuzeit*. Munich: Beck.
- Duval, Frédéric (2016). *Miroir des classiques-2. Les traductions galloromanes du Corpus iuris civilis. Répertoire électronique*, http://elec.enc.sorbonne.fr/miroir_des_classiques.
- Duval, Frédéric / Guillot-Barbance, Céline / Zinelli, Fabio (dir.) (2019). *Les introductions linguistiques aux éditions de textes*. Paris: Garnier.
- Eberenz, Rolf (1991). 'Castellano antiguo y español moderno: reflexiones sobre la periodización en la historia de la lengua', *RevFilEsp* 7: 79–106.
- Elcock, William D. (²1975). *The Romance Languages. Revised with a New Introduction by J.N. Green*. London: Faber [¹1960].
- Ernst, Gerhard (1970). *Die Toskanisierung des römischen Dialekts im 15. und 16. Jahrhundert*. Tübingen: Niemeyer.
- Ernst, Gerhard (1985). *Gesprochenes Französisch zu Beginn des 17. Jahrhunderts. Direkte Rede in Jean Héroards Histoire particulière de Louis XIII, 1605–1610*. Tübingen: Niemeyer.
- Ernst, Gerhard (2019). *Textes français privés des XVII^e et XVIII^e siècles*, 2 vols. Berlin; Boston: de Gruyter.
- Ernst, Gerhard / Wolf, Barbara (2005). *Textes français privés des XVI^e et XVII^e siècles*, CD-ROMs 1–3. Tübingen: Niemeyer.
- Fabra, Pompeu (1912). *Gramàtica de la lengua Catalana*. Barcelona: L'Aveve.
- Fanciullo, Franco (1996). *Fra Oriente e Occidente. Per una storia linguistica dell'Italia meridionale*. Pisa: ETS.
- Faraoni, Vincenzo (2018). 'L'origine dei plurali italiani in -e e -i. Alessandria: Edizioni dell'Orso.
- Faré, Paolo A. (1972). *Postille italiane al Romanisches etymologisches Wörterbuch, de W. Meyer-Lübke, comprendenti le Postille italiane e ladine di Carlo Salvioni*. Milan: Istituto Lombardo di Scienze e Lettere.
- Ferguson, Charles A. (1959). 'Diglossia', *Word* 15: 325–40.
- Fernandez Ramirez, Salvador (1951). *Gramática española I.: los sonidos, el nombre y el pronombre*. Madrid: Revista de Occidente.
- Fernandez Ramirez, Salvador (²1991). *Gramática española*. Madrid: Arco Libros [¹1985–86 (posthumous)].
- Fillmore, Charles J. (1968). 'The case for case', in Bach, Emmon / Harms, Robert T. (eds), *Universals in linguistic theory*. New York et al.: Holt, etc., 1–88.
- Fillmore, Charles J. et al. (1988). 'Regularity and idiomaticity in grammatical constructions: The case of 'Let alone'', *Language* 64: 501–38.
- Fishman, Joshua (1967). 'Bilingualism with and without diglossia; diglossia with and without bilingualism', *Journal of Social Issues* 23: 29–38.
- Fleischman, Suzanne (1982). *The future in thought and language. Diachronic evidence from Romance*. Cambridge: CUP.
- Flydal, Leiv (1951/52). 'Remarques sur certains rapports entre le style et l'état de la langue', *Norsk tidsskrift for sprogvidenskap* 16: 240–57.
- Fradin, Bernard / Kerleroux, Françoise / Plenat, Marc (eds) (2009). *Aperçus de morphologie du français*. Paris: PU Vincennes.
- Francard, Michel et al. (³2021). *Dictionnaire des belgicisms*. Brussels: De Boeck; Duculot [²2010].
- Frank, Barbara (1994). *Die Textgestalt als Zeichen. Lateinische Handschriftentradition und die Verschriftlichung der romanischen Sprachen*. Tübingen: Narr.
- Gabelentz, Georg von der (²1901). *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: Tauchnitz [¹1891] (reprint. e.g. Tübingen: Francke, 2003).
- Gamillscheg, Ernst (1934–1936). *Romania Germanica. Sprach- und Siedlungsgeschichte der Germanen auf dem Boden des alten Römerreiches*, 3 vols. (vol. 1: ²1970). Berlin; Leipzig: de Gruyter.
- Gargallo Gil, José Enrique / Bastardas, Maria Reina (eds) (2007). *Manual de lingüística románica*. Barcelona: Ariel.
- Gartner, Theodor (1879). *Die Grödnertal-Mundart*. Linz: self-published.
- Gauger, Hans Martin / Oesterreicher, Wulf / Windisch, Rudolf (1981). *Einführung in die romanische Sprachwissenschaft*. Darmstadt, Wissenschaftliche Buchgesellschaft. [Spanish trans. of the first part: Gauger, Hans Martin (1990). *Introducción a la lingüística románica*, Madrid: Gredos.]

- Gauger, Hans Martin / Pöckl, Wolfgang (eds) (1991). *Wege in der Sprachwissenschaft: vierundvierzig autobiographische Berichte: Festschrift für Mario Wandruszka*. Tübingen: Narr.
- Geeraerts, Dirk (2004). *Theories of Lexical Semantics: a Cognitive Perspective*. Oxford: Oxford UP.
- Georgescu, Simona (2021). *La regularidad en el cambio semántico. Las onomatopeyas en cuanto centros de expansión en las lenguas románicas*. Strasbourg: ELiPhi.
- Georgescu, Simona / Georgescu Theodor (2020). 'Fr. *trouver*, occ. *trobar* etc.: un dossier étymologique ouvert à nouveau', *RLiR* 84: 83-98.
- Gévaudan, Paul (2013). 'Les rapports entre la modalité et la polyphonie linguistique', in *id.*, Detges, Ulrich / Atayan, Vahram (eds), *Modalität und Polyphonie. Modalité et polyphonie. Modalidad y polifonía*. Tübingen: Stauffenburg, 39-59.
- Gilliéron, Jules ([ca. 1880]). *Petit atlas phonétique du Valais roman, (sud du Rhône)*. Paris: Champion.
- Gilliéron, Jules (1918). *Généalogie des mots qui désignent l'abeille: d'après L'Atlas linguistique de la France*. Paris: Champion.
- Gilliéron, Jules / Edmont, Edmond (1912). *Table de l'Atlas Linguistique de la France*. Paris: Champion.
- Gilliéron, Jules / Edmont, Edmond (1920). *A.L.F. Suppléments. Tome premier*, Paris: Champion.
- Gilliéron, Jules / Mongin, Jean (1905). *Étude de géographie linguistique. «Scier» dans la Gaule romane du Sud et de l'Est*. Paris: Champion.
- Giraud, Cédric / Renault, Jean-Baptiste / Tock, Benoît-Michel (eds) (2010), *La base de données ARTEM. Chartes originales antérieures à 1121 conservées en France*, <www.cn-telma.fr/originaux>.
- Glessgen, Martin (1996). *Die Falkenheilkunde des Moamin im Spiegel ihrer volgarizzamenti: Studien zur Romania Arabica*, 2 vols. Tübingen: Niemeyer.
- Glessgen, Martin ([1996] 1997a). 'Das Französische im Maghreb: Bilanz und Perspektiven der Forschung', *Rjb* 47: 28-63.
- Glessgen, Martin (1997b). 'Der Beitrag der kanarischen Ostinseln zur hispano-amerikanischen Sprachentwicklung', *ZrP* 113: 418-65.
- Glessgen, Martin (2000). 'Les manuels de linguistique romane, source pour l'histoire d'un canon disciplinaire', *RKXIV*: 189-259.
- Glessgen, Martin (2005). 'Diskurstraditionen zwischen pragmatischen Regeln und sprachlichen Varietäten', in Schrott, Angela / Völker, Harald (eds), *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*. Göttingen: Göttinger Universitätsverlag, 207-28.
- Glessgen Martin (2008). 'Les lieux d'écriture dans les chartes lorraines du XIII^e s.', *RLiR* 72: 413-540.
- Glessgen Martin (2011). 'Le statut épistémologique du lexème', *RLiR* 75: 391-468.
- Glessgen Martin (2012). 'Trajectoires et perspectives en scriptologie galloromane', *Medioevo Romanzo* 36: 5-23.
- Glessgen, Martin (2017). 'La genèse d'une norme en français au Moyen Âge: mythe et réalité du 'francien'', *RLiR* 81: 313-98.
- Glessgen, Martin (2019). 'L'apport des 'Inconnus' du FEW à la recherche étymologique', *Italia Dialettale* 80: 633-676.
- Glessgen, Martin (2021). 'Pour une histoire textuelle du gascon médiéval', *RLiR* 85: 325-384.
- Glessgen, Martin (2022a). 'L'étude linguistique du gascon médiéval: analyse scriptologique des genres textuels', *RLiR* 86: 35-94.
- Glessgen, Martin (2022b). 'La langue des premiers troubadours. Une approche scriptologique', *Medioevo Romanzo* 46: 265-329.
- Glessgen, Martin / Giolitto, Marco (2007). 'Los vectores de la romanística a través del tiempo', in Gargallo Gil, José Enrique / Bastardas, Maria Reina (eds), *Manual de lingüística románica*. Barcelona: Ariel, 375-93.
- Glessgen, Martin / Lebsanft, Franz (eds) (1997). *Alte und neue Philologie*. Tübingen: Niemeyer.
- Glessgen, Martin / Schøsler, Lene (2018). 'Repenser les axes diasystématiques: nature et statut ontologique', in Glessgen, Martin / Kabatek, Johannes / Völker, Harald (eds), *Repenser la variation linguistique: Actes du Colloque DIA IV à Zürich (12-14 sept. 2016)*. Strasbourg: ELiPhi.

- Glessgen, Martin / Schweickard, Wolfgang (eds) (2002). *Ex traditione innovatio. Miscellanea in honorem Max Pfister septuagenarii oblata*. Vol. 1: *Max Pfister. Scripta minora selecta. De rebus Galloromanicis et Italicis*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Glessgen, Martin / Schweickard, Wolfgang (eds) (2014). *Étymologie romane: objets, méthodes et perspectives*. Strasbourg: SLiR/ÉLiPhi.
- Glessgen, Martin / Thibault, André (eds) (2005). *La lexicographie différentielle du français et le Dictionnaire des régionalismes de France*. Actes du colloque en l'honneur de Pierre Rézeau. Strasbourg: PU Strasbourg.
- Glessgen, Martin / Trotter, David (eds) (2016). *La régionalité lexicale au Moyen Âge*. Strasbourg: ELiPhi.
- Goebel, Hans (1984). *Dialektometrische Studien anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, 3 vols. Tübingen: Niemeyer.
- Goebel, Hans (2003). 'Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantificatives et structures de profondeur', *ER* 20: 59–118.
- Goebel, Hans (2008). 'La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati', *RLiR* 72: 25–113.
- Gosselin, Laurent (1996). *Sémantique de la temporalité en français. Un modèle calculatoire et cognitif du temps et de l'aspect*. Louvain-la-Neuve: Duculot.
- Gosselin, Laurent (2005). *Temporalité et modalité*. Bruxelles: Duculot-De Boeck.
- Gougenheim, Georges / Michéa, René / Rivenc, Paul / Sauvageot, Aurélien (1956). *L'élaboration du français élémentaire*. Paris: Didier.
- Grelois, Emmanuel / Chambon, Jean-Pierre (2008). *Les noms de lieux antiques et tardo-antiques d'Augustonemetum/Clermont-Ferrand*. Strasbourg: SLiR.
- Grevisse, Maurice (¹³1993). *Le Bon Usage*. Paris; Louvain-la-Neuve: Duculot [¹1936].
- Grice, Herbert P. (1975). 'Logic and conversation', in Cole, Peter / Morgan, Jerry L. (eds), *Speech acts, syntax and semantics*, vol. 3. New York: Academic Press, 41–58.
- Grice, Herbert P. (1991). *Studies in the way of words*. Harvard et al.: HUP.
- Grimm, Reinhold et al. (eds) (1991). *Mittelalter-Rezeption: zur Rezeptionsgeschichte der romanischen Literaturen*. Heidelberg: Winter.
- Gröber, Gustav (^{1/2}1902–1906). *Grundriss der romanischen Philologie*, 2 vols. Strassburg: Trübner (¹1886–1888, vol. 1: linguistique, new edition ²1904–1906; 1902 vol. 2: histoire littéraire; reprint. Berlin: de Gruyter, 1985).
- Gross, Gaston (1989). *Les Constructions converses du français*. Geneva: Droz.
- Gross, Gaston et al. (2009). *La sémantique de la cause*. Leuven: Peeters.
- Grossmann, Maria / Rainer, Franz (eds) (2004). *La formazione delle parole in italiano*. Tübingen: Niemeyer.
- Gruber, Teresa / Grübl, Klaus / Scharinger, Thomas (eds) (2021). *Was bleibt von kommunikativer Nähe und Distanz? Mediale und konzeptionelle Aspekte sprachlicher Variation*. Tübingen: Narr.
- Guéricos, Rosário Farâni Mansur (1979). *Dicionário de etimologias da língua portuguesa*. São Paulo: Companhia editora nacional.
- Haerle, Phillip (1955). *Captivus – cattivo – chétif; zur Einwirkung des Christentums auf die Terminologie der Moralbegriffe*. Bern: Francke.
- Halliday, Michael A. K. (1967, 1968). 'Notes on transitivity and theme in English', *Journal of Linguistics* 3, 37–81, 199–244; 4, 179–215.
- Halliday, Michael A. K. (1978). *Language as social semiotic*. London: Edward Arnold.
- Hallig, Rudolf / Wartburg, Walther von (²1963). *Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas: Système raisonné des concepts pour servir de base à la lexicographie*. Berlin: Akademie-Verlag [¹1952].
- Harris, Martin / Vincent, Nigel (eds) (1988). *The Romance Languages*. London; Sydney: Helm (reprint. Routledge, 1997).
- Hauser, Marc D. / Chomsky, Noam / Fitch, Tecumseh (2002). 'The faculty of language: what is it, who has it, and how did it evolve?', *Science* 298, 1569–79.

- Hausmann, Frank R. (2000). *'Vom Strudel der Ereignisse verschlungen': deutsche Romanistik im 'Dritten Reich'*. Frankfurt am Main: Klostermann.
- Heger, Klaus (1967). *Die Bibel in der Romania: Matthäus 6, 5–13*. Tübingen: Niemeyer.
- Heger, Klaus (²1976). *Monem, Wort, Satz und Text*. Tübingen: Niemeyer [¹1971].
- Heinimann, Siegfried (1988). *Oratio dominica romanice: das Vaterunser in den romanischen Sprachen von den Anfängen bis ins 16. Jahrhundert mit den griechischen und lateinischen Vorlagen*. Tübingen: Niemeyer.
- Herman, Jozsef (1990). *Du latin aux langues romanes: études de linguistique historique (I)*. Tübingen: Niemeyer.
- Herman, Jozsef (2005). *Du latin aux langues romanes II: Nouvelles études de linguistique historique*. Tübingen: Niemeyer.
- Hernanz, María L. / Brucart, Josep M. (1987). *La sintáxis: principios teóricos*. Barcelona: Ed. Crítica.
- Hickok, Gregory / Poeppel, David (2007). 'The cortical organization of speech processing', *Nature Reviews Neuroscience* 8/5, 393–402.
- Hilmer, Hermann (1914). *Schallnachahmung, Wortschöpfung und Bedeutungswandel: auf Grundlage der Wahrnehmungen von Schlag, Fall, ... dargestellt an einigen Lautwurzeln der deutschen und der englischen Sprache*. Halle, Niemeyer.
- Hilmer, Hermann (1918). «The origin and growth of language», *The Journal of English and Germanic Philology* 17, 21–60.
- Hilty, Gerold (2001). 'Das *Indovinello veronese* ist kein Rätsel', in Wunderli, Peter / Werlen, Iwar / Grünert, Matthias (eds), *Italica–Raetica–Gallica. Studia linguarum litterarum artiumque in honorem Ricarda Liver*. Tübingen; Basel: Francke, 293–301.
- Hjelmslev, Louis (1961). *Prolegomena to a Theory of Language*, Revised English Edition. Madison: University of Wisconsin Press (trans. by Francis J. Whitfield from the Danish *Omrking sprogteoriens grundlaeggelse*, Copenhagen: Akademisk Forlag, 1943).
- Höfler, Manfred (1982). *Dictionnaire des anglicismes*. Paris: Larousse.
- Holtus, Günter (1985). *La versione franco-italiana della Bataille d'Aliscans: codex Marciianus fr. VIII [252]: testo con introd., note e glossario*. Tübingen: Niemeyer.
- Hudson, Richard A. (1980). *Sociolinguistics*. Cambridge: CUP.
- Huizinga, Johan (1996). *The Autumn of the Middle Ages* (trans. R.J. Payton and U. Mammitzsch). Chicago, University of Chicago Press (Dutch original: *Herfsttij der Middeleeuwen*, Haarlem, 1919).
- Humboldt, Wilhelm von (1988). *On Language: The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*. Cambridge: CUP (German original: *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Schneider, 1836).
- Jaberg, Karl (1960). *Index zum Sprach- und Sachatlas Italiens und der Südschweiz*. Bern: Stämpfli.
- Jaberg, Karl / Jud, Jakob (1928). *Der Sprachatlas als Forschungsinstrument: kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz*. Halle: Niemeyer.
- Jud, Jakob (1934). 'Sur l'histoire de la terminologie ecclésiastique de la France et de l'Italie', *RLiR* 10, 1–62 [= Jud, 1973, 213–78].
- Jud, Jakob (1973). *Romanische Sprachgeschichte und Sprachgeographie: ausgewählte Aufsätze*. Konrad Huber et al. (eds). Zürich: Atlantis.
- Juneau, Marcel (1976). *La jument qui crotte de l'argent: conte populaire recueilli aux Grandes-Bergeronnes (Québec)*. Quebec: PU Laval.
- Kabatek, Johannes (2005). *Die Bolognesische Renaissance und der Ausbau romanischer Sprachen. Juristische Diskurstraditionen und Sprachentwicklung in Südfrankreich und Spanien im 12. und 13. Jahrhundert*. Tübingen: Niemeyer.
- Kaehlbrandt, Roland (1989). *Syntaktische Entwicklungen in der Fachsprache der französischen Wirtschaftswissenschaften, untersucht an der Textsorte 'Lehrwerk' im Zeitraum von 1815–1984*. Stuttgart: Steiner.
- Keller, Rudi (²1994). *Sprachwandel. Von der unsichtbaren Hand in der Sprache*. Tübingen; Basel: Francke.
- Kerbrat Orecchioni, Catherine (1992–1995). *Les Interactions verbales*, 3 vols. Paris: Colin.

- Klapp, Otto / Klapp-Lehmann Astrid (1960–). *Bibliographie der französischen Literaturwissenschaft*. Frankfurt am Main: Klostermann.
- Kleiber, Georges (1981). *Problèmes de référence: descriptions définies et noms propres*. Paris: Klincksieck.
- Kleiber, Georges (1990). *La Sémantique du prototype. Catégories et sens lexical*. Paris: PUF.
- Klein, Franz-Josef (1997). *Bedeutungswandel und Sprachendifferenzierung. Die Entstehung der romanischen Sprachen aus wortsemantischer Sicht*. Tübingen: Niemeyer.
- Klemperer, Victor (2010). *LTI – Notizbuch eines Philologen*. Stuttgart: Reclam [1947].
- Klinkenberg, Jean-Marie (21999). *Des langues romanes*. Louvain-la-Neuve: Duculot [1994].
- Kloss, Heinz (21978, 1952). *Die Entwicklung neuer germanischer Kultursprachen seit 1800*. Düsseldorf: Schwann.
- Koch, Peter (1993). 'Pour une typologie conceptionnelle et médiale des plus anciens documents/monuments des langues romanes', in Selig, Maria / Frank, Barbara / Hartmann, Jörg (eds), *Le Passage à l'écrit des langues romanes*. Tübingen: Narr, 39–81.
- Koch, Peter (2012). 'Es gibt keine Konstruktionsbedeutung ohne Bedeutungswandel. Valenz – Konstruktion – Diachronie', in Dörr, Stephen / Städtler, Thomas (eds), *Ki bien voldreit raison entendre. Mélanges en l'honneur du 70^e anniversaire de Frankwalt Möhren*. Strasbourg: SLiR/ELiPhi, 147–74.
- Koch, Peter / Oesterreicher, Wulf (1985). 'Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte', *Rjb* 36, 15–43.
- Koch, Peter / Oesterreicher, Wulf (1994). 'Schriftlichkeit und Sprache' in Günther, Hartmut / Ludwig, Otto (eds), *Schrift und Schriftlichkeit* (HSK 10.1), vol. 1. Berlin/New York: de Gruyter, 587–604.
- Koch, Peter / Oesterreicher, Wulf (1996). 'Sprachwandel und expressive Mündlichkeit', *Zeitschrift für Literaturwissenschaft und Linguistik* 102, 64–96.
- Koch, Peter / Oesterreicher, Wulf (21991). *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*, Berlin; New York: de Gruyter [1990] (Spanish trans. of the first ed. *Lengua hablada en la Romania: español, francés, italiano*, Madrid: Gredos, 2007).
- Kramer, Johannes (1986). *English and Spanish in Gibraltar*. Hamburg: Buske.
- Kramer, Johannes (1998). *Die Sprachbezeichnungen Latinus und Romanus im Lateinischen und Romanischen*. Berlin: Schmidt.
- Kramer, Johannes (1999). 'Sind die romanischen Sprachen kreolisiertes Latein?', *ZrP* 115, 1–19.
- Kramer, Johannes (1999–2000). 'Sprachwissenschaft und Politik. Die Theorie der Kontinuität des Rumänischen und der balkanische Ethno-Nationalismus im 20. Jh.', *Balkan-Archiv* 24/25, 105–63.
- Kramer, Johannes (2000). 'Die Zahl der Sprecher der Sprache Roms in Geschichte und Gegenwart', *RGG* 6, 1–31.
- Krefeld, Thomas (1987). *Das französische Gerichtsurteil in linguistischer Sicht: Zwischen Fach- und Standessprache*. Frankfurt am Main: Lang.
- Kuhn, Alwin (1951). *Romanische Philologie*. Bern: Francke.
- Laca, Brenda (1986). *Die Wortbildung als Grammatik des Wortschatzes*. Tübingen: Niemeyer.
- Laca, Brenda / Bertolotti, Virginia (forthcoming). *Orden de constituyentes*, in Concepción Company Company (dir.), *Sintaxis histórica de la lengua española: Cuarta parte: Orden de constituyentes, estructura argumental y discurso. Cambio, tradiciones y soportes textuales* (2 vol.), México, D.F.: Fondo de Cultura Económica.
- Lafage, Suzanne (1985). *Français écrit et parlé en pays Éwé (Sud-Togo)*. Paris: SELAF.
- La Fauci, Nunzio (2009). *Compendio di sintassi italiana*. Bologna: il Mulino.
- Lanher, Jean (1983). *Les Contes de Fraimbois*. Nancy: PU Nancy.
- Lausberg Heinrich (21967–1972). *Romanische Sprachwissenschaft*, 4 vols. Berlin: de Gruyter (Spanish trans. *Lingüística románica*, 2 vols., 1966–70).
- Lebsanft, Franz (1982). 'Hören und Lesen im Mittelalter', *ZfSL* 92, 52–64.
- Lebsanft, Franz (1987). *Studien zu einer Linguistik des Grußes. Sprache und Funktion der altfranzösischen Grußformeln*. Tübingen: Niemeyer.
- Lebsanft, Franz (1990). *Spanien und seine Sprachen in den Cartas al Director von El País (1976–1987)*. Tübingen: Narr.

- Lebsanft, Franz (1997). *Spanische Sprachkultur: Studien zur Bewertung und Pflege des öffentlichen Sprachgebrauchs im heutigen Spanien*. Tübingen: Niemeyer.
- Lebsanft, Franz (1999 [2001]). ‘Lingüística popular’ y cultivo del idioma en Internet. Los resultados de una encuesta a participantes de la lista de correo ‘Apuntes’, *EA* 72, 47–58.
- Lebsanft, Franz (2000). ‘Die eigene und die fremden Sprachen in romanischen Texten des Mittelalters und der frühen Neuzeit’, *RK* XIII, 3–20.
- Lebsanft, Franz (2005). ‘Kommunikationsprinzipien, Texttraditionen, Geschichte’, in Schrott, Angela / Völker, Harald (eds), *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*. Göttingen: Göttinger Universitätsverlag, 25–43.
- Lebsanft, Franz / Glessgen, Martin (2004). *Historische Semantik in den romanischen Sprachen*. Tübingen: Niemeyer.
- Ledgeway, Adam (2009). *Grammatica diacronica del napoletano*. Tübingen: Niemeyer.
- Ledgeway, Adam (2016). ‘From coordination to subordination: The grammaticalisation of progressive and andative aspect in the dialects of Salento’, in Pratas, Fernanda / Pereira, Sand / Pinto, Clara (eds), *Coordination and Subordination. Form and Meaning. Selected Papers from CSI Lisbon 2014*. Newcastle: Cambridge Scholars Publishing, 157–84.
- Ledgeway, Adam (2017). ‘The Romanian definite article in a comparative Romance perspective’, in Dragomirescu, Adina / Nicolae, Alexandru / Stan, Camelia / Zafiu, Rodica (eds), *Sintaxa ca mod de a fi. Omagiu doamnei profesoare Gabriela Pană Dindelegan, la aniversare*. Bucharest: Editura Universității din București, 231–47.
- Ledgeway, Adam (2021). ‘V2 beyond borders: The *Histoire Ancienne jusqu’à César*’, in Meklenborg, Christine / Wolfe, Sam (eds), *Secrets of Success*, special issue of *Journal of Historical Syntax* 5(29), 1–65.
- Ledgeway, Adam (2023). ‘Aspectual and irrealis marking: the distribution of the Wallon *temps surcomposés*’, in Tommaso, Balsemin *et al.* (eds), special of *Quaderni di lavoro ASIt*.
- Ledgeway, Adam / Maiden, Martin (2022) → *CambrHandb*, 13.1
- Ledgeway, Adam / Smith, John Charles / Vincent, Nigel (eds) (2022). *Periphrasis and Inflection in Diachrony: A View from Romance*, Oxford: Oxford UP.
- Leonardi, Lino / Minervini, Laura / Burgio, Eugenio (2022). *Filologia romanza. Critica del testo, linguistica, analisi letteraria*. Florence: Mondadori.
- Le Roy Ladurie, Emmanuel (1975). *Montaillou, village occitan de 1294 à 1324*. Paris: Gallimard.
- Le Roy Ladurie, Emmanuel (³2003). *Histoire du climat depuis l’an mil*. Paris: Flammarion [¹1967].
- Lodge, R. Anthony (2004). *A sociolinguistic history of Parisian French*. Cambridge: CUP.
- Loporcaro, Michele (1988). *Grammatica storica del dialetto di Altamura*. Pisa: Giardini.
- Loporcaro, Michele (2013). *Profilo linguistico dei dialetti italiani*. Rome: Laterza.
- Machado, José Pedro (²1967). *Dicionário etimológico da língua portuguesa [...]*, 5 vols. Lisbon: Confluência.
- Maiden, Martin (2000). ‘Il sistema desinenziale del sostantivo nell’italo-romanzo preletterario. Ricostruzione parziale a partire dai dialetti moderni (il significato storico di plurali del tipo “amici”); in Herman, József / Marinetti, Anna (eds), *La preistoria dell’italiano*, Atti della tavola rotonda di linguistica storica (Università Ca’ Foscari di Venezia, 11–13 giugno 1998). Tübingen: Niemeyer, 167–179.
- Maiden, Martin (2016). ‘The Romanian alternating gender in diachrony and synchrony’, *Folia Linguistica Historica* 37, 111–144.
- Manzini, M. Rita / Savoia, Leonardo M. (2005), *I dialetti italiani e romanci. Morfosintassi generativa*, 3 vols. Alessandria: Edizioni dell’Orso, 2005.
- Marchello-Nizia, Christiane (1985). *Dire le vrai: l’adverbe «si» en français médiéval: essai de linguistique historique*. Geneva: Droz.
- Marchello-Nizia, Christiane / Combettes, Bernard / Prévost, Sophie / Scheer, Tobias (eds) (2021). *Grande Grammaire historique du français (GGHF)*. Berlin; Boston: de Gruyter-Mouton.
- Martineau, France / Remysen, Wim / Thibault, André (2022). *Le français au Québec et en Amérique du Nord*. Paris: Ophrys.

- Martinet, André (²1970). *Économie des changements phonétiques: traité de phonologie diachronique*. Bern: Francke [¹1955].
- Matthews, Peter H. (²1991). *Morphology: An introduction to the theory of word-structure*. Cambridge: CUP [¹1974].
- Mayr, Ernst (2001). *What Evolution is*. New York: Basic Books.
- Melis, Federico (1962). *Aspetti della vita economica medievale: studi nell'archivio Datini di Prato*. Siena: Monte dei Paschi di Siena.
- Menéndez Pidal, Ramón (¹1926). *Orígenes del español. Estado lingüístico de la Península ibérica hasta el siglo XI*. Madrid: Hernando.
- Menéndez Pidal, Ramón (1929). *La España del Cid*, 3 vols. Madrid: Plutarco.
- Messner, Dieter (1976). *Dictionnaire chronologique des langues ibéroromanes*, vol. 1: *Dictionnaire chronologique portugais*. Heidelberg: Winter.
- Meyer-Lübke, Wilhelm (1890–1902). *Grammatik der romanischen Sprachen*, 4 vols. Leipzig: Fues.
- Migne, Jacques Paul (ed.) (1879–1890). *Patrologia latina*, 221 vols. Paris: Garnier.
- Möhren, Frankwalt (1986). *Wort- und sachsgechichtliche Untersuchungen an französischen landwirtschaftlichen Texten 13., 14. und 18. Jahrhundert: Seneschaucaie, Menagier, Encyclopédie*. Tübingen: Niemeyer.
- Morlet, Marie-Thérèse (1968–1985). *Les Noms de personne sur le territoire de l'ancienne Gaule du vi^e au xi^e siècle*, 3 vols. Paris: CNRS.
- Morlet, Marie-Thérèse (²1997). *Dictionnaire étymologique des noms de famille*. Paris: Perrin [¹1991].
- Morf, Heinrich (1911). *Zur sprachlichen Gliederung Frankreichs*. Berlin: Preussische Akademie der Wissenschaften.
- Morris, Charles (1938). *Foundations of the theory of signs*. Chicago: Chicago UP.
- Müller, Bodo (1985). *Le Français d'aujourd'hui*. Paris: Klincksieck (Original German edition: *Das Französische der Gegenwart*. Heidelberg: Winter, 1975).
- Müller, Bodo (1994). 'Die 'Sprachen' der romanischen Sprachwissenschaft', in Richard Baum *et al.* (eds), *Lingua et Traditio, Festschrift für H.H. Christmann zum 65. Geburtstag*. Tübingen: Narr, 643–50.
- Nègre, Ernest (1990–1991 + 1998). *Toponymie générale de la France*, 3 vols. + 1 vol. *Errata et addenda*. Geneva: Droz.
- Nisbet, Hugh B. / Rawson, Claude (eds) (1997). *The Cambridge History of Literary Criticism. Volume IV: The Eighteenth Century*. Cambridge: CUP.
- Noomen, Willem (1983–1998). *Nouveau recueil complet des fabliaux (NRCF)*, 10 vols. Assen: Van Gorcum.
- Nyrop, Kristoffer (1903–1930). *Grammaire historique de la langue française*, 6 vols. Copenhagen: Gyldendal.
- Ogden, Charles K. / Richards, Ivor A. (1923). *The Meaning of Meaning: a Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul.
- Origo, Iris (1960). *The merchant of Prato*. London: Cape.
- Parry, M. Mair (2005). *Sociolinguistica e grammatica del dialetto di Cairo Montenotte. Parluma 'd còiri*, Savona: Società savonese di Storia Patria.
- Parry, M. Mair (2007). 'The Interaction of Semantics and Syntax in the Spread of Relative *che* in the Early Vernaculars of Italy', in Bentley, Delia / Ledgeway, Adam (eds), *Sui dialetti italo-romanzi. Saggi in onore di Nigel B. Vincent*. Norfolk: Biddles 200–19.
- Pei, Mario A. (1976). *The story of Latin and the Romance Languages*. New York: Harper & Row.
- Perlmutter David (1978). 'Impersonal passives and the unaccusative hypothesis', in *id.*, *Proceedings of the 4th Annual meeting of the Berkeley linguistic Society*, 157–89.
- Petrucchi, Livio (1994). 'Il problema delle Origini e i più antichi testi italiani', in Serianni, Luca / Trifone, Pietro (eds), *Storia della lingua italiana*, 3. *Le altre lingue*. Turin: Einaudi, 5–73.
- Pfister, Max (1970). *Lexikalische Untersuchungen zu Girart de Roussillon*. Tübingen: Niemeyer.
- Pfister, Max (1983, 1985). 'Compte rendu de Anonimo Romano, Cronica, edizione critica a cura di G. Porta', *ZrP* 99: 526–29; 'Replica a Giuseppe Porta', *StMed* III. 26, 365–68.
- Pfister, Max (1987). 'Zur Chronologie der Palatalisierungserscheinungen in der östlichen Galloromania', in Lüdi, Georges / Stricker, Hans / Wüest, Jakob (eds), *Romania ingeniosa, Festschrift für Gerold Hilty zum 60. Geburtstag*. Bern: Lang, 179–90 [= Pfister, 2002, 409–20].

- Pfister, Max (2002). *Scripta minora selecta* → Glessgen/Schweickard 2002.
- Pfister, Max / Lupis, Antonio (2001). *Introduzione all'etimologia romanza*. Catanzaro: Rubbettino (new version of M. Pfister (1980). *Einführung in die romanische Etymologie*. Darmstadt: Wiss. Buchgesellschaft).
- Picoche, Jacqueline / Marchello, Nizia Christiane (²1994). *Histoire de la langue française*. Paris: Nathan.
- Pillet, Alfred / Carstens, Henry (eds) (1933). *Bibliographie der Troubadours*. Halle: Niemeyer.
- Pinker, Steven (1994). *The language instinct*. New York: Harper Collins.
- Pöckl, Wolfgang / Rainer, Franz / Pöll, Bernhard (²2003). *Einführung in die romanische Sprachwissenschaft*. Tübingen: Niemeyer.
- Poplack, Shana (1980). 'Sometimes I'll start a sentence in English y termino en español: toward a typology of code-switching', *Linguistics* 18, 581–618.
- Posner, Rebecca (1996). *The Romance Languages*. Cambridge: CUP (reprint. 2002; Spanish trans.: *Las lenguas romances*, Madrid: Cátedra, 1998).
- Pottier, Bernard (1963). *Recherches sur l'analyse sémantique en linguistique et en traduction mécanique*. Nancy: PU Nancy.
- Prati, Angelico (1931). *I vocabolari delle parlate italiane*. Rome: Caponera.
- Puscariu, Sextil (²1975). *Etymologisches Wörterbuch der rumänischen Sprache*, vol. 1: *Lateinisches Element*. Heidelberg: Winter.
- Pusch, Claus / Raible, Wolfgang (eds) (2002). *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache/ Romance Corpus Linguistics: Corpora and Spoken Language*. Tübingen: Narr.
- Pusch, Claus / Kabatek, Johannes / Raible, Wolfgang (eds) (2005). *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft*. Tübingen: Narr.
- Quadri, Bruno (1952). *Aufgaben und Methoden der onomasiologischen Forschung*. Bern: Francke.
- Radermacher, Ruth (1998). *Le 'Trésor de la langue française'. Une étude historique et lexicographique*, available on the website of the ATILF.
- Raible, Wolfgang (1983). 'Zur Einleitung', in Stimm, Helmut / Raible, Wolfgang (eds), *Zur Semantik des Französischen*. Wiesbaden: Steiner, 1–24.
- Raible, Wolfgang (1991). *Die Semiotik der Textgestalt. Erscheinungsformen und Folgen eines kulturellen Evolutionsprozesses*. Heidelberg: Winter.
- Raible, Wolfgang (2004). '¿Qué es un texto?', *Función* 21–4, 9–29.
- Raible, Wolfgang (2006). *Medien-Kulturgeschichte. Mediatisierung als Grundlage unserer kulturellen Entwicklung*. Heidelberg: Winter.
- Rainer, Franz (1993). *Spanische Wortbildungslehre*. Tübingen: Niemeyer.
- Real Academia española/Asociación de academias de la lengua española (2009). *Nueva Gramática de la lengua española*, 2 vols. (vol. 1: *Morfología y sintaxis I.*, vol. 2: *Sintaxis II.*). Madrid: Espasa Libros.
- Rembert, Eufe (2012). 'Merovingian Coins and Their Inscriptions: A Challenge to Linguists and Historians', in Langer, Nils / Davies, Steffan / Vandenbussche, Wim (eds), *Language and History, Linguistics and Historiography. Interdisciplinary Approaches*. Oxford et al.: Lang, 293–322.
- Renson, Jean (1962). *Les dénominations du visage en français et dans les autres langues romanes: étude sémantique et onomasiologique*. Paris: Belles-Lettres.
- Renzi, Lorenzo (²2001). *Grande grammatica italiana di consultazione*, 3 vols. Bologna: il Mulino [¹1988–1995].
- Renzi, Lorenzo / Andreose, Alvisè (2003). *Manuale di linguistica e filologia romanza*. Bologna: il Mulino.
- Renzi, Lorenzo / Salvi, Giampaolo (²1994). *Nuova introduzione alla filologia romanza*. Bologna: il Mulino, (Spanish trans. of first edition: *Introducción a la filología románica*, Madrid: Gredos, 1982; cf. Renzi/Alvisè 2003).
- Revol, Thierry (2000). *Introduction à l'ancien français*. Paris: Nathan.
- Rézeau, Pierre (2013). *Les Noël's en France aux XV^e et XVI^e siècles. Édition et analyse*. Strasbourg: ELiPhi/SLR.
- Rézeau, Pierre (2018). *Les Mots des Poilus*. Strasbourg: ELiPhi.
- Rézeau, Pierre (2023). *L'Etymologicon françois* de Jean le Bon (1572). Strasbourg: ELiPhi.
- Rheinfelder, Hans (1928). *Das Wort 'Persona': Geschichte seiner Bedeutungen mit besonderer Berücksichtigung des französischen und italienischen Mittelalters*. Halle: Niemeyer.

- Richter, Elise (1934). *Beiträge zur Geschichte der Romanismen, I. Chronologische Phonetik des Französischen bis zum Ende des 8. Jahrhunderts*. Halle: Niemeyer.
- Riegel, Martin / Pellat, Jean-Christophe / Rioul, René (2009). *Grammaire méthodique du français*. Paris: PUF [1994].
- Rieger, Angelica (1993). 'Les troubadours-fantômes en Italie', in Giuliano Gasca Queirazza (ed.), *Atti del Secondo Congresso Internazionale della 'Association Internationale d'Études Occitanes'*, 2 vols. Turin: Dipartimento di Scienze Letterarie e Filologiche, vol. 1, 327–47.
- Rohlf, Gerhard (1937). *La struttura linguistica dell'Italia*. Leipzig: Keller.
- Rohlf, Gerhard (1966–1969). *Grammatica storica della lingua italiana e dei suoi dialetti*, 3 vols (I. *Fonetica*. II. *Morfologia*. III. *Sintassi e formazione delle parole*). Torino: Einaudi (new ed. Bologna: il Mulino, 2021).
- Rohlf, Gerhard (1971). *Romanische Sprachgeographie. Geschichte und Grundlagen, Aspekte und Probleme mit dem Versuch eines Sprachatlas der romanischen Sprachen*. München: Beck.
- Rohlf, Gerhard (1986). *Panorama delle lingue neolatine. Piccolo atlante linguistico panromanzo*. Tübingen: Narr.
- Roques, Gilles (1999). 'Parler d'autre Martin', *TraLiPhi* 37, 109–22.
- Rossebastiano, Bart Alda (1983). *Vocabolari veneto-tedeschi del secolo xv*. Rome: L'Artistica savigliano.
- Roudet, Léonce (1921). 'Sur la classification psychologique des changements sémantiques', *Journal de Psychologie* 18, 676–92.
- Sabatini, Francesco (1968). 'Dalla 'scripta latina rustica' alle 'scriptae' romanze', *StMed* 9, 320–58.
- Salvi, Giampaolo / Renzi, Lorenzo (2010). *Grammatica dell'italiano antico*, 2 vols. Bologna: il Mulino.
- Sapir, Edward (1921). *Language: an introduction to the study of speech*. New York: Harcourt, Brace and Co.
- Saussure, Ferdinand de (1959). *Course in General Linguistics* (English trans. by W. Baskin). New York: Philosophical Library (French original: *Cours de linguistique générale*, Lausanne; Paris: Payot, 1916; critical edition by Rudolph Engler, Wiesbaden: Harrassowitz, 3 fasc., 1967–1968).
- Savj-Lopez, Paolo (1920). *Le Origini neolatine*. Milan: Hoepli.
- Schaden, Gerhard (2007). *La Sémantique du parfait. Étude des «temps composés» dans un choix de langues germaniques et romanes*. Université Paris 8: doctoral thesis.
- Scheuermeier, Paul (1943–1956). *Bauernwerk in Italien, der italienischen und rätoromanischen Schweiz: eine sprach- und sachkundliche Darstellung landwirtschaftlicher Arbeiten und Geräte*, 2 vols. Erlenbach: Rentsch; Bern: Stämpfli.
- Schlegel, August Wilhelm von (1818). *Observations sur la langue et la littérature provençales*. Paris (repr. Tübingen: Narr, 1971).
- Schlemmer, Gerd (1983). *Die Rolle des germanischen Superstrats in der Geschichte der romanischen Sprachwissenschaft*. Hamburg: Buske.
- Schlieben-Lange, Brigitte (1996). *Idéologie, révolution et uniformité de la langue*, Mardaga: Sprimont.
- Schmid, Stephan (2002). *Einführung in die allgemeine Phonetik und Phonologie für Studierende der Romanistik. Materialien*. Zürich: Phonetisches Laboratorium [unpub.].
- Schmitt, Christian (1974). *Die Sprachlandschaften der Galloromania: eine lexikalische Studie zum Problem der Entstehung und Charakterisierung*. Bern: Lang.
- Schmitt, Christian (1996). 'Euromorphologie: Perspektiven einer neuen romanistischen Teildisziplin', *RK* IX, 469–82.
- Schrott, Angela / Völker, Harald (eds) (2005). *Historische Pragmatik und historische Varietätenlinguistik in den romanischen Sprachen*. Göttingen: Göttinger Universitätsverlag.
- Schwarze, Christoph (1988). *Grammatik der italienischen Sprache*. Tübingen: Niemeyer.
- Schweickard, Wolfgang (1985). 'Etymologia est origo vocabulorum'. Zum Verständnis der Etymologiedefinition Isidors von Sevilla', *HistLing* 12, 1–25.
- Schweickard, Wolfgang (1987). *Die 'cronaca calcistica'. Zur Sprache der Fußballberichterstattung in italienischen Sporttageszeitungen*. Tübingen: Niemeyer.
- Schweickard, Wolfgang (1990). '«Il se croit de la côte d'Adam». Reflexe der biblischen Vorstellungswelt im übertragenen Sprachgebrauch der romanischen Sprachen', *RK* III, 23–45.

- Schweickard, Wolfgang (1992). *'Deonomastik'. Ableitungen auf der Basis von Eigennamen im Französischen unter vergleichender Berücksichtigung des Italienischen, Rumänischen und Spanischen*. Tübingen: Niemeyer.
- Searle, John (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Seidl, Christian (1995). 'Le système acasuel des protoromans ibérique et sarde: dogmes et faits', *VR* 54, 41–73.
- Selinker, Larry (1969). 'Language Transfer', *General Linguistics* 9, 67–92.
- Serianni, Luca (1988). *Grammatica italiana: italiano comune e lingua letteraria: suoni, forme, costrutti*. Turin UTET.
- Serianni, Luca (1991). *Il secondo Ottocento: dall'Unità alla guerra mondiale*. Bologna: il Mulino.
- Serianni, Luca / Trifone, Pietro (eds) (1993–1994). *Storia della lingua italiana*, vol. 1: *I luoghi della codificazione* (1993); vol. 2: *Scritto e parlato* (1994); vol. 3: *Le altre lingue* (1994). Turin: Einaudi.
- Söll, Ludwig (1974). *Gesprochenes und geschriebenes Französisch*. Berlin: Schmidt.
- Sombart, Werner (1922). *Luxus und Kapitalismus*. Munich: Duncker & Humblot (new edition: *Liebe, Luxus und Kapitalismus. Über die Entstehung der modernen Welt aus dem Geist der Verschwendung*, Berlin: Wagenbach, 1986).
- Stark, Elisabeth (1997). *Vorstellungsstrukturen und topic-Markierung im Französischen. Mit einem Ausblick auf das Italienische*. Tübingen: Narr.
- Stefenelli, Arnulf (1981). *Geschichte des französischen Kernwortschatzes*. Berlin: Schmidt.
- Stefenelli, Arnulf (1992). *Das Schicksal des lateinischen Wortschatzes in den romanischen Sprachen*. Passau: Rothe.
- Stein, Achim (2005). *Semantische Repräsentation italienischer Verben: automatische Disambiguierung mit Konzepthierarchien*. Tübingen: Niemeyer.
- Stein, Achim (2010). *Einführung in die französische Sprachwissenschaft*. Stuttgart; Weimar: Metzler.
- Steiner, Linda (2016). *I centri di espansione nel cambio semantico. Per un'interpretazione cognitiva del Französiches Etymologisches Wörterbuch*. Strasbourg: ELiPhi.
- Stewart, William A. (1968). 'A sociolinguistic typology for describing national multilingualism', in Joshua A. Fishman (ed.), *Readings in the sociology of language*. The Hague: Mouton, 531–45.
- Straka, Georges (1979). *Les sons et les mots. Choix d'études de phonétique et de linguistique*. Paris: Klincksieck.
- Stussi, Alfredo (1994). *Introduzione agli studi di filologia italiana*. Bologna: il Mulino.
- Tagliavini, Carlo (1982). *Le Origini delle lingue neolatine. Introduzione alla filologia romanza*. Bologna: Patron [1949, 21952].
- Tappolet, Ernst (1895). *Die romanischen Verwandtschaftsnamen. Mit besonderer Berücksichtigung der französischen und italienischen Mundarten. Ein Beitrag zur vergleichenden Lexikologie*. Strasbourg: Trübner.
- Terreaux, Louis (1968). *Ronsard correcteur de ses œuvres: les variantes des 'Odes' et des deux premiers livres des 'Amours'*. Geneva: Droz.
- Tesnière, Lucien (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Thibault, André et al. (2019). *Nouveaux regards sur la variation dialectale*. Strasbourg: ELiPhi.
- Thornton, Anna (2005). *Morfologia*. Rome: Carocci.
- Tiktin, Hariton / Miron, Paul (2001–). *Rumänisch-deutsches Wörterbuch*, 3 vols. Wiesbaden: Harrassowitz, [21985–1989].
- Tock, Benoît-Michel (2005). *Scribes, souscripteurs et témoins dans les actes privés en France (VII^e–début XII^e siècle)*. Turnhout: Brepols.
- Tortora, Christina (2014). *A Comparative Grammar of Borgomanerese*. Oxford: Oxford UP.
- Tramater (1829–1840). *Vocabulario universale italiano, compilato a cura della Società tipografica Tramater*, 7 vols. Naples: Tramater.
- Troubetskoï, Nicolai S. (1949). *Principes de phonologie*. Paris: Klincksieck [1939].
- Trovato, Paolo (1991). *Con ogni diligenza corretto: la stampa e le revisioni editoriali dei testi letterari italiani (1470–1570)*. Bologna: il Mulino.
- Ullmann, Stephen (1952). *Précis de sémantique française*. Bern: Francke.
- Vachon, Claire (2010). *Le changement linguistique au XVI^e siècle*. Strasbourg: SLR/ELiPhi.

- Varvaro, Alberto (2001). *Linguistica romanza. Corso introduttivo*. Naples: Liguori (French trans. *Linguistique romane*, 2010, Namur: PU Namur).
- Vendler, Zeno (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell UP.
- Videsott, Paul (2009). *Padania scrittologica*. Tübingen: Niemeyer.
- Vidos, Benedek Elentheus (1956). *Handboek tot de Romaanse taalkunde*. 's Hertogenbosch, Malmberg (Italian trans.: *Manuale di linguistica romanza*. Florence: Olschki, 31974; Spanish trans. *Manual de lingüística románica*. Madrid: Aguilar, 21977).
- Vieliard, Françoise (2000). 'Compte rendu de l'*Inventaire systématique* [...]', *CCM* 43, 294–98.
- Vincent, Auguste (21984). *Toponymie de la France*. Brussels: Librairie générale [1937].
- Vincent, Nigel (2016). 'Causatives in Latin and Romance', in Adams, James N. / Vincent, Nigel (eds), *Early and Late Latin: Continuity or Change?* Cambridge: CUP, 294–312.
- Vitale, Maurizio (1992). *Studi di storia della lingua italiana*. Milan: LED.
- Völker, Harald (2003). *Skripta und Variation. Untersuchungen zur Negation und zur Substantivflexion in altfranzösischen Urkunden der Grafschaft Luxemburg (1237–1281)*. Tübingen: Niemeyer.
- Wartburg, Walther von (1936). 'Die Ausgliederung der romanischen Sprachräume', *ZrP* 56, 1–48.
- Wartburg, Walther von (1950). *Die Ausgliederung der romanischen Sprachräume*. Bern: Francke.
- Wartburg, Walther von (21970). *Einführung in Problematik und Methodik der Sprachwissenschaft*. Tübingen: Niemeyer ([1943], Halle: Niemeyer; French trans. *Problèmes et méthodes de la linguistique*, Paris: PUF, 11946, 31969; English trans. *Problems and methods in linguistics*, Oxford: Blackwell, 1969).
- Wartburg, Walther von / Keller, Hans Erich / Robert, Geuljans (21969). *Bibliographie des dictionnaires patois galloromans (1550–1967)*. Geneva: Droz.
- Weinhold, Norbert, (1985). *Sprachgeographische Distribution und chronologische Schichtung*. Hamburg: Buske.
- Weinreich, Uriel (1954). 'Is a structural dialectology possible?', *Word* 10, 288–400.
- Wolf, Heinz-Jürgen (21991). *Französische Sprachgeschichte*. Heidelberg: Quelle & Meyer.
- Wolf, Lothar (1975). *Aspekte der Dialektologie: eine Darstellung von Methoden auf französischer Grundlage*. Tübingen: Niemeyer.
- Wolfe, Sam (2021). *Syntactic Change in French*, Oxford: Oxford UP.
- Wolfe, Sam / Maiden, Martin (eds) (2020). *Variation and Change in Gallo-Romance Grammar*. Oxford: Oxford UP.
- Wüest, Jakob (1979). *La Dialectalisation de la Galloromania: problèmes phonologiques*. Bern: Francke.
- Zauner, Adolf (1902). *Die romanischen Namen der Körperteile. Eine onomasiologische Studie*. Erlangen: Junge.
- Ziltener, Werner (1972). *Repertorium der Gleichnisse und bildhaften Vergleiche der okzitanischen und der französischen Versliteratur des Mittelalters*. Bern: Francke.
- Zinelli, Fabio (2016). 'Espaces franco-italiens: les italianismes du français médiéval', in Glessgen, Martin / Trotter, David (eds). *La régionalité lexicale au Moyen Âge*. Strasbourg: ELiPhi, 207–267.
- Zink, Gaston (1986). *Phonétique historique du français*. Paris: PUF.

Indexes

Subject index

A

- abbreviations 502, 519
Ablaut. cf. vowel alternation
Abstandsprache. cf. language by distance
abstract 297, 322, 329, 368, ~ concept 310, 337, 448, ~ reality 320, ~ referent 318, 325
Abtönung. cf. modalisation
academia 526
academy (language ~) 203, 404, **467**–469
 Academia das Ciências de Lisboa 467
 Academia de la Llingua asturiana 84
 Academia română 467
 Academia Valenciana de la Llengua 85
 Académie française 203, 461, 467, 468
 Accademia della Crusca 203, 467
 Real Academia Española 203, 467, 522
accent 519, ~ position 241
 diacritic ~ 201, dynamic ~ 196
acceptability 11, **110**
acculturation **408**, 440
accusative 237, 239, ~ with infinitive 411
acoustic phonetics 178
acquisition (language ~) 44
acrolect **133**, 468
acronym 361
act, (performative) illocutionary ~ 7, **150**, 278, speech ~ 10, 41, **150**, 278, speech ~ person 293, 297
actant, actantial role (Tesnière) 169, 275, 292, second ~ 296
action mood 266, completion of ~ 275, mode of ~ (*Aktionsart*) 274sq., potential ~ 269
active articulation 181
activity (brain ~) 166
actual 322, the ~ 321, ~ referent 320
actualisation 149, **261**, 325
adaptation of Latin texts 442, graphic ~, morphological ~, phonetic ~ 374sq.
address (form of ~) 20, 246sq.
addressee **245**, 250
adjectival compound 357, ~ head 265, ~ inflection 234, **241**
adjective 225, 235, **241**, 244 phrase (AP) **265**, 279
 denominal ~, deverbal ~ 354, epicene ~ 235, qualifying ~ **261**–264, verbal ~ 284
adjunct (clause) 302, 304, circumstantial ~ 258sq., 276sq., 291, **299**, 302
administration 407, 438, centralised ~ 474
administrative document **441**sq., ~ infrastructure 410, ~ language 147
adnominal 239
adstratum 372, cultural ~ 408
adverb 225, 271, 279, ~ formation 243, ~ phrase 279, ~ position 279, ~ suffix 243, deadjectival ~ 354
adverbial clause 302, ~ expression 226
affix **226**sq., 230, 233, 349sq., 364, ~ function, ~ position 226
 derivational ~ **226**, 231, 310, 348sq., 378, grammatical ~ 168, **226**, inflectional ~ **226**, 231, 233, 243, 295, 310
affricate 183, **184**, 189–191, 215, 218, 220
 dental ~, palato-alveolar ~ 216
affrication 217
Africa 4, 68, 86, 402, 406sq., 455, 477, 529, equatorial ~ 475, northern ~ 411, 420, 434, sub-Saharan ~ 62, 125, 365, 452sq., 476sq.
age 129, 132, 316
Agens 292
agent 260, 272–278, 289, **292**–297, 300sq., 336, ~ subject 289
agglutination **230**, 233sq., 242, 260
agglutinative 233, 241, ~ language 232, ~ marker/ marking 230, 286
aggregation 141
agreement **228**, 266, 277, 295sq., ~ marker/ marking 245, noun-adjective ~ 241
agriculture 436, 438
Aktantenfunktion. cf. actant
Aktionsart. cf. action, mode of ~
Al-Andalus 420
Albanian 58sq., 60, 74, 415, 418
Albigensian Crusade 71, 435, 452
Alemannic dialect(s) 60, 69
Algeria 62, 529
Alghero 59, 64, 78, 84sq.
allegory 489, 491

- allograph 202
- allomorph **226sq.**
- allomorphy 226, **230**, 233, contextual ~, root-~, stem ~ **227**, 233
- allophone 107, **177**–193, 196, 209, 219
- allophonic distribution 205, ~ spirantisation 193, ~ variation 179, 192
- alphabet 203, 502. cf. also script, writing system
 - Arabic ~ 200, Cyrillic ~ 88, 200, 470, 502,
 - Greek ~ 25, 200, 502, Hebraic/Hebrew ~ 83, 200, Latin ~ 200, 470, phonetic ~ 182
- alphabetic writing system 25, 199–201, 400, 406
- Alps 417, Eastern ~ 213
- Alsace 58, 60, 69, 102, 119, 126, 415, 452
- Alsatianism 102
- alveolar articulation 181, ~ consonant 182, 189–191, 220, ~ ridge 181sq.
- Alvernian 119
- American, Native ~ languages 373, Meso-~ writing 200
- America(s) 4, 61, 81–83, 90, 101, 106sq., 110, 125, 128, 135, 365, 402, 410, 412, 434, 454–459, 475sq.
 - Hispanic ~ 269, 529, Latin ~ 90, 477, North ~ 477, South ~ 455
- analogy 25, 221, 233, 235, 240, **250**–255, false ~ 105
- analysis. cf. discourse
- analytical language 234, ~ principle 242
- anaphor 148, 249
- Ancien Régime* 373, 452, 482, 513
- Andalusia 407, 431, Andalusian 84, 108, **124**
- Angevin(s) 438
- Angles 415
- Anglicism 27, 102, 316, 369, **372**, 374, 377, 477, 479
- Anglo-American linguistics 129
- Anglo-Norman 69, 103, 433, 444, 504
- Angola 62, 86, 455, 475
- animacy hierarchy 237, **293**, 294, 297, 329, 337, 344
- animate 245, 265, 292sq., 297, 329
- antepenultimate syllable 199
- anteriority 267
- anterior. cf. past
- anthropogeography **431**, 434
- anthropological. cf. concept
- anthropology (historical ~) 344, 345
- anthroponym(y) 312, 379, **380–382**, 420, 427
- anticipation 194
- Antiguo Régimen* 453
- Antilles 61, 452, 477
- antiphrasis 332, **338**
- Antiquity 58, 172, 386, 405, 412, 445, 492
- antonomasia 313, **380**
- antonym(y) 311, 317, 328, 329, auto-~ 339
- Aosta Valley 60, 64, 73, 74
- Apennines 64, 417, 422
- aphasia 44, **166**, Broca's ~, Wernicke's ~ 166
- apheresis 194
- apical 186
- apico-alveolar 111
- apocope 194
- apostrophe 233, 519
- apparatus (critical ~) 512
- applied linguistics **41**, 533
- apposition 264, **304**
- approximation 352
- Apulia, Apulian 64, 73, 77, 206
- Arab 431, ~ immigration 476
- Arab-Berber armies/conquest 81, 415, **420**, 421
- Arabic 82, 102, 133, **372**, 415, 420sq., 440, 476, 492, ~ alphabet 200, ~ characters 502
- Arabism 103, 373, 377
- Arabo-Muslim expansion 416, 420
- Aragonese 72, 81, 83sq., **90**, 402, 432, 444, 453, 458, 475
 - ~ conquest 78, ~ kingdom 438, ~ rule 434
- Aranese **72**, 81, 91, 402, 475
- arbitrariness **8**, 230, 319
- archaic form 121, ~ spelling 222, ~ vowel system 206
- archaism 15, **316**, 317
- archetypal sound 194
- archetype 509
- archilexeme 328
- architecture 111, language ~ (Flydal) 128
- Argentina 61
- argot* 136
- argument 257–261, **273**–276, 291–302, ~ clause 302, 304, ~ constellation 272, ~ hierarchy 292, 295, ~ marker/marking 227, 276, **295**–299, ~ number 259, 272, 275, 366, ~ structure **260**, 276
 - second ~ 296, zero ~ 291
- Aromanian 59, 87, 92, **125**, 402, 418
- artefact 334
- Arthurian romance 440, 489
- article 98, 241, **261sq.**, 379, ~ omission 262
 - definite ~ 150, 169, 171, 241sq., **262sq.**, 379, 384, 414, indefinite ~, partitive ~ 262, 297, postposed ~, zero ~ 262

- article (publication) 50, newspaper ~ 140, 146, 158sq.
 articulation, active ~, alveolar ~, passive ~ 181
 manner of ~, place of ~, position of ~ 180, **181**, 183, 186–188, 193, 198
 articulatory apparatus 175, ~ organ 181, ~ phonetics **178**, 194, ~ variant 178
 artificial language 307
 Arverno-Limousin 444
 Asia 86, 434, 455, 529, ~ Minor 406, 433, East ~ 475
 Asian languages 373, ~ immigration 476
 aspect 43, 228, 250, 266–270, 275, 283, 285,
 imperfective ~ 253, **268**, perfective ~ 268,
 progressive ~ 271
 aspiration 218sq.
 assertion 258sq., **285**
 assertive content/sentence 149, 198, **258sq.**, 270, 285
 assimilation **214sq.**, 223, 411
 assimilation **193**, 217sq., long-distance ~,
 progressive ~, regressive ~ 193, reciprocal ~ 193, 432
 association 314, **316**, 327, 330, 334, 337, 339, 344,
 380, cognitive ~ 332, 344, emotional ~ 324,
 326, shared ~ 327, syntagmatic ~ 366
 Asturian 81, **90**, 402, 458, 475, Asturian-Leonese 83
 asymmetry 233, morphological ~ 230
 atlas (linguistic ~) 35, 98, 114–120, 125, 138, 182sq.
 attitude of discourse 129
 attribute 277, 304
 attributive 239, ~ adjective 249, ~ verb 277
 auditory phonetics **178**, 194
Ausbauprache. cf. language by elaboration
Ausdruck 149. cf. expressive function
 Australia 74, 529
 authenticity 202, 503, 505, **507**, 512
 auto-antonymy 332, **339**, ~converse 336,
 ~derivation 380
 autobiographical writing 459, 473, 493
 autograph 506
 autonomous status 73
 autonomy 91, internal (linguistic) ~ 17, 57, 89sq.
auxiliarii 410
 auxiliarisation 270
 auxiliary verb 250, 254sq., 266sq., 270–273,
 277–279, 298, causative ~ 274
 availability 278
 awareness of diglossia 423, 427, 431, linguistic ~ 429, 453, speaker ~ 56, 309, 312, 358, 363
 axiological inversion 338
 Aymara 81
 Azores 61
B
 back-formation 355, 359
 backness (vowels), back vowel **182**, 186sq.
 Balearic Islands 416, 444, ~ Catalan 84sq., 261, ~ variety 59
 Balkan(s) 83, 87, 237, 406sq., 415, 432
 ~ system (vowels) **206sq.**
banlieues 133, **136sq.**
 Barcelona 84sq., 444
 Bartsch's law 210
 base, derivational ~ 360, free ~ 349, lexical ~ 43, **310**, 350, 360, verbal ~ 362
 Latin ~, pre-Latin ~ 388
 basilect 133
 Basque 59sq., 69, 72, 81, 206, 371, 408, 415sq.
 Béarn 452, Béarnese 98
 Bédier's method 510
Begriffssystem (Hallig/Wartburg) **328**, 391
 Belgium 58, **68sq.**, 90, 415, 453, 530
Benefaktiv 292
 beneficiary 260, 275, **292**
 Beneventan script 500
 Berber 415, 420sq.
 Bible translation 460, 470, Gutenberg ~ **460**, 500,
 Lutheran ~ 460
 bilabial consonant **181**, 189–191, 220
 bilateral marking 287
 bilingual dictionary 464sq.
 bilingualism 69, 71, 73, 80–85, **101sq.**, 370, 406,
 408, 442sq., 475
 bivalent verb **275**, 291, 301
 blending **362**, 386
 Bolivia 61, 81
Bolognese 208
 book hand 500sq.
Bordelais 72
 border region 415
 borrowing 19, 20, 27, 83, 102–104, 127–131, 173,
 205, 221sq., 316, 330, 353, 360, 365, 369–379,
 384–388, 392, 420, 430, 433sq., 477, 479
 diasystematic ~ 342, direct ~ 221, false ~ 375,
 formal ~ 222, 369, **374**, intralinguistic ~ **378**,
 388, luxury ~, necessity ~ 377, variational ~ 130
 grammatical ~ 369, lexical ~ 103–105, 131,
 369, **374**, 376, 443, phonetic ~ 369, semantic

~ 222, **369**, 376, structural ~ **376**, syntactic ~ **377**, 443, syntagmatic ~ 377
 bound form/unit **225**, 227
 boundary 421, 431, chronological ~ 170, dialect ~ 118, 121, geographical ~ 416, (geo)linguistic ~ 62, 64, linguistic ~ 417, 434, phonetic ~ 193
 ~ of the Romània 416
Bourguignon. cf. Burgundian
 Boyar (class) 456
 brain 44, 166sq., 322, 330, 337, ~ anatomy, ~ injury 166
 Brazil, Brazilian 61, 86sq., 90, 455, 475, 529
 Breton 58–**60**, 69, 416
 Britain 406–409, 415
 British Isles 69
 Brittany 58, 70, 120, 452
 Broca's. cf. aphasia
 buccal cavity 180, 187
 Burgundian(s), Bourguignon 120, **419**, 444
 Burgundy 438, 444, 452
 byname 382
 Byzantine Church 418, ~ period 78

C

cacuminal consonant 182
 Calabria 64, 77, Calabrian 116, 206
 calamus 500
 calque 376
Campidanese 59, 77sq.
 Canada 69, 110, 452
 Canadian French 108, 452
 Canarian Spanish 108, Canary Islands 434
 capacity (language ~) 7
 Cape Verde Islands 61, 86, 106, 455
 capital
 ~ city 448
 ~ letter 201, square ~s 500
 Caribbean 83, 454
 Carolingian(s) 421sq., 498, ~ minuscule 500sq., ~ period 429, ~ Renaissance/reform 414, **427**–429, 443, 500
 Carthage 406sq.
 cartulary 503sq.
cas régime, ~ *sujet* 239
 case **234**, 237–241, 244, 411, ~ distinction/reduction of ~ **238**–241, 248, ~ marker/marking 239, ~ role 292, ~ system (Latin) 234
 absolute oblique ~ 265, accusative ~ 239, deep ~ 292, marked ~ 241, 295, marker/

marking ~ 241, oblique ~, subject ~ **237**–240, 286, unmarked ~ 241, 286
 Castilian/*castellano* **82**sq., 402, 432, 444, 449, 453, 458, 463, ~ Spanish 108
 Catalan 3, 59–72, **78**–89, 248, 373, 387, 421, 431, 435, 439, 471, 474, 481, 483, 528, Balearic ~ 261, Eastern ~, Western ~ 84sq., 444
 Catalanism 102
 Catalonia 102, 421, 471, 528
 categorisation 379
 category, folk ~, scientific ~ 334
 causal contiguity 335
 causative 293, ~ auxiliary 274, ~ construction 298, ~ diathesis **272**, 274, 291
 causativity 294
 cause – effect 336
 cavity, buccal ~ 180, 187, glottal ~ 182, nasal ~ 184
 cedilla 201
 Celtiberian 408
 Celtic 30, 60, 100, 370, 408sq., 415
 Celto-Iberic 370
 central group 120, ~ vowel **182**, 186sq.
 centralisation 38, 75, 399, 407, 449, 452, 482
 centralised state/administration 407, 438, 462, 474
 centrifugal force 40, 388, 413, 482, ~ paradigm 45, ~ phenomena 18, ~ syntax 288, ~ tendency 45, 474–477, 531
 centripetal syntax 288
 Chabacano 83, 106
 chain of variation **130**, 137, ~ of vertical communication **413**, 430
 phonetic ~ 175
 Champagne 120
 chancery 443, 502sq., ~ cursive 501sq.
 royal ~ 402, 442, **444**, 449, 458, 462
 change
 internal ~ 170, language ~ 14, **15**, 19sq., 23, 107, 154, 170, 172, 307, 402sq., 430, 464, 479sq., 506, grammatical ~ **305**, 330, morphosyntactic ~ 426
 lexical ~ **330**, 385, 426, onomasiological ~ 346, semantic ~ 9, 294, 327, **330**, 331–334, 339–341, 344, 347, 364, 369, 375, 388, taxonomic ~ 331, 332, 334
 phonetic ~ 15, 104, **178**, 177, 182, 193, 199–205, 223, 250, 330, 426, phonological ~ 178, **192**, sound ~ **32**, 178sq., 194
 ~ of word class 313, 348
 Channel Islands 70
chanson de geste **440**, 489

- chansonnier* 434, 439, **507**
 character 517, Arabic ~, Hebrew/Hebrew ~, printed ~ 502
 characterisation 261, 264
 charter, chartae 439, 442, **494**, 498sq., 503, 513, 517
 Chile 61
 Chinese 200
 chirograph 503
 chronicle 473, 492sq.
 chronological connotation 316
 chronology (relative ~) 179, 430, 524
 Church 403, **437**sq.
 circumfix 286
 circumstantial adjunct 258sq., 276sq., **291**, 299, 302, ~ complement 292
 city 74, 398, 438, 448, 450, **457**
 class, social ~ 457
 word ~. cf. part of speech
 classification (of Romance languages) 95, 100sq.
 clause 163, 168, 257, **259**, 288, 302–304
 adjunct ~, adverbial ~, argument ~, infinitive ~, main ~, object ~, participial ~ 302–304
 complement ~ 291, 302, 304, independent ~ 259, relative ~ **264**, 302, subordinate ~ 249, 269, **286**, 302–304
 cleft, pseudo~ 301
 clefting 301
 clerical text 491
 cliché. cf. phraseologism
 clipping 361
 clitic **227**, 243–246, 298, ~ climbing, ~ position 298, ~ pronoun 245, **255**, 295, 297
 object/subject ~ 244, **245**, 251, 295, 296, 298, phonological/syntactic ~ 295
 closed. cf. syllable
 Cluniac monks 429, ~ reform 442
 cluster
 concept ~ 103, 279
 consonant ~ 196, 210, 213sq., **217**, **218**
coda 195
 code-mixing, ~-switching 102sq.
 graphic ~, phonic ~ 137, **138**sq., 158, 483
 codex 439
 codicology 496
 codification (pluricentric ~) 18, 127, **128**, 475–477
 cognition 43, 325
 cognitive association 332, 344, ~ capacity 165, ~ function (of language) 5, **6**sq., 149, ~ linguistics 22, 42, 331, 345, ~ relationship 386, ~ semantics **328**sq., ~ structure 387, ~ theory 329, ~ turn 41, 526
cognomen 381
 coherence **148**, 285
 cohesion 110sq., 121, **148**, 285, 397, 412, 430, 436, 438
 diasystematic ~ 113sq., 124, geographical ~ 421, geolinguistic ~ 413, infrastructural ~ 422
 cohesiveness 160, semantic ~ **367**
 co-hyponym(y) 311, 317, 328, 332, **334**
 collective
 ~ memory 151, ~ unconscious 344
 ~ noun 351, ~ plural 238
 collocation 276, **311**, 321, 366–369, 521
 colloquial conversation 159, ~ variety 413, 472, 474, 524
 Colombia 61
 colonial empire/expansion 451sq., ~ language **406**, 412, 432, 452
 colonisation 42, 61, 105sq., 372, 402, 407, **431**, 433, 454sq., 475
 combinatory variant 177
 comment 148, 198, 300
 common noun **312**sq., 379sq., 427
 communication, interregional ~ 421, long-distance ~ 403, 422, 440sq., 451, 472, vertical ~ **413**, 430
 communication model/Organon model (Bühler) **6**
 communication system 8
 approximative ~ 42, minimal ~ 105sq.
 communication verticale 413
 communicative context 143, **315**sq., 397, 470, ~ distance **139**, 155, ~ function 145, 458, ~ marker/marking 154, ~ proximity 139, ~ situation 109, 140, 144, 151, ~ space **58**, 60, 92, 110, 204, 459, 474, 477
 community (speech ~) 11, 29, 151, 326
 commutation **13**, 176
 comparative approach 30sq., **44**, 163, 178, 284, 388
 comparative (degree of comparison) 242, synthetic ~ 242
 comparison (degree of ~) **242**, 280
 competence **10**–12, 25–29, 151, 157
 complement 265sq., 280, 288, 304, ~ clause 291, 302, 304
 circumstantial ~, indirect ~ 292, noun ~ 264, 302, object ~ 262, 276, 277, 292
 complementiser/~ deletion 303sq.
 completion (of action) 275

- complex sentence **259**, 285, 307, ~ syntax 107, ~ word 311, 363
- complexity (syntactic ~) 21, 460, 462, 492, 513sq.
- composition 375
- compositional morphology 164
- compound 224, 357, 366, ~ meaning 358sq., ~ past **268**, 270, 282sq., ~ past perfect 267, ~ sentence 259, ~ tense 283
- adjectival ~ 357, copulative ~, cumulative ~ 359, derivation within ~(s) 361, determinative ~, endocentric ~, exocentric ~, improper ~, lexicalised ~ 358sq., syntagmatic ~ 357, **356**, 366, 368, verbal ~ 357
- compounding 8, 326, 347sq., **356**
- comprehensibility 443, 507sq., mutual ~ 443
- concept 7, 9, 21, 165–167, 225, 311, **314**, **317**–337, 342–348, 357, 377, 385, 388, 443
- ~ cluster 103, 279, ~ emergence 325, ~ extension 379, ~ meaning 329, ~ verbalisation 324, 325, **326**sq.
- abstract ~ 310, 337, 448, anthropological ~ 285, concrete ~ 310, 337, non-activated ~, non-verbalised ~ 324, source ~, target ~ 337
- conception (of a message) **139**
- oral/phonetic ~, written ~ **139**sq., 143, 425
- conceptual feature 332sq., ~ memory 324, ~ network **323**sq., 330, 344
- conceptualisation 325
- concord 228
- concordance 520sq.
- concrete 322, 329, 368, ~ concept 310, 337, ~ reality 320, ~ referent 318, **325**sq.
- conditional **255**, 268sq., 281sq., 303, 411, ~ perfect 283
- conditioned diphthongisation 208, **210**
- condition (prosodic ~) 286
- configuration 153, 155, 170, **287**
- configurationality 286
- conflict, language ~ 69, 113, 482
- modal ~ 303, semantic ~ 302
- congruence 287
- conjugation 228, ~ class 229
- object ~ 283, **297**, subject ~ 283, **295**
- conjunction, coordinating ~, subordinating ~ 225, 302, 304
- connotation **315**sq., 342, 377
- chronological ~ 15, **316**, diaphasic ~ 130, 315sq., diastatic ~, diatopic/regional ~ 315sq., lexical ~ 131, popular ~, pragmatic ~ 315sq., semantic ~ 169, 262, 271, 290, 312, 315sq., 340, variational ~ 311, **316**, 321, 389
- connotative value 131
- conquest 406, **433**sq., Arab-Berber ~ 81, **420**, 421, Aragonese ~ 78, extra-European ~ 453, Muslim ~ 82, 420, 431
- Napoleonic ~ 455, Norman ~ 431–433
- consonant 66, 180, **181**–185, 214, 219, ~ system 189–191, 219
- affricate ~ 218, alveolar ~ 182, bilabial ~ **181**, 189–191, 220, cacuminal ~ 182, dental ~ **181**, 189–191, 220, double ~ 66, 97, glottal ~ **182**, 220, interdental ~ 181, labial ~ 216, labiodental ~ **181**, 189–191, 220, labiovelar ~ 220, nasal ~ 66, **184**, 188–192, 220, palatal ~ **182**, 189–191, 210, 220, palato-alveolar ~, prepalatal ~, retroflex ~ **182**, transitional ~ 218, uvular ~, velar ~ **182**, 189, 190sq., 220, voiced ~, voiceless ~ 180, **185**, word-final ~ **217**, 299, 411
- consonant cluster 196, 210, 213, ~ simplification 214, **218**
- secondary ~ **217**sq.
- constituency grammar 257
- constituent 168, 239, 257, **258**, 288, 521, ~ marker/marking 239
- free ~ 358
- constituent order 43, 103, **288**–290, 298–304, 308
- prototypical ~ 300
- SOV (subject-object-verb), SVO, VOS, VO, VSO 103, 288–290, 296, 300
- construction grammar 22, **165**, 257, ~ head 288
- presentative ~ 142
- contact language 67, 71, 81, 173, **370**–372, 385, 409, 418, 421, 432, ~ zone 102
- cultural ~ 82, 102–104, 369, 371, 434, language ~ 20, 42sq., 81, **101**–104, 112, 369, **370**–373, 401sq., 408–412, 419, 421, 432–438, 447, 476
- container-content 332
- contamination 509
- content 275, 331, ~ (Hjelmslev) 319
- assertive ~, denotative ~, semantic ~ **258**, **285**
- context 159, 300, 321
- communicative ~ **315**sq., 397, 470, cultural ~ 344, diasystematic ~ 177, 311, **321**, extralinguistic ~ 149, 261, historical ~ 512, Latin ~ 423, situational ~ 110, 157, socio-historical ~ 344, sociological ~ 470, syntactic ~ 224, syntagmatic ~ **311**, 321, 324, 366, usage ~

349, 357, use ~ 346, usual ~ 368, utterance ~ 267, variational ~ 397

contextual allomorphy 231, ~ ellipsis 273, ~ meaning 318, ~ morphosyntactic feature 228

contextualisation 131, 135, 224

contiguity 8, 329–**334**, causal ~, formal ~, spatial ~, temporal ~ 335, 339

continuity 112, 171, 174, 220, 250, 254, 280, 284, 421, 435, 476, ~ of form/function 280, dialectal ~ 422

continuum 474

 diastratic-diaphasic ~ 137

 immediacy-distance ~ **140**–146, 151–158

 linguistic/dialect ~ 57, 63, 88sq., 95, 99, 111, **121**, 447, 472

contraction 247, phonetic ~ 282

contrast 15, 177, 332sq., 338

control 294

controller 228

conventionality 8–10, 17, **319**

conversation ~ analysis 147, 150

 colloquial ~. familiar ~, informal ~, telephone ~ 140, 146, 158sq.

conversational maxim 150

conversion 313, **354**–356, 380

cooperative principle (Grice) 41, **150**

coordinating conjunction 302

coordination 225, pseudo-~ 271

Copenhagen linguistic circle 39

co-presence 334, **335**

copular verb 277

copulative compound 359

copy 205, 490, 502, **503**–506, 513, first-hand ~, multiple ~ 495

copyist 444, 504, 508

coreference 270

corpus **317**, 515, 524, ~ linguistics **515**sq., 521, 523

 oral ~ 138

Corsica 59, 77, 406, 416, 452

Corsican 57, 64, 68, 75, 78, 91, 116, 475

Cortina d'Ampezzo 79

Costa Rica 61

Council of Tours **429**, 490, ~ of Trent 382, **491**, Second Vatican ~ 477

countable 329, ~ noun/count noun 234

counterfactual 282

court 438, 453, 456, **457**, ducal/princely ~ **438**, 457sq., 462, royal ~ 75, 127, 438, 444, 448–459, 461–463, **466**, 468, 482

creativity **7**, 315, 481, 523

creole 36, 42sq., 105, **106**, 239, 289, 299, 409, 471, 476

 ~ studies **42**, 526

 Arabic-based ~, English-based ~, French-based ~, Hispano-Portuguese-based ~, Luso-Spanish-based ~, Romance-based ~ 105sq., Portuguese-based ~ 83, 86, **106**, 476, Spanish-based ~ 83, **106**

 Haitian ~ 300

creolisation 409, 453

critical edition 508

criticism, source ~ 521, 525, textual ~ 512

Croissant 121

crusade **433**, 435, Albigensian ~ 71, 435, 452

Cuba 61, 412, 454

cultural adstratum 408, ~ centre 412, **458**, ~ contact 82, 102–104, 369, 371, 434, ~ context 344, ~ elaboration 458, ~ factor 459, ~ knowledge 334, ~ model 369

culture, language ~ **404**, 479, linguistic ~ 307, **445**, religious ~ 403, written ~ **401**, 404, 413, 423, 433, 439sq., 459, 478, 479

cumulation (of functions)/cumulativeness 230, 232sq.

cumulative compound 359, ~ encoding 241, ~ morpheme 232

cuneiforms 200, 496

cursiva 200

cursive 500–**502**, chancery ~ 501sq., Merovingian ~ 500

customary law 493, 495

cycle, future ~, linguistic ~ 281

Cyrillic alphabet/script 88, 200, 418, 470, 502

D

Dachsprache. cf. umbrella language

Dacia 406sq.

Daco-Romance **100**, 239, 431

Daco-Romanian 59, 88, **125**

Dalmatia(n) 59, **61**, 68, 89, 96, 100, 208, 243, 433sq., 474

Danube, Danubian (south ~) 418

Darstellung 6, 149

data collection 524, management 498, 523

 empirical ~ 310, oral ~ 138, textual ~ 524

database 92, 129, 368, 389–393, 515, 521sq.

dative, Dativ 239, 292

Dauphinois 70, 119, 444

deadjectival adverb 354, ~ noun 353, ~ verb 356

deanthroponymic proper noun 380

declarative 259, ~ memory 167, 322

- declension 94, **228**, 231–241
 declension class, reduction of ~ 229, 411
 decoding 286, 290, 364
 decolonisation 476
 decompositionality (non~) 358
 de-dialectalisation 70, 449sq.
 deduction 135
 deep case 292
 defining relative clause 264
 definite article 150, 169, 171, 241sq., **262**sq., 379, 384, 414 ~ determiner 261
 degemination 97sq., 214, **215**
 degree of comparison **242**, 280
 deixis 142, 149, 225, 248, 262, temporal ~ 271
 de-Latinisation 415
 delexical proper noun **380**, 384
 delimitation of varieties 111, 113, 171
 geographical ~ 68, 421, semantic ~ 386
 delocutive 361, discursive ~, pronominal ~, suffixal ~, verbal ~ 362
 democratisation 477
 demographic decline 416, ~ expansion 451
 demographics **401**, 403, 451
 demonstrative 248, ~ determiner 247, ~ pronoun 225, 244, **247**, 262
 denominal adjective 354, ~ toponym 379, **384**, ~ verb 354, 356
 denotation 131, 315–317, denotatum 321
 denotative content 258, 285
 density (population ~) 399, **403**, 416, 422, 454
 dental affricate 216, ~ consonant **181**, 189–191, 220, ~ fricative 215sq.
 deonomastics, deonymisation 313, **380**
départements d'outre mer (DOM) 62, 68, 476
 dependency grammar 38, **257**, 288, head ~ 257
 deponent verb 228, **272**, 284
 depopulation 435
 derivation 8, 224, 313, 326, 330, **347**–350, 359sq., 365, 375, 379sq., ~ within compounds 361
 auto~ 380, improper ~ 355, regressive ~ 359, zero ~ 355
 derivational affix **226**, 231, 310, 348sq., 378, ~ base 360, ~ innovation 387, ~ morphology **164**, 243, 385, ~ suffix 356
 descriptive grammar 307, ~ norm 134
 desemanticisation **169**, 200, 271, 312
 designated, designation 320sq.
 determination **261**, 379
 grammatical ~, taxematic ~ 286
 nominal ~, partitive ~ 261
 determinative compound 358
 determined 359, 376
 determiner 225, 241–249, 261, 266, 359, 376, ~ phrase (DP) 288
 definite ~ 261, demonstrative ~ 247, indefinite ~ 261, nominal ~ 263, 265, optional ~ 279, possessive ~ 247, unstressed ~ 244
 detoponymic proper noun 380
 development, language by ~ 88, 92
 deverbal adjective 354, ~ noun 352
 devoicing 214, 217
 diachronic 23, ~ lexicography 34, ~ linguistics **19**, 23, ~ phonetics 200, ~ variation **23**, 107
 diachrony **19**, 45
 diacritics 183, 201, accent, cedilla, tilde 201, subscript ~ 201
 diaeresis 185
 dialect 143
 dialectal continuity 422, ~ differentiation 426, ~ literature 478, ~ variation 419, 432, ~ variety 62, **447**
 dialectology 32, **33**, 35, 45, 113sq., 138, 156, 308, 480, 525, 532
 dialectometry **118**, 525
 dialect(s) 28, 31, 45, **55**, 58, 62, 91, 130, 159, 171, 307, 398, 435, 458, 479, ~ boundary/frontier 118, 121, ~ continuum 57, 63, ~ d'oil 92, ~ elimination 479, ~ zone 437
 elaborated ~ 91, extinction ~ 474, hereditary ~ 108, modern ~ 430, oil ~ **55**, 92, 119, primary ~ 83, 92, **108**, **112**, 113, 117, 124, 126, 129, 474, 483, regional ~ **55**–58, 62, 67, 82, 88, 90sq., 108, 112, 130, 145, secondary ~ 83, 108, 112, **124**, 125, 432, 452–455, tertiary ~ 126
 diamesy 140, **142**, 157
 diaphasic 129, ~ connotation 130, 315sq., ~ marker/marking 152, 159, 514, ~ variation/variety **109**, 131–136, 143–147, 157, 472, 478, 482, 513
 diaphasy 143, 156
 diastratic 129, ~ connotation 315sq., ~diaphasic continuum 137, ~ dimension 472, ~ marker/marking 152, 514, ~ variation/variety 36, 40, **108**, 109, **131**–137, 143, 157, 171, 411, 448, 472, 478, 483,
 diasystem 12, 111, 124, **129**, 136, 144–146, 151–160, 173, 400sq., 405, 469, 482, 515
 diasystematic 159, ~ borrowing 342, ~ choice 151, ~ cohesion 113sq., 124, ~ context 177, 311, **321**, ~ marker/marking 154, 157, **311**, 317,

- 324, ~ restriction 311, ~ variation/variety 126, **142**–146, 179, 482
- diathesis **228**, 250, 266, **267**, 272, 294, 308, causative ~ 272, **274**, 291, recessive ~ **272**, 273, 295, 301
- diathetic alternation 273, ~ passive 273, 301
- diatopic 129, ~ connotation 315sq., ~ marker/ marking 152, 157, 159, ~ neutrality 426, ~ variation **108**, 112, 135, 157, 411sq., 432, 443, 454sq., 472, 483, 513, ~ variety 143
- dichotomies de Saussure*, dichotomies of the sign **8**sq., **319**sq., *langue/parole* 11, 38, signifier/ signified 38, synchrony/diachrony **19**, 38sq.
- dictionary 114, 154, 203, 310, 386–394, 445, 461, 467, 481, 514, bilingual ~ 464sq., etymological ~ 31, 35, 121, 124, 306, 386, **389**–394, 465, 514, historical ~ 31, 276, 390–393, monolingual ~ 465, pan-Romance ~ 35, semasiological ~ 465, toponym ~ 384, valency ~ 276
- differential analysis 285, 398, ~ description 135, ~ method **307**, ~ object marking (DOM) 297, 299
- differentiation 3, 18, 62sq., 72, 95, 113, **194**, 198, 414, 416, 436, ~ of Latin 419
- dialectal ~ 426, group ~ 7, internal ~ 472, lexical ~ 412, linguistic ~ 417, 419, regional ~ 288, 409, **412**, 420–422, 430
- digital humanities 113, 525, ~ philology 515
- diglossia* **101**–103, 130, 402, 408, 410, 414, 427, 429, 458, 472, awareness of ~ 423, 431
- digram 201
- diminutive 351
- dimotiki* 415
- diocese boundary 38, 403, 437
- diphthong **185**, 187, 208–211, ~ simplification 254
- falling ~, nasal ~, rising ~ 187
- diphthongisation **208**sq., 223, conditioned ~ 208, **210**, spontaneous ‘French’ ~, spontaneous ‘Romance’ ~ 208
- diplomatic edition 426, **508**–512, 518
- direct object 260, **277**, 296, 298
- discontinuity (linguistic ~) 458
- discontinuous negation 279
- discourse **12**, 153, ~ analysis 147, 150, ~ marker/ marking, ~ particle, ~ planning, ~ structuring 140sq., ~ tradition 109, 137, **144**–147, 151–157, 285, 309, 403, 478, 488, 503sq., 513
- attitude of ~ 129
- discursive delocutive 362
- dislocation 290, 296sq., **301**
- dissimilation 194
- distal 247
- distance 248, ~ expression 472, ~ language 474
- communicative/linguistic ~ 137, 139–143, 155, 158, 160, 426, 441, 443, 468, 472, geographical ~ **121**, 475, 482, language by ~ **56**sq., 88, 139, linguistic ~ 56, 133, social/personal ~ 246sq.
- distant etymology 377
- Distanzsprache*. cf. language by distance
- distinctive feature 176, **188**, 196, 205, 220, **294**, 325–329, 337sq., 343, 380, ~ unit 176
- distortion (formal ~) 362
- distribution 287, allophonic ~ 205, complementary ~ 177
- distributionalism 39
- ditransitive verb 277
- dittography 504
- diversity, lexical ~ 142, 493, linguistic ~ 92, 451, syntactic ~ 493
- division (dialectal ~) 121
- doctrine (legal ~) 494
- documentary text/writing 33, 72, 78, 439, 440, **488**, 494, 495, 500, 503, 513, 522
- Dolce Stil Novo* 489, 513
- Dolomitic Ladin 60, 74, **80**, 89, 92, 387
- DOM. cf. differential object marking
- domain 332, **334**, ~ mapping 332, 337
- language ~ 44, 107, **163**, 165, 168, 169–171, royal ~ 452
- Dominican Republic 61, 454
- donor language 102
- dorsal 186, ~ stream 167
- Douai 439
- double compound past **268**, 282sq., ~ consonant 66, 97, ~ nature of language (Humboldt) 29, ~ negation 279, ~ processing 364, ~ subject 296
- doublet (etymological ~) 222
- dropping (schwa~) 193
- D-system 229, **243**
- ductus* 504
- Dutch 83, 106, 373
- dynamic accent 196, ~ verb 278
- dysphasia 44
- dysphemism 131, 136, **340**
- E**
- ecdotics 487
- economy (principle of ~) 12–16
- Ecuador 61, 81

- edict of Moulins 466
 edition 487, 508–514, 524, critical ~ 508, diplomatic ~ 426, **508**, 510, 512, interpretative ~ 510, 511 (semi-)diplomatic ~ 518, synoptic ~ 510
 education 404, 478–480
 effect (expressive ~) 130sq.
 Egyptian 200
 elaborated dialect 91, ~ language 474
 elaboration, abandonment of ~ 470, cultural ~ 458, full~ 89, graphemic ~ 430, hypotactic ~ 494, language ~ 56, 67, 75, 81, 85, **112**, 126, 143, 160, 304, 307, 402, 404, 431, 433, 439, 446, 449sq., 455, 458–460, 466, 470–472, 479, 482, 492, language by ~ 56, **90**, partial/ongoing ~ **89**, 470, scriptural ~ 170
 elimination (dialect ~) 479
 elite 436, 440, 443, 459, **468**, 478, ~ variety 307
 ellipsis 243, 267, 331sq., 339sq., contextual ~ 273
 El Salvador 61
 email (personal ~) 140, 158
 emancipation of women 477
 Emilian 208, 425
 emotion 198
 emotional association 324, 326, ~ charge 377, ~ connotation 131
 emotive marker/marking 141
 emphasis 149, 239, 244, 264, 289, 300sq.
 empirical approach 111, ~ data 310, ~ knowledge 516, 534, ~ study 114
 empiricism 525
 enclave 104
 enclisis 262, 298
 encoding, cumulative ~ 241, grammatical ~ 286, linguistic ~ **12**, 364, 496, material ~ **496**, 498, multiple ~ 14
 encyclopedia 46sq., 306, 492, **493**, 530sq.
 encyclopedic current 465, ~ element 323sq., ~ knowledge 151, 321, 325sq.
 endangered language 18, 73, 80, 83, 86sq., 92, 402
 endocentric compound 358
enérgeia (Humboldt) 11, 29
 Engadin 297, Engadinian 79
 England 60, 415, 431sq., 439sq., 493
 English 28, 81, 103sq., **372**, 433, 476–479
 Enlightenment 457sq., 493
énoncé. cf. utterance
 environment (positional ~) 179
 epenthesis **194**, 213, 218, 222
 epicene adjective 235
 epic poetry 440, 489
 epidemic 435sq., 457
 epigraphy 426, 499
 episodic memory 323
 epistemic form/verb **278**, 302sq.
 epistemology 445, 525
 epistolary genre 495
 equivalence (functional ~) 168, 286
érgon (Humboldt) 11, 29
 error. cf. mistake
 ethnic group 412
etimologia prossima, ~ *remota* 377
 Etruscan(s) 200, 371, 400, 406
 etymological doublet 222, ~ history 36, ~ motivation 120. cf. dictionary
 etymology 26, 36, 97, **385**–387, 465, 524sq., distant ~ 377, folk/popular ~ 25, immediate ~ 377, popular/folk ~ 331sq., 339sq., 362, 386
 euphemism **340**, 362
 Europe 58, 454, 457, 474, 477, western ~ 170, 436
 event 270, ~ memory 324, 326
 natural ~ 278, verbal ~ 266, 274sq.
 evidence, metalinguistic ~ 25, 205, 401, 423, **429**
 Évolène 73
 evolution, external/internal ~ 398, hereditary ~ 173, language ~ 16. cf. language change
 example (best ~) **333**
 exclamatory 259
 existential construction 261, ~ verb **278**, 301
 exocentric compound 359
 expansion 458, Arabo-Muslim ~ 416, 420, colonial ~, demographic ~, extra-European ~ 451sq., geopolitical ~ 402, variety of ~ 432
 experienter 293
 explicit memory 167
 expression 331, adverbial ~ 226, fixed ~ 155, 321, prepositional ~ 226, subject ~ 308
 expression (Hjelmslev) 319
 expressive effect 130sq., ~ function (of language) 6, **7**, 149, ~ need 377
 expressiveness 15, 142, 411, weakening of ~ 341
 expressivity 169, 247sq., 264, 271, **327**, 330, 340, 345, 360, 362, 422
 external aspects 23, ~ characteristics 19, **67**, 524, ~ criteria 56, ~ evolution 398, ~ factors 19, 220, 430, **532**, ~ history 35, **37**, 49, **397**–400, 405, 482, ~ linguistics 401, ~ motivation 369, ~ period 170
 extinction (language ~, language death) **18**, **402**, 474

extra-European conquest/expansion expansion
452sq., ~ *România* **36**, 61, 127, 475
extralinguistic 165, 320, ~ context 149, 261, ~ factor
398, 405, **482**, ~ reality 17, 318sq., ~ referent
312
extreme vowels 191, 208, **210**

F

fabliau 504
Fachprosa 492
Fachsprachenforschung 147
falling (diphthong) 187
false analogy 105, ~ borrowing 375, ~ documents
505
familiar conversation 140, 158, ~ Spanish 475, ~
variety **135**, 307, 317
family, language ~ 524, semantic ~ 331sq., word ~
339, 348, **385**
famine 436
farce 497
Fascism 533
fauna 334
Félibrige 71, 481
feminine 235–238
field 129, lexical ~ 328, semantic ~ 321, **328**, 332,
347, 376
figurative meaning 358, 367sq.
first object 239, **260**, 277, 291sq., 296sq., 299, ~
position 289, 300
fixed expression 155, 321. cf. phraseologism
fixedness 358, 366sq.
Flanders 452
Flemish 69, 415
flora 334
Florence, Florentine 64, 74sq., 455, 463
focalisation **148**, 244, 289
focus 296, ~ marker/marking 239
Foix 452
folio 499
folk category 334, ~ etymology 25, 331, **339**sq.,
362, ~ linguistics 27, **480**, ~ literature 472
folklore 28, 532
form 164, **310**sq., 331, bound ~ 225, **227**,
disappearance of ~ 124, epistemic ~ 303,
graphic ~ 202, 310, **321**, lexical ~ 164, 167, 311,
317sq., 320–327, lexicalised ~ 243, mixed ~
427, orthographic ~ 322, phonic/phonetic ~
310, **321**, 326, phonological ~ **9**, 166, 322, 324,
326, simple ~ 250

formal borrowing 222, 369, **374**, ~ contiguity
339, ~ distortion 362, ~ integration 374, ~
motivation 332sq., **339**, ~ proximity **339**, 376,
~ unity 226, 357
formula (stereotypical ~) 144, 151, 440
foundation (pragmatic ~) 149
fragmentation, geolinguistic ~ 416, territorial ~ 413
frame 332–334, **335**, 337, semantic ~ 321
Franc-Comtois 70, 120
France 68, **448**–451, 479, 529sq., northern/
southern ~ 438
Franche-Comté 415, 452
Francien 449
Franco- ~Italian 69, **434**, ~Piedmontese 490
Franconian 60, 69, 104
Francophonie 476
Francoprovençal 33, 55, 58–**73**, 74, 89–92, 118–124,
209, 213, 239, 394, 402, 419, 422, 444, 458, 471,
474, 483, 492
Franks 410, 419, 433
free base 349, ~ constituent 358, ~ lexeme 360,
~ syllable 195, ~ syntagm 366, ~ unit 227, ~
variant 178sq., 207, 276, ~ variation **177**
French 51, 56–**68**, 69, 73, 89–92, 110, 118sq., 126,
136, 171sq., 189, 209, 236, 248, 252, 254, 262,
268, 296, 378, 389, 420–424, 430–433, 439,
442, 449, 452sq., 456, 459–462, 465–470, 476,
479, 482
~ consonants 219, ~ consonant system 189,
~ dialect 402, 474, ~ Guiana 62, ~ Revolution
437, 442, 452sq., ~ vowel system 189, 208
Canadian ~ 108, 110, 452, informal ~ 296,
medieval ~ 126, 172, 217, Old ~ 265,
popular ~, standard ~ / *français courant*, ~ *des
banlieues/des jeunes*, ~ *familier*, ~ *populaire*,
soigné/littéraire, *standard/commun* 55, 127, 131,
136, 290, 307, 365
frequency 312, **314**, 322, 346, 364, 367, 388,
fundamental ~ 196
fricative 66, 183, 184, 188–194, 198, 218, 220, dental
~, interdental ~, palato-alveolar ~ 215215sq.
Friuli 59sq., 79, 81
Friulian 59, 60, 64, 67, 74, **79**–81, 89–92, 116, 208,
243, 394, 402, 415, 460, 471, 474, 483
fronting (stylistic ~) 271
front vowel **182**, 186
full elaboration 89, ~ pronoun **244**, ~ subject 295,
~ verb 298
function word **8**, **163**, 168, **225**, 227, 232, 286, 314

communicative ~ 145, 458, continuity of ~ 280, modal ~ 282
 functional equivalence 168, 286, ~ languages (*langues fonctionnelles*) (Coseriu) 128, ~ opposition 176, ~ syntax 257
 functionalism 22, 288
 functionality 323
 function (of language) 5, 16, 19, 21, cognitive ~, expressive ~, social/conative ~ 5–7, 149
 fundamental relationship 258
 fusion 230, **231**
 future **254**, 255, 281, 303, ~ cycle 281, ~ I. cf. simple future, ~ in the past 268, **282**, 283, 411, ~ participle 270, 284, ~ perfect 282
 infinitive ~ 284, periphrastic ~ **255**, 267, 271, 275, 412, simple ~ 254, 267, synthetic ~ 280, 411

G

Galicia(n) 59, 60, 67, 81–84, **85**–89, 125, 373, 387, 431, 435sq., 471, 474, 481, 483, 528, 439, standard ~ 89
 Galician-Portuguese/Gallego-Portuguese 3, 83, **85**, 425, 439, 444, 489
 Gallia cisalpina 406
 Gallicisation 104
 Gallicism 373
 Gallo 70
 Gallo-Italian 77, 417, 420
 Gallo-Romance **62**, 77, **100**, 118, 120, 295, 409, 417, 431sq., 438sq.
Gallurese **78**
 Gascon 3, 59–**72**, 81, 90, 92, 118sq., 215, 217, 430, 444, 471, 483
 Gascony 433
 Gaul 406, 408, 411, 413, 420, northern ~ 407, 419
 gender 228, **234**, **237**sq., 241, 244, 311, ~ change 237, 355, 356, ~ marker/marking 375
 genealogical approach 529
 generalisation 179, 312, 332, **333**
 Generalized Phrase Structure Grammar (GPSG) 42
 generativism, generative grammar 11, 22, 40–42, 165, **224**, 257, 266, 288, 305sq.
 genetic approach 30, 32
 genitive 239
 Genoa, Genoese 64, 74, 458
 genre 159, literary ~ 488sq., textual ~ 12, 132, 143–145, **146**–156, 285, 304, 309, 347, 397, 401, 433–447, 458sq., 472, 478, 487sq., 493, 498, 503–505, 513–516, 523

gentilicium (nomen ~) 381
genus verbi 272
 geographical ~ space **121**, 157, 401sq.
 ~ variation. cf. diatopic variation, variety
 geography 412, 430
 geolinguistics 32, **401**, 403, historical ~ 386
 geolinguistic configuration 121, ~ order 435
 geopolitics 456
 German 59, 60, 69, Swiss ~ 79
Germania 407, 409, 415
 Germanic 30, 415, 420, ~ invasion 371, 415, **419**, ~ language/dialect 239, **371**, 373, 410, 415, 419sq., ~ names 381, 420, ~ Studies 31, ~ superstratum **38**, 533, ~ toponyms 383
 Germanism **371**, 392, 420
 gerund 284
 gerundive 270, 284
Gestalt psychology 293
 Gibraltar 104
 glide **182**–192, 220, palatal ~, velar ~ 182
 glossary 425, **445**, 491
 glossematics 39
 glosses 425, 445, 491, interlinear ~, marginal ~ 446
 glossography 26, 310, 445–447
 glottal cavity 182, ~ consonant **182**, 220
 glottophagy 18
 goal **292**, 293, pragmatic ~ 153
 Gothic 30, ~ script 501
 government (= rection) 228
 Government and Binding (GB) 42
 graffiti 424
 grammar 41, 154, 164–166, 306sq., 349, 462, 467, 471
 constituency ~ 257, construction ~ **22**, **165**, 257, dependency ~ 38, **257**, 288, descriptive ~ 307, dialectological ~ 308, generative ~ 11, **224**, 257, 266, historical ~ 305–308, 480, normative ~ 164, **307**, Port-Royal ~ 27, 305, 462, prescriptive ~ 307, 308
 grammatical affix 168, **226**, ~ analysis 461sq., ~ change **305**, 330, ~ description 34, 36, **305**, 307, 404, 445, 480, ~ determination, ~ encoding 286, ~ innovation 173, ~ marker/marking 233, 260, 286, 299, 302, 308, ~ norm 444, ~ semantics **164**, 278, ~ sign 8
 grammaticalisation 149, **169**, 243sq., 262, 267–271, 275, 298, 308sq., 330, 340, ~ of definite article 414
 re~, secondary ~ 196, 296.
 grammaticality **109**, 154, un~ 110, 141, 152

grammaticalness (Chomsky) 109
 grammaticography 307
 grammeme **225**–234, 295
 Granada 431
 grapheme **199**–204, 426, ~ -phoneme relationship 201–204
 graphemic elaboration 430, ~ neutralisation 504
 graphemics 199
 graphic adaptation 374, ~ code 137–139, 158, 483, ~ form 202, 310, **321**, ~ sign 201, ~ utterance 140, ~ variation 202, ~ word 357
graphie classique (Occitan) 71, ~ *mistralienne* (Occitan). cf. *Félibrige*
 Greece 87, 406, 408, 418
 Greek 30, 59sq., 77, 239, 303, 371, 400, 406, 408, 415–418, ~ alphabet/script 25, 200, 502
 Grisons 60, 69, 78–80, 415
 group differentiation 7, ~ formation 13
 ethnic ~ 412, semantic ~ 278, social ~ 7, 55, 108, 132, 135, socio-cultural ~ 6, 18, 108
 Guadeloupe 62, 106
 Guaraní 81, 454
 Guardia Piedmontese 64, 71
 Guatemala 61, 81
 Guillaumian approach 306
 Guinea-Bissau 61, 86, 106, 455
 Guyana 106, 452

H

hagiography 490
 Haiti 68, 106, Haitian creole 300
 hand (book ~) 500sq.
 handwriting 497, 502
 haplography 504
 hard palate 180, 181, 182
 head **261**, 266, 288, ~-dependency 257
 adjectival ~ 265, construction ~ 288, phrase ~ 265, sentence ~ 288
 Hebraic/Hebrew alphabet/script 83, 200, 502
 Hebrew 83
 height (vowel ~) **186**sq., 205
 hereditary dialect 108, ~ phonetic evolution 104, 173, ~ word/form 221sq., 378, 387
 heritage (textual ~) 506
 hermeneutic approach 397
 hesitation 11, **141**, 149
 hiatus **185**, 213
 hierarchical structure/order 257, 286sq., 492
 hieroglyphs 200, 496
 high style 132, ~ variety (HV) 102, 132

hijra 416
 Hispanic America(s) 269, 529, pre-~ 410
hispanidad 475
 Hispanism 103, 373
 historical ~ approach 4, ~-comparative paradigm **28**sq., 32–34, 42, 45, 529, ~ dictionary 31, ~ language (Coseriu) 55, 124, **129**, 316, 400, ~ level (Coseriu) 153, ~ linguistics 23, ~ periodisation 173, 405
 historicity 5
 historiography 493, linguistic ~ 34, 404, 480, 483, 515, 532
 history 480, ~ of language 461
 etymological ~ 36, external ~ 35, **37**, 49, **397**–400, 405, 482, internal ~ 37, 163, 405
 h (loss of ~) 218
 holophrase 168
 homogenisation 124, 438, 444, 453, 459, 460, 462, 471, 477, 504
 homonymy 327
 homophone, homophony **202**sq., 251, 255, 345
 Honduras 61
 HTML 517
 human 293, 329, 338, non-~ 297, 329, 337
 Hundred Years' War 433, 435, 452
 Hungarian 59sq.
 hypercorrection 204
 hypernym, hypernymy 317, **328**–334
 hypocoristics 360
 hyponym, hyponymy 311, 317, **328**–334
 hypotaxis 140, **302**
 hypotactic elaboration 494
 hypothetical sentence 282

I

Iberia 420
 Iberian Peninsula 81, 82, 406, 408, 420sq., 429–432, 438
 Iberic (Celto-~) 370
 Ibero-Romance **100**, 417, 431sq.
 iconicity, iconic sign **8**, 17
 idealism 35, **37**, 399
 identification 249, 263
 identifier 249
 identity 7, 102, 109, 203, 449, 481sq., 532sq., ~ conflict 85sq., 101, ~ marker/markings 126
 partial ~ 331sq.
 ideogram 200
 ideographic writing system 200

- ideology 403, 418, 449, 533, nationalist ~ 38, pan-Germanic ~ 419
- idiom, idiomatic expression 367, 368. cf. phrase-ologism
- IEO (Institut d'Estudis Occitans) 71, 481
- Île-de-France 120, 444, 448sq.
- illiteracy 468, 475
- illocutionary act **7**, 150, 278
- illumination 502
- Illyria 406, 415
- immediacy-distance continuum **140**–158
- immediacy (linguistic/communicative ~) 139–143, 158, 422, 440, 443, 472
- immigrant language 69
- imperative 198, ~ negation 283, present ~ 280
- imperfect (past ~) 233, **253**sq., 268, ~ indicative 253, 280, ~ subjunctive 281
- imperfective 253, **268**, 275
- impersonal construction 273
- implement (writing ~) 500
- implicature 150
- implosive [l] 218
- improper compound 358, ~ derivation 355
- inalienability 437
- inanimate 245, 292, 329
- incomprehensibility of Latin 443, mutual ~ 471
- incomprehension 105, 443
- incunabula* 497
- indefinido* 269
- indefinite article 262, ~ determiner 261, ~ plural 227, ~ pronoun 244, **249**
- independent clause 259
- indexical sign 8
- indicative 303, past perfect ~ 280, pluperfect ~ 281, present ~ **251**sq., 280
- indigenous language 408, 410, 420, 454
- indirect complement 292, ~ object 260, **277**, 298
- individualisation 150, 312
- individuality 5, 7
- individual knowledge 314
- individuation 262
- Indo-China 476
- Indo-European 299, ~ language **2**sq., 30, 239, 406, ~ philology 178, ~ studies 28, 31, Proto~ 299
- induction 135
- industrialisation 477
- infallibility (of language) 37
- inferred meaning 348
- infinitive 270, ~ clause 304, ~ nominalisation 355
- accusative with ~ 411, active present ~, future ~, inflected ~, passive present ~, perfect ~ 283sq.
- infinitivo pessoal* 303
- infix 226, 286
- inflected form **224**, 227, 233, ~ infinitive 283
- inflectedness **233**sq., 254, 297, reduction of ~ 230, 233, 240
- inflection 163, **224**, **228**, 232–243, 260, 266, 270, 286, 288, adjectival ~ 234, **241**, nominal ~ 228, **234**–238, obligatory ~ 228, verbal ~ 228, 233
- inflectional affix **226**, 233, 243, 295, 310, ~ class **229**, 234–238, 252, ~ language 232, ~ marker/ marking 262, 286, ~ morphology 224, ~ paradigm 229
- informal conversation 146, ~ variety 133–**136**, 143
- information structure 148, 198, 285, **300**sq., ~ technology 487, 515, referential ~ 6, 8
- informational emphasis 301
- infrastructure **401**, 403, 406, 410, 421sq., 430
- infrastructural cohesion 422, ~ disintegration 381, 412–**416**
- initialism 361
- initial phase 274, pre~ 274
- injunctive 149, **259**
- injury (brain ~) 166
- innovation 121, 369, 469, 475, 481
- derivational ~ 387, grammatical ~ 173, lexical ~ **123**, 125, 385, phonetic ~ 422, 430, semantic ~ 365, 387
- Romance ~ 270, **282**sq., 423, 428
- inscription(s) 411, 425sq., 496, 498, ~ of Pompei 204
- instigator* 260, 274, **292**, 294
- Institut d'Estudis Catalans 85
- Institut d'Estudis Occitans (IEO) 71
- instrument 293
- integration 141, 375sq., formal ~ 374
- intellectualisation 480
- intelligibility **13**sq., 109, 152
- intensification (semantic ~) 169, 296, 360
- intensity 196sq.
- intentionality 294
- intentional source 494sq.
- intention (pragmatic ~) 149, 151
- interdental consonant 181, ~ fricative 216
- interference (linguistic ~) **17**, 20, 82, **102**, 105, 134, 267, 369, 376, 408, 442sq.
- interjection 141, 225
- interlanguage 42, **105**sq.

- interlinear glosses 446
- internal changes 170, ~ characteristics **19**, 23,
524, ~ differentiation 472, ~ evolution 398,
~ hiatus 185, ~ history 37, 163, 405, ~
metonymy 336, ~ periodisation 170, 172
- internationalism 378
- International Phonetic Association (IPA) 182, 183
- interpretative edition 510sq.
- interpunct 519
- interregional communication 421
- interrogative pronoun 244, 304, ~ sentence 149,
198, 259
- interview (job ~) 140, 146, 158
- intervocalic [n, l] 219, ~ plosive 66, 97sq., 214sq.,
409
- intimacy 130, 246sq.
- intonation 196sq.
- intralinguistic borrowing **378**, 388
- intransitive verb **276**sq.
- introspection 11, 317
- intuition 154sq.
- invasion 407, 415, Germanic ~ 371, 415, **419**,
Slavonic ~ 88, 418
- inversion 194, axiological ~ 338, syllable ~ 363
- invisible hand (A. Smith) 20
- IPA (International Phonetic Association) 182
- irony 198
- irrealis **282**, 303
- irregular verb 253
- isochrony (syllabic ~) 195
- isogloss **64**–66, 96, 98, 417, 420, ~ bundle **64**, 118,
121
- isolation, morphological ~, morpho-semantic ~ 346
- isomorphism (morphematic ~) 230
- Istro-Romanian 59, 87, 92, **125**
- Italian 56–60, 67, 69, **74**–80, 89, 110, 127, 135, **171**,
182, 190, 236, 238, 248, 251, 254, 268, 282, 303,
305, 420, 459, 469sq., 528
~ norm 463, ~ consonant/vowel system 190
Franco-~ 69, Gallo-~ 77, 417, 420, northern ~
209 popular ~ 136, standard ~ 55, 57, 64, 74,
296sq., 307, **450**, 455, **462**, standardisation of
~ 483
italiano dell'uso medio, italiano popolare,
italiano regionale, italiano standard 135sq.
- Italianism 103, **373**, 434
- italianisti* 462
- Italic system (vowels) 206sq.
- Italo-Romance 96, **100**, 409, 431sq., 439, ~ atlas 35,
~ dialect **57**, 64, 67, 74–77, 92, 110, 124, 127,
136, 209, 218, 261, 424, 430, 439, 450, 456, 458,
474, 528
- Italy 74sq., 408, 420, 442, 444, 450, 455–458, 462,
512, 528sq.
central ~ 438, northern ~ 67, 69, **96**, 406, 434,
regionalisation of ~ 474, Republic of ~ 455,
southern ~ 269, 408, 433sq., 438
- item-based model 111
- ## J
- Japan 529
- jargon **136**, 363
- Jerusalem 69, **433**
- Jews 83, expulsion of ~ **454**, 464
- job interview 140, 146, 158
- journal (scientific ~) 530
- Judeo-Spanish 83, 91, **125**, 455
- judgement 26, acceptability ~ 11
- judicial documents **441**sq., 466
- Junggrammatiker* **32**, 178
- Jura 70
- Justinian Plague 413, 416
- juxtaposition 302
- ## K
- key word in context (KWIC) 520
- kharja* **425**, 502
- knowledge, cultural ~ 314, 334, encyclopedic ~ 151,
321, 325sq., implicit ~ 11, individual ~ 314,
shared ~ 110, 151, 311, 326, specialised ~ 21,
459, 461, 472, **491**–494, 504
- Kulturdialekt*. cf. elaborated dialect
- KWIC (key word in context) 520
- ## L
- labial consonant 216sq.
- labialisation **182**, 193
- labiodental consonant **181**, 189–191, 220
- labiovelar consonant 220
- Lachmann method 510
- Ladin 33, 59, 64, 67, **79**, 90, 116, 239, 394, 402, 460,
470, 474, 483, ~ Dolomitan 81, Dolomitic ~ 60,
74, **80**, 89, 92, 387, standard ~ 81, questione
ladina 80
- Laienlinguistik*. cf. lay/folk linguistics
- lais* 440, 489
- langage* (Saussure) 11
- language **5**, **55**, 88, 91, 110, ~ acquisition 44, ~ by
development 88, 92, ~ by distance **56**sq., 88,
139, ~ by elaboration 56, **90**, ~ capacity 7, 11,

- ~ exchange 434, ~ family 524, ~ legislation 466, ~ loss 410, ~ manifestation 524, ~ mixing 103, ~ origin 464, ~ region 398, ~ teaching 533, ~ use 461
- analytical ~ 234, artificial ~ 307, distance ~ 474, donor ~ 102, elaborated ~ 474, immigrant ~ 69, literary ~ 314, minimal ~ 434, natural ~ 327, origin/emergence of ~ 16, recipient ~ 102, sectorial ~ 478, semi-artificial ~ 426, source ~ 374, 376, specialised ~ 147, synthetic ~ 233, target ~ 374, 376, territorial ~ 92
- cf. academy, architecture, change, colonial, conflict, contact, culture, domain, endangered, evolution, extinction (death), functional, historical, indigenous, learning, minority, mixed, national, particular, planning, politics, processing, regional, second, sign, umbrella
- langue* **10**sq., 30, 151, 153, 157, 175, 179, 223, 320, 323
 - ~ *historique* (Coseriu). cf. historical
- Languedoc 71, 466
- Languedocien 70sq., 119, 435, 444
- lapsus calami* 504
- La Réunion 62, 106
- La Spezia-Rimini line **64**–67, 96, 417, 420
- lateral 180, **185**, 189–194, 220
- Latin 30, 130, 219, 370, 405sq., 410, 412, 415, 426–428, 432, 441sq., 445, 450, 458sq., 464, 466, 482, 513, 528sq., ~ alphabet 200, 470, ~ base 388, ~ context 423, ~ model 400, 407, 442, 444, 488, ~ regionalisation 410–412, 419, ~ root 360, ~ toponym 383, ~ writing 440
- Archaic ~ 172, Classical ~ 172, 239, 288, 414, 427–429, creolised ~ 409, fragmentation of ~ 409, incomprehensibility of ~ 443, Late ~ 172, 371, 419, medieval ~ 414, non-standard ~ 411, normative ~ 423, 427, **428**, post-Classical ~ 172, pre-~ 383, rustic ~ 289, 423, 428, 429, spoken ~ **173**, 411, 414, spoken late ~ **3**, 100, 104, 112, 124, 221, 350, 387sq., 409, 411sq., 420, standard ~ 409, **428**, vernacular ~ 423, Vulgar ~ 173, written ~ 104, 387, 414, 421, 427
- abandonment of ~ 407, 443, reform of ~ 414, 427–429, standardisation of ~ 413, substitution of ~ 440sq.
- Latin America 90, 477
- Latinisation 58, 77, 88, 221sq., 406sq., 418, 427, 443sq., 449, de-Latinisation 415, re-~ 104, 221
- Latinism 104, 130, **221**sq., 316, 353, 360, 365, 370, 378, 387sq., 448
- Lautgesetze* 32
- law 492
 - Bartsch's ~ 210, customary ~ 493, 495
- lay linguistics 27
- layout 201
- leaflet 491, 497
- learned form/word 221sq., 378, ~ language 414, ~ spelling 222, ~ word/form **221**, semi-~ **222**, 360, semi-~ word/form 353
- learning (language ~) **105**, 314, 446, 526
- Lebenswelt* 347
- lecture (scientific/University ~) 139, 140, 143
- legal doctrine 492, ~ language 147, ~ text(s) 140, 146, 158sq., 433, 450, 513, ~ theory 492, ~ vocabulary 494
- legislation (language ~) 466
- lemma 313
- lemmatisation 520sq.
- lengthening (positional ~) 199
- length (vowel ~) 195, **196**sq., 205, 207, 411
- Leonese* 432, 444, 458
- letter
 - ~ patent 494
 - capital ~ 201, 500, official/private ~ 140–146, 158
- Lëtzebuergsch* 69, *Letzeburger Sproch* 104. cf. Luxembourg
- level, basic semantic ~ 329, historical ~ (Coseriu) 153
- lexeme 8, 112, **163**, 168, 224–227, **230**, 233, 258, 310–313, 324, 382, 426, ~ -inherent feature 228, 311
- archi- 328, free ~ 360. cf. common noun
- lexical base 43, 106, **310**, 350, 360, ~ borrowing **369**, 443, ~ change **330**, 385, 426, ~ connotation 131, ~ differentiation 412, ~ diversity 142, 493, ~ field 328, ~ form 164–167, **311**, 317–327, ~ innovation **123**, 125, 385, ~ meaning 9, 167, 278, 310, **311**, 317–320, 326–329, 342, ~ memory 364, ~ morpheme 360, ~ semantics 164, 166, **317**, ~ sign 8, ~ unity 358, ~ variant 107, ~ variation 126, ~ verb 298
- Lexical-Functional Grammar (LFG) 42
- lexicalisation 336, 341
- lexicalised compound 358, ~ word/form 243, 312
- lexically determined variation 230, 231

- lexicography **310**, 404, 445sq., 461, 464, 471,
 diachronic ~ 34, etymological ~ 35, 386, **389**,
 394, historical ~ 386–**389**, 394, 480, 524,
 synchronic ~ 34, 386, 480
 lexicology **310**, 386, 521, historical ~ **385**, 387
 lexicon 310
 lexis 142, 157, 164, 168, 224, 257, **310**, 349, 364, 411,
 513
 Lia Rumantscha 79
libraria 200, **501**
 life expectancy 478
 Ligurian 210
Limousin 119, ~*Auvergnat* 70, *Arverno*~ 444
 linear marker/marking 285, ~ structure/order 287,
 492
 linearity **12**sq., 257, 285sq., 290
lingua franca (Mediterranean ~) 62, 69, 106, 406sq.,
 433sq., 476
 linguist 27
 linguistic awareness 429, 453, ~ boundary 417,
 434, ~ continuum 89, 95, 99, **121**, 447, 472, ~
 culture 307, **445**, ~ cycle 281, ~ differentiation
 417, 419, ~ discontinuity 458, ~ polyphony
 269, ~ sign 8, 318, 321, 397, 517, ~ theory 524,
 ~ thought 24–27, 401, **404**sq., 423, 439, 445,
 447, 461, 464, 467, 479, 483, 536, ~ unity **480**.
 cf. change (language ~)
 linguistics 27, 480, 528sq., Anglo-American ~ 129,
 applied ~ **41**, 533, cognitive ~ 42, 331, 345,
 corpus ~ **515**, 516, 521, 523, diachronic/
 historical ~ **19**, 23, external ~ 401, folk ~ 27,
 480, geo~ 386, **401**, 403, lay ~ 27, neuro~/
 psycho~ 22, **44**, 345, socio~ **40**, 132,
 synchronic ~ **19**, 39, text ~ 147, variational ~
 22, 41, 107, **110**, 128. cf. generativism
 link (neural ~) 318, 323sq., 337
 lip(s) (lower/upper ~) 181sq., 186
 liquid **185**, 188, 198, 214, *muta cum liquida* 196, 199
 Lisbon 457
 literacy 70, 170, 404, 451, 469, 473, 477–481, 502
 literary genre 488sq., ~ language 147, **314**, ~
 studies 528sq., ~ text(s) 156, 440, 444, 450,
 487, 501, 504, ~ theory 478, ~ tradition 478
 literature 436, 440, 442, dialectal ~ 478, folk ~ 472,
 religious ~ **490**sq., 497, secular ~ **489**, 497,
 travel ~ 493
 liturgical text 490
 Livinallongo 79
 Lleida 85
 loanword. cf. borrowing
 location 293
 locative 277
Logudorese 59, 77, 78
 Loi Guizot, ~ Toubon 479
 Lombardian 425
 Lombard(s) 210, 410, 419sq., 422, 444
 long-distance assimilation 193
 ~ communication 403, 422, 440sq., 451, 472
 lord(s) 438
Lorrain 70, 120, 444
 Lorraine 58, 69, 415, 444, 452
 loss of forms 124, word ~ **345**, 364, language ~ 410
 Louisiana 452
 low style 132, ~ variety (LV) 102, **132**
 lowercase 500
 Lusism 103
lusitanidade 475
 Lutheran Bible 460
 Luxembourg 68sq., 104, *Luxembourgish* 69. cf.
 Lëtzebuergsch
 luxury borrowing 377
 Lyon 73, 419, 422, 457
 lyric poetry 489
- M**
 Macau 61
 Madagascar 62
 Madeira Islands 61
 Madrid 457, 464
 Maghreb 62, 83, 125, 452sq., 476sq.
 main clause 302, ~ verb 270sq., 298
 majuscule 519
 Maltese 104
 manifestation (language ~) 151, 524
 manner 293, ~ of articulation 180, **183**, 186
 manual 48, 530sq., language learning ~ 445
 manuscript **496**, 498, 502, 506–510, ~ dating/ ~
 localisation 496, ~ production 506, reference
 ~ 510
Marca Hispanica 421
 marginal glosses 446
 marked case 240sq., 295
 markedness (linguistic ~) 472
 marker/marking, agglutinative ~ 230, 286,
 agreement ~ 245, argument ~ 227, 276,
 295–299, bilateral ~ 287, case ~ 239, 241,
 communicative ~ 154, constituent ~ 239,
 diaphasic/diastratic/diatopic ~ 152, 157, 159,
 514, diasystematic ~ 154, 157, **311**, 317, 324,
 discourse ~ 141, emotive ~ 141, focus ~ 239,

- gender ~ 375, grammatical ~ 233, 260, 286, 299, 302, 308, identity ~ 126, inflectional ~ 262, 286, linear ~ 285, maximal ~ 302, metatextual ~ 148, minimal ~ 302, noun ~ 233, number ~ 234, 375, positional ~ 248, 297, 299, pragmatic ~ 141, prepositional ~ 297, prosodic ~ 148, 264, **286**, role ~ 308, temporal ~ 148, unilateral ~ 287, variational ~ 157, verb ~ 43, zero ~ 43, 227, 286, **295**
- Martinique 62, 106
- masculine 237sq.
- mass
~ media 478
~ noun **234**, 262
- Massif Central 417
- material (support ~) 496, **498**, 500
- mathematico-astronomical/astrological science/
writing 492
- matronym 382
- Mauritania 409
- Mauritius 68, 106
- maxim (conversational ~) 150
- meaning 8, 164, 311, 324, 329sq., 358, compound ~ 358sq., conceptual ~ 329, contextual ~ 318, figurative ~ 358, 367sq., inferred ~ 348, lexical ~ 9, 167, 278, 310–**311**, 317–329, 342, restriction of ~ 333, weakening of ~ 340, word ~ **348**, word formation ~ **348**, 365
- media 147, 478, mass ~ 478, medium 137, 159
mediality 158
- medicine 438
- medico-biological science/writing 450, 492
- medieval English 103, ~ Francoprovençal 90,
~ French 126, 172, 217, ~ Gascon 217, ~
Italo-Romance **74**, 127, ~ language 437, ~
Latin 414, ~ Occitan 89, 217, ~ orthography
202, ~ punctuation 205, ~ *România* 172, 436, ~
source 417, 419, ~ textual genre 487, ~ written
language 445
- medium cf. media
- Megleno-Romanian 59, 87, 92, **125**, 418
- melody (sentence ~) 178
- memorisation 21, 323, ~ test 312
memory, collective ~ 151, conceptual ~,
declarative ~, episodic ~, event ~, explicit ~,
procedural ~, semantic ~ 167, 322–324, 326,
386, lexical ~ 364
- mendicant order 438
- mental referent 326, ~ representation 9, 163, 175,
319, 321
- merchant(s) 434sq., 438, 446, 502
- Merovingian(s) 422, 498, ~ cursive 500, ~ period
427
- mesolect 133
- metalinguistic evidence 25, 205, 401, 423, **429**, ~
texts 493, ~ witness 411, ~ work 466
- metaphony 193, 210, 231
- metaphor 15, 136, **331**sq., 337sq., 359, 380
- metaphorisation 313
- metatextual marker 148
- metathesis **194**, 214, 218
- metonymy 330–**336**, 359, 380
- Mexican Spanish 108
- Mexico 61, 81
- microtext 168, 423, 425
- middle style 132, ~ voice 272
- Middle Ages (High/Low ~) 170, 431
- migration 17, 82sq., 88, 124, 372, 382, 402, 406, 412,
418sq., 431–435, 450, 454, 463, 476, 483, 533
- Milan, Milanese 74, 455, 456–458
- military service 474
- miniature 331, 502
- minimal language 434, ~ pair 163, **176**, ~ system
of communication 105sq., ~ marker/markings
302
- Minimalist Program (MP) 42
- minority language 62, 68, 101, 408, **418**, 447sq.,
452, 474, 529
- minuscule 500, ~ script/Carolingian ~ 500sq.
- mistake 13, 15, **134**, 504, 508, 512
- mistrallienne* (*graphie* ~) 71
- mixed form 427, ~ language **104**, 369, 402
- mixing, code ~, language ~ 103
- mobility 62, 70, 108, 403, 413, 434, 447, 477, social
~ 477, 481
- modal conflict 303, ~ form 302, ~ function 282, ~
verb 270, **278**
- modalisation 149, 269
- modality **269**, 283, graphic/phonic ~ **21**, 137, 140,
sentence/utterance ~ 259, verbal ~ 298
- mode 129. cf. action
- model, cultural ~ 369, Latin ~ 400, 407, 442, 444,
488
- modern era/period 170, **451**, 456–471, 506
- modification (nominal/pronoun ~) 261, **263**–265, 270
- modularity 167
- modulation 141
- Moldova, Moldovan 87–90, 203, 456, 502
- Monaco 68
- monastic network 421

- monoglossia* 402
 monolingual dictionary 465
 monolingualism 101
 monophthongisation **210**–212, 411
 monosemy 351
 monotransitive verb 277
 monovalent verb **275**, 301
 mood 43, 250, 266–**270**, 303, action ~ 266
 Morocco 62
 morph 227, 233
 morpheme 227, 233, cumulative ~ 232, lexical ~ 360, zero ~ 227
 morphological adaptation 374sq., ~ asymmetry 230, ~ isolation 346, ~ motivation 8, ~ neutralisation 504, ~ structure 233, 311, 324
 morphology **163**, 167sq., 224sq., 308, 349, 513, compositional ~ 164, derivational ~ **164**, 243, 385, Euro~ 221, inflectional ~ 224
 morphome, morphomic variation 252sq.
 morpho-semantic isolation 346
 morphosyntactic change 426, ~ feature **228**, 234, 249, 266, **311**, ~ variation 443
 morphosyntax **243**, 513
 mortmain (law of ~) 437
Moselromania 415
 mother tongue 414, 472, 474
 motivation 8, 340, 368, etymological ~ 120, external ~ 369, formal ~ 332sq., **339**, morphological ~ 8, pragmatic ~ 153, semantic ~ 35, 326, 333
 motive (political ~) 533
 Moulins (edict of ~) 466
 mountain chain 64, 416sq., ~ name 312
mouvance 503–506, 511
 Mozambique 62, 86, 455, 475
 Mozarabic 489
 multilingualism 74, 324, 440, 450
 multi-word unit. cf. phraseologism
muta cum liquida 196, 199
- N**
Nähesprache (Koch/Oesterreicher). cf. language by proximity
 name of origin, by~, double~ system 381sq., Germanic ~ 381, 420, mountain ~ 312, single ~ 381
 personal ~. cf. anthroponym
 place ~. cf. toponym
 Naples 74, 438, 455, 457
 Napoleonic conquests 455
 nasal cavity 184, ~ consonant 66, **184**, 188–192, 220, ~ diphthong 187, ~ sound 186, ~ vowel 183, **187**, 191, 198, 208
 nasalisation **193**, 211
 nation 449, 474, 480, 532, ~ state 448, 456, 463
 nationalism 466
 nationalist ideology 38, ~ myth 449
 national language 57, 60, 402, 404, **448sq.**, 459, 461, 469, 471–474, 480, 529, 533
 Native American language 373, 471
 native speaker 154
 Nativity play 497
 natural language 327, ~ phonology 223
 Navarrese 83, 98, 215
 Neapolitan 64, 75, 444, 450, 456, 458
 necessity borrowing 377
 negation 249, 267, **279**, ~ particle 107, 271, 279, discontinuous/double ~ 279, imperative ~ 283
 negator (preverbal ~) 279
 Neogrammarians **32**, 37, 178
 Neolithic era 406
 neologism 15, **312**, 316sq., 475, 481
 network, conceptual ~ **323sq.**, 330, 344, metonymic ~ 336, monastic ~ 421, neural ~ 157, semantic ~ 311, 317sq., 327–**329**, social ~ 439, taxonomic ~ 336
 neural link 318, 323sq., 337, ~ network 157
 neurolinguistics, neurology 22, **44**
 neuro-psychology 166
 neuter 234, 237sq., 411
 neutralisation 69, 74, 81, 83, 112, 124–127, 206sq., 251, 402, 432, 445, 450sq., 459, 504, ~ of conjugation classes 229, ~ of number distinction 262
 graphemic/morphological ~ 504
 neutrality (diatopic ~) 426
 New Caledonia 106
 newspaper(s) 478, 497, 515, ~ article 140, 146, 158sq.
 New Zealand 529
 Nicaragua 61
 Nice 452
 nickname 382
 nobility 438
 nobles 502
 noeme 320
 nomadism 403
 nomen 320sq., 381
 nominal determination **261**–265, ~ inflection 228, 234sq., 238, ~ modification **263sq.**, ~ subject 296. cf. common noun

nominalisation 152, 352, infinitive ~ 355
 nominandum 321
nomina sacra 362
 nominative 237, 239, 240
 non-actual. cf. non-perfective
 non-assertive sentence 270
 non-conjugated verb form 284
 non-countable 238, 297
 non-count noun **234**, 262
 non-decompositionality 358
 non-defining relative clause 264
 non-participant referent 245, 250
 non-perfective 268
 non-professional writing 503
 non-restrictive relative clause 264
 non-standard 133sq., ~ variety 134, 398, 411, **471**, 481
 non-verbalised concept 324
 norm **134**sq., 459, 461, 479, cross-regional ~ 437, descriptive ~ 134, grammatical/orthographic ~ 444, prescriptive ~ 26, **134**, 461, usage ~ (Coseriu) **11**, 152, 366, violation of ~ 13
 normalisation 480, 511sq.
 Norman 70, ~ conquest 431–433, Anglo~ 69, 103, 433, 504
 Normandy 120
 normative grammar 164, 307, ~ Latin 423, 427, **428**
 notary 438, 503
 noun 225, ~ complement 264, 302, ~ marker/marking 233, ~ phrase (NP) 244, **259–262**, 263, 288, collective ~ 351, common ~ **312**, 313, 379sq., 427, count/countable ~ 234, deadjectival/deverbal ~ 352sq., mass ~ **234**, 262, non-count/non-countable ~ **234**, 262, 297, proper ~ 200, 261, 312sq., **379**sq., 384, 427, relational ~ 351
 novel 144
 nucleus (syllable ~) 195
 number 228, **234**, 236, 241, 244, 250sq., 266, ~ marker/marking 234, 375, argument ~ 366, speaker ~ 67, 401, word ~ 313
Nuorese 77
 nursery language 360

O

oath formula 425
 object **260**, 297, 302, ~ clause 303, ~ clitic **244**, **245**, 298, ~ complement 262, 276sq., ~ conjugation 283, **297**, ~ position 277

complement ~ 292, direct ~ 260, **277**, 296, 298, first/second ~ 239, **260**, 277, 291sq., 296sq., 299, indirect ~ 260, **277**, 298
 obligatory inflection 228, ~ subject expression 244, 251, 273, 295sq.
 oblique case **237–240**, absolute ~ 265
 obstruction 181, 183
 occasional word 312
 Occitan, langue d'oc 3, 31, 59–64, 67–68, **70–71**, 92, 118sq., 124, 402, 417, 421sq., 425, 433, 439, 442, 444, 466, 471, 474, 481, 483, 529, ~ lexicography 391, medieval/Old ~ 89, 217, 265, southern ~ 435
 Occitanism 373
 occlusive 183
 oc (langue d'~). cf. Occitan
 official status 60, 80
 oil domain 444, 458, dialects/langue d'~ **55**, 62, 67–70, 92, 119, 417, 419
 omission, article ~ 262, negation particle ~ 107, subject ~ 295sq.
 onomasiological change 346, ~ order 116, ~ system 328, ~ trend 35
 onomasiology 35, 120, **328**, 342, historical ~ **342**, 345–347, 386
 onomastic evidence 427
 onomastics 313, **379**sq., 385
 onomatopoeia **8**, 17, 386
 onset (syllable ~) 195
 onymisation 380
 opaqueness 312, **364**, 368
 open
 ~ access 531
 ~ syllable **195**, 207sq.
 openness (vowel ~) 183, **186**sq.
 opposition, functional ~ 176, phonological ~ 188sq., 205, 223
 oral conception **139**sq., 143, 425, ~ corpus/data 138, ~ sound 186, ~ source **137**, 473, 524, ~ vowel 198
 orality **140**, 288, 409, 428
 order, constituent ~ 43, 103, **288–290**, 296–304, 308, geolinguistic ~ **435**, mendicant ~ 438, onomasiological ~ 116
 Ordinance of Villers-Cotterêts 466
 organ (articulatory ~) 181
 organisation, social/sociopolitical ~ 402sq.
 Organon model (Bühler) 6
 Orient 406
 origin, language ~ 464, name of ~ 382

original 495, 503, 506, 513, ~ text(s) 508
 proximity to the ~ 512, reconstruction of the
 ~ 508, 510
 oronym 312
 orthographic form 322, ~ norm 444, ~ pronun-
 ciation 203, ~ reform 15, 202sq., 479
 orthography 202, **203**, 461sq., 467, medieval ~ 202,
 semi-phonetic ~ 462, standardised ~ 202
 oscillogram 175, 197sq.
 Osco-Umbrian 371, 408
 Ostrogoths 410, 413, 419

P

palatal consonant **182**, 189–191, 210, 220, ~ glide/
 vowel 182
 palatalisation 182, 193, 210, **214**–217, 223, 411
 palate (hard/soft ~) 180–184
 palato-alveolar 182, ~ affricate 216, ~ consonant
 182, 189–191, 220, ~ fricative 215
Palenquero 83, 106
 Palermo 457
 pamphlet 491, 497, 515
 Panama 61
 Papal States 438, 455
 paper 496, **499**, scientific ~ 146
 Papiamento **83**, 106, 300
 papyrus 496, **498**
 paradigm 229, verb ~ 250
 paradigmatic axis 13, 169
 Paraguay 61, 81, 454
 paraliturgical text 490
 paraphrase 242
 parasyntesis 356
 parataxis 140sq., **302**
 parchment 496, **499**
 Paris 38, 70, 448sq., 452, 457
 parish 437, ~ register 382, 491, **495**
 parole **10**–12, 30, 151, 153, 175, 223, 320sq., 523
 paronomasia 339
 paroxyton 199
 parsing 286, 291, 521
 participant role 292
 participial clause/construction 304
 participle (future/past/present ~) 226, 270, 284
 particle, discourse ~ 141, negation ~ 107, 271, 279
 particular languages (Coseriu) **10**, 153
 partitive article 262, 297, ~ determination 261
 part of speech 26, **225**, 311, 324, 521
 part-whole 332, **335**
passato remoto/passé simple 269

passive articulation 181, ~ present infinitive 284, ~
 reflexive 272, ~ vocabulary 314, ~ voice 266,
 270, **272**, 411
 diathetic ~ 273, 301, periphrastic/reflexive ~
 272, 273, synthetic ~ 284
 past 253, ~ anterior 282sq., ~ imperfect 233, 254, ~
 participle 226, ~ perfect indicative/subjunctive
 280sq., ~ subjunctive 281
 compound ~ **268**, 270, 283, compound ~
 perfect 267, double compound ~ **268**, 283,
 future in the ~ 268, 282sq., 411, simple ~
 268sq.
 patient **260**, 275, 278, 292–296, 301, 336, ~ subject
 289, 295
patois 55
 patriariate (urban ~) 421, 468
 patronym 382, 495
 pejorative 131, 340
 pentagon, enhanced semiotic ~ (Glessgen) 322,
 324, semiotic ~ (Raible) 320
 penultimate syllable 199
 perceptive phonetics **178**, 194
 perfect 275, ~ infinitive 284, ~ stem 284
 compound past ~ 267, conditional ~
 283, future ~ 282, past ~ indicative 280,
 periphrastic ~ 282sq., simple ~ 268
 perfective 268, 275
 performance 11, 12, 151
 performative illocutionary acts 7
 periodisation, historical ~ **173**, 405, internal ~ 170,
 172
 periphrasis 149, 242, **250**, 254, 255, 266, 270sq.,
 275, 280–282, 308, 411, verbal ~ 270sq.
 periphrastic future **255**, 267, 271, 275, 412, ~ passive
 272sq., ~ perfect 282sq., ~ verb 287
 Persian 30
 person 228, 244, **250**, 266, speech act ~ **293**, 297
 personal distance 246, 247, ~ pronoun 225,
 244–**246**, ~ writing 497
 personal name. cf. anthroponym
 Peru 61, 81, 412, 454
pharmacopoeia 492
 pharynx 181
 phase (verb ~), initial ~, post-terminal ~, pre-initial
 ~, progressive ~, terminal ~ 274sq., 268
 Philippines 83, 106, 475
 philology **404**sq., 487, 496, 498, 512–516, 529,
 digital ~ 515, editorial ~ 32sq., **507**, 523
 philosophical doctrine 492
 philosophy 492

- Phoenician 371, 400, ~ writing system 200
- phoneme **163**, **175**–177, 180, 188, 201, 204, 219, 223,
grapheme~ relationship 201sq., 204
- phonetic adaptation 374, ~ alphabet 182, ~
boundary 193, ~ chain 175, ~ contraction
282, ~ evolution 378, ~ form 326, ~
innovation 422, 430, ~ reduction 15, 281,
348, ~ segmentation 200, ~ sequence 319, ~
shortening 15, ~ signal 167, ~ transcription
182sq., 524, ~ variant 107, ~ variation 15,
114, 192, ~ weakness 123, **345**. cf. change
(phonetic ~)
- phonetics 112, **163sq.**, 167, 178, acoustic ~,
articulatory ~, auditory ~ **178**, 194, diachronic
~ 200, historical ~ 524, 525, perceptive ~ **178**,
194
- phonic code 137, **139**, 158, 483, ~ conception 140,
~ form 310, **321**, ~ utterance 140, ~ volume
196, 223
- phonographic writing system 200
- phonological change 178, **192**, ~ clitic 295, ~
form **9**, 166, 322, 324, 326, ~ image 320, ~
opposition 188sq., 205, 223, ~ representation
167, ~ system **188**, 192, ~ transformation 192,
220, ~ variant 107, **223**, ~ variation 178, 192
- phonologisation 179, **194**
- phonology 38, **163**, 164, 167, 178, natural ~ 223,
suprasegmental ~ 195
- phonotactic structure 195
- phrase 163, **257**, ~ head 265
adjective ~ (AP) **265**, 279, adverb ~ 279,
determiner ~ (DP) 288, noun ~ (NP) 244,
259–**263**, 288, verb ~ (VP) 258sq., **266**, 274,
291
- phrase 368. cf. phraseologism
- phraseologisation 326
- phraseologism 313, 362, **366**–369, 376, 379sq.,
387, 521
- phraseology **366**, 385, 389
- Phrase-Structure Grammar 257
- phylogeny 16
- Picard* 70, 120, 434, 444
- Picardy 444
- pidgin 42, **106**, 434, 455
- pidginisation 402, 409, 453
- Piedmont 64, 71
- Piedmontese 73, 210, 444, 450, Franco~ 490
- pilgrimage 448
- pitch (sound/syllable ~) **196sq.**
- place name. cf. toponym
- ~ of articulation **180sq.**, 188, 193, 198
- plague 435, 451, Justinian ~ 413, 416
- planning, discourse ~ 140, language ~ 404, 466,
479
- Plantagenet 432
- plosive **183**, 188–191, 220, intervocalic ~ 66, 97sq.,
214sq., 409
- pluperfect 281sq.
- plural 98, 228, 236, ~ of plural 238, ~ suffix 236
collective ~ 238, polite ~ 308, sigmatic ~ **97**,
215, 236, zero indefinite ~ 227
- pluralia tantum* 234
- pluralis maiestatis* 246
- plurality (non-universal ~) 249
- pluricentric codification 18, 127, **128**, 475–477
- plurilingualism 101
- poetry 460, epic ~ 440, 489, lyric ~ 489,
troubadour ~ 31sq., 71, 434, 437, 439, 489, 513
- politeness 20, 198, **246sq.**, ~ strategy 250
- polite plural 308
- political, ~ language 147, ~ motive 533, ~
structures 430
- politics 479, **531**, language ~ 404, 453, 474, 479,
481
- polyglossia **101**, 402
- polymorphy 14, 254
- polyphony (linguistic ~) 269
- polysemy 9, 232, **311**, 321, 326sq., 348, 364, 376
- Pompeii 204
- popular connotation 315, ~ etymology 331sq.,
339, 386, ~ French 136, 290, ~ Italian 136, ~
Spanish 135, ~ variety 130sq., **135sq.**, 143,
307, 474, 481, 524
- population, ~ density **403**, 416, 422, 454, ~ growth
418, 457, 473, 477, ~ mixing 124, 432
- portmanteau* word 362
- Port-Royal grammar 27, 305, 462
- Portugal 86, 191, 431, 434, 438, **455**, 458, 528
- Portuguese 56, 59–61, 67, **86**–90, 125, 128, 191sq.,
210, 255, 261, 417, 434, 436, 459, 470, 475, 512,
528, ~ consonant/vowel system 191
- Galician~ 439, Gallego~ 3, 83, **85**, 425, 444,
489, standard ~ 432, standardisation of ~ 483
- position 260, **286**, 296, ~ of articulation 187
accent ~ 241, adverb ~ 279, affix ~ 226, first
~ 289, 300, object ~ 277, postverbal ~ 295,
preverbal ~ 295, rule of ~ 211, tongue ~ 186,
unmarked ~ 295, verb ~ 301, word ~ 286,
word-final ~ 212

- positional environment 179, ~ freedom 239, 300, ~ lengthening 199, ~ marker/marking 248, 297, 299, ~ rule 168, ~ variant 177, 202
- positive (degree of comparison) 242
- possessive determiner 247, ~ pronoun 244, **247**, stressed ~ 247
- posteriority 267
- postposition 263
- post-terminal phase 268, 274sq.
- post-tonic vowel 213
- postverbal position 295
- potential action 269
- potentialis 282
- pragmatic connotation 150, 169, 361, 423, 489, ~ foundation 149, ~ goal/motivation 153, ~ intention 149, 151, ~ marker/marking 141
- pragmatics 428
- pragmatics 41, 149–151, 164
- Prague school/circle 22, 38, 177
- predicate 257–261, **266**, 270, 291, 294
- predication 259
- prefix 226, 276, **286**, 356
- prefixation 286, **350**, 356
- pre-Hispanic languages 410
- pre-initial phase 275
- pre-Latin 383, 388, 409sq.
- prepalatal consonant 182
- preposition 225, 248, 277, 292, 299
- prepositional construction 264, ~ expression 226, ~ marker/marking 297
- prescriptive grammar 307sq., ~ norm 26, **134**, 461
- present imperative 280, ~ indicative **251**sq., 280, ~ infinitive/~ participle 284, ~ progressive 270, ~ stem 284, ~ subjunctive 280
- presentative construction 142, ~ verb 301
- present-day period 473
- press (printing ~) 459, 493, 497, 500
- pre-standardisation 398, 446, **504**
- prestige 102, 108, 130–**135**, 142, 144, 155–160, 308, 369, 378, 408, 410, 433, 439, 445, 447–450, 452, 458, 465, 471sq., 476
- pre-textual period 172, ~ Romance 426
- pretonic vowel 66
- preverbal negator 279, ~ position 295
- primary dialect 83, 92, **108**, **112**, 113, 117, 124, 126, 129, 474, 483
- princely court **438**, 457sq., 462
- principle (cooperative ~) (Grice) 41, **150**
- Principles and Parameters model 42
- printed text/characters 460, 497, 502, 511
- printer 459, 497
- printing 497, ~ press 459, 493, 497, 500, woodblock ~ 496
- private writing 459, 506
- procedural memory 167
- processing, double ~ 364, language ~ 164–169, 327, 330, 348, vocabulary ~ 348
- proclisis 298
- production (level of ~) 153
- productivity 361–365
- professional variety 137
- progression (of an action) 148, 268
- progressive aspect 271, ~ assimilation 193, ~ phase 274sq., present ~ 270
- pronominal delocutive 362, ~ redundancy 297, ~ repetition 301, ~ subject 290, ~ verb 277
- pronominalisation 245, 265
- pronoun 225, 238, **243**–247, 265sq., ~ modification 265
 clitic ~ 245, **255**, 295sq., clitic subject ~ 295, demonstrative ~ 225, 244, **247**, 262, full ~ **244**, identifier ~ 249, indefinite ~ 244, **249**, interrogative ~ 244, 304, personal ~ 225, 244–**246**, possessive ~ 244, **247**, quantifier ~ 249, reflexive ~ 246, relative ~ 244, 248, 302, stressed ~ 244sq., 295, subject ~ 296, unstressed ~ **245**, 295, zero~ 245
- pronunciation 202, 444, ~ reform 428, orthographic ~ 203
- proparoxyton **199**, 213
- proper noun 200, 261, 312sq., **379**, 380, 427, deanthroponymic ~ 380, delexical ~ **380**, 384, detoponymic ~ 380
- proposition 258, 259
- prose, ~ romance 489, religious ~ 493, scientific ~ 443, specialised ~ **491**, 493, 497
- pro-sentence form 259
- prosodic condition 286, ~ curve 197sq., ~ factor 298, ~ group 197, ~ marker/marking 148, 264, **286**, ~ transcription 197
- prosody 178, **195**–199, 205, 298, 301
- Protestant Reformation 457, 460, 491
- Proto-Indo-European 299
- Proto-Romance **173**sq., 254sq., 409, 421
- prototype 333, semantic ~ 329
- prototypicality 294
- prototypical subject 300
- provenance (text) 512
- Provençal 70sq., 119, 444, 481
 old ~ (= Occitan) 71

Provence 452
 proverb (collection of ~) 493
 proximal 247
 proximity, communicative/linguistic **137**–140,
 143, degree of ~ 247sq., formal ~ **339**, 376,
 hierarchy of ~, relative ~ 266, semantic ~ 327,
 329
 proximity (temporal, spatial, causal ~) 8, 330sq.,
 334
 pseudo-cleft 301, ~coordination 271, ~reflexivity
 273
 psycholinguistics 22, **44**, 345
 psycholinguistic test 166, 318
 psychology 44, *Gestalt* ~ 293, neuro~ 166
 psychomechanics 36
 Puerto Rico 61, 81, 102, 454
 punctuation 201, 205, 501, medieval ~ 205, 519
 Punic 371, 408
Putér (Romansh) 78
 Pyrenees 68, 417

Q

qualifying adjective 261, **262**–264
 quality (vowel ~) **180**, 205sq.
 quantification **249**, 515sq., 520, 523
 quantifier 249
 quantity (vowel ~) 195
 Quebec 61, 68sq., 90, 125, 127, 476, 528
 Quechua 81, 454
questione della lingua 462sq.
questione ladina 80
 quill 500

R

Raetia 80
 Real Academia española 306, 467
 realisation of a message 139
 reality, abstract/concrete ~ 320, extralinguistic ~
 17, 318sq.
 receiver 6, 292
 recessive diathesis **272**sq., 295, 301
 recipient 260, **292**, ~ language 102
 reciprocal assimilation 193, 432
Reconquista 81sq., 85sq., 124, 421, 431, 438, 449,
 455, 463
 reconstitution (text ~) 508
 reconstruction 255, 387, 412, 430, 507, comparative
 ~ 173, 524, text ~ 508, 510
 recording 524
 rection 228

recto 499
 reduction of case distinction **238**–241, ~ of
 declension classes **229**, 411, ~ of inflectedness
 230, 233, 240, ~ of inflectional classes 236, ~
 of valency 272, 295
 phonetic ~ 15, 281, 348, semantic ~ 248
 redundancy 14, pronominal ~ 297
 reduplication 226, **360**
 reference 261, ~ manuscript 510
 everyday ~ 329, **347**
 referent 312, 318–321, 324–326, 332sq., 379,
 abstract ~ 325, actual ~ 320, concrete ~ **325**,
 326, extralinguistic ~ 312, mental ~ 326,
 non-participant ~ 245, 250
 referential information 6, 8, self~ 7, **10**
 reflexive construction/passive 272sq., ~ pronoun
 246
 reflexivity (pseudo~) 273
 reform of Latin 414, **427**, 429
 Carolingian ~ 414, **427**, 443, 500, Cluniac ~
 442, orthographic ~ 15, 202sq., 479, pronun-
 ciation ~ 428
 Reformation, Counter~-/Protestant ~ 457, 460, 491
 reformulation 141
 regiolect **108**, 125–130
 region, border ~ 415, language ~ 398
 regional connotation 315sq., ~ dialect
 55–58, 62, 67, 82, 88–91, 108, 112, 130, 145, ~
 differentiation 288, 409, **412**, 420–422, 430,
 ~ language 75, 452, 466, 472, ~ variety 398,
 444, **447**, 458
 cf. diatopic variation
 regionalisation of Latin 410–412, 419
 regionalism 126
 register 109, **129**, 130, 147, 159, 160, 316, 423, 428,
 478, parish ~ 382, 491, **495**
 re-grammaticalisation 169, **296**
 regressive assimilation 193, ~ derivation 359
 reinforcement 341, ~ of temporality 308, semantic
 ~ 169
 re-Latinisation 104, 221, 427
 relational noun 351
 relative chronology 179, 430, 524, ~ clause **264**,
 302, ~ construction 263, ~ pronoun 244, **248**,
 302, ~ proximity 266, ~ subordination 304
 reliability 525
 religion 403
 religious culture 403, ~ language 147, ~ literature
 490sq., 497, ~ structures 430, ~ text/writing
 440, 450, 460, 493

- remote taxynomy 332
 Renaissance 170, 457, 493, Carolingian ~ 414,
 427–429, 443, 500
renaiçença 85
 repetition (pronominal ~) 301
 representation, mental ~ 9, 163, 175, 319, 321,
 phonological ~ 167
 representativeness 515
 reproducibility 525
 re-Romanisation 470
 resemblance, formal ~ 339, visual ~ 8
 restriction of meaning 333
 diasystematic ~ 311, geographical ~ 472,
 semantic ~ 368, usage ~ 369
 restrictive relative clause 264
 restructuring of inflectional classes 238, ~
 of nominal inflection 234, 236sq., ~ of
 phonological system 194, ~ of subordinate
 clauses 249, ~ of verbal system 284
 result 293, 352
 retroflex consonant 182
 Revolution (French ~) 437, 442, 452sq.
rexurdimento 86
 Rhaeto-Romance 80. cf. Romansh
 rhematisation **301**
 rheme 285, 290, **300**
 Rhetia 406sq.
 rhetoric 467, 492, 495
 Rhodian Provençal 71, 481
 rhyme 195, 205
 rhythm 197
 ridge (alveolar ~) 181sq.
 rising diphthong 187
 role, actantial ~ (Tesiñère) 292, case ~ 292, marker/
 marking ~ 308, participant ~ 292, semantic ~
 260, 273, **292–294**, 300, 336, syntactic ~ **260**,
 273, 292, 294, 300, 308, theta ~ 292
 Roma-Ancona line 64, 66
 Roman 444, 450, 458, ~ period 170, **405**
 romance (Arthurian ~, prose/vers ~) 437, 440, 489
 Romance 428sq., ~ element 423, 425–427, 441,
 ~ innovation 270, **282**, 283, 423, 428, ~
 linguistics/studies 31sq., 40, 480, 526–530,
 535, ~ texts/writing 439sq., ~ toponym 384, ~
 vocabulary 388, ~ vowel system 207, ~ writing
 425, 500
 Daco~ **100**, 239, 431, early ~ 4, 173, 414,
 Gallo~ **62**, 77, **100**, 118, 120, 295, 409, 417, 431,
 432, 438sq., Ibero~ 417, 431sq., Italo~ 64,
 67, 74–77, 92, 96, **100**, 110, 124, 127, 136, 209,
 218, 409, 424, 430, 431sq., 439, 450, 456, 458,
 474, 528
 pre-textual ~ 423, 426, Proto~ **173**sq., 254,
 255, 409, 421, written ~ 423, 436
Romandie 69
 Roman Empire 58, 170, 402, 405, 410, 415, decline/
 disintegration/fall/fragmentation of ~ 409,
 414–416, Eastern/Western ~ 415, 418
Romani 373
 Romania (= country) 58, **87**, 88, 418, 456, 528
România 347, ~ *continua* **58–61**, 103, 416–419, 435,
 ~ *creolica* 29, 106, ~ *nova* 29, **61**, 81, 125, 382,
 475, ~ *submersa* **60**sq., 402
 boundaries of the ~ 416, Eastern ~ **96**, 100,
 212, 215, Eastern Alpine ~ 80, 100, emergence
 of the ~ 412, extra-European ~ **36**, 61, 127,
 475, fragmentation of the ~ 38, **400**, 408,
 medieval ~ 172, 436, modern/present-day ~
 172, Western ~ **96**, 100, 215, 439
 Romanian 57, 60, 67, **87–89**, 96, 104, 125, 236,
 238, 243, 248, 262sq., 297, 303, 372sq., 393,
 416–418, 425, 460, 470, 512, 528, ~ case system
 241, ~ territory 432
 Aromanian 59, 87, 92, **125**, 402, 418, Daco~
 59, 88, **125**, Istro~ 59, 87, 92, **125**, Megleno~
 59, 87, 92, **125**, 418, south Danubian ~ 418,
 standardisation of ~ 483
 Romanisation 412, re~ 470
 Romansh 59, 60, 64, 67, 69, **78–80**, 89, 92, 209sq.,
 297, 394, 402, 425, 439, 460, 470, 474, 481, 483,
 528, ~ dialect 78, standard ~ 79, **481**
 Romanticism 28
 Rome 74, 456sq.
 root 233, ~allomorphy 231, Latin ~ 360
 roundedness (vowels) 186
 Roussillon 85, 452
 routinisation 341
 royal chancery 402, 442, **444**, 449, 458, 462, ~ court
 75, 127, 438, 444, 448sq., 459–463, **466**, 468,
 482, ~ domain 452
 rubric 502
 rule of position 211, genre ~ 152, positional ~ 168
Rumantsch Grischun 79, 387, 481
 Russian 87, 372
 rustic Latin 289, 423, **428**, 429
- S**
 Saarland 415
Sachverhaltsausdruck 258
 Saint Pierre and Miquelon 62

- Salentino 116, 206
 salience 244, 293sq., 329, **337**
 salient feature **294**, 325, 327, 337, 343, 380
 Sanskrit 27, 30, 178, 239
 Santiago de Compostela 433
 São Tomé-and-Príncipe 61, 86, 106
 Sardinia 59, 64, 74, 406, 416
 Sardinian 59sq., 64, 67, 74, **77**sq., 89sq., 92, 116,
 243, 394, 439, 471, 474, 483
 Sardinian system (vowels) 206sq.
 satire 489
 Savoy 452
 scholasticism 26
 school 403, 406, 429, 437, 502
 schooling 476, compulsory ~ 469, 474, 479
 schwa-dropping 193
 science (mathematical-astronomical ~) 492
 scientific language **144**, 147, 317, ~ prose 443, ~
 terminology 327, 371, **387**, 442, 460, ~ text(s)
 440, 442, 444, **491**, 501, ~ vocabulary 347
 scribe 495, public ~ 503
 script 200, Beneventan ~ 500, cursiva 200, Cyrillic
 ~ 88, 418, Gothic ~ 501, Greek ~ 25, Hebraic/
 Hebrew ~ 83, *libraria* 200, **501**, minuscule ~,
 Visigothic ~ 500
 script (= domain) 332, 334
scripta 126, **443–445**, 449sq., 458, ~ *latina rustica*
428, 442
scriptorium 437, 449, 502sq.
Scuola siciliana 442, 489, 499, 513
 second argument/actant 296, ~ language 62, **105**,
 474sq., ~ object 239, **260**, 292, **293**, 299, ~
 Vatican Council 477, 490
 secondary consonant cluster **217**sq., ~ dialect 83,
 108, 112, **124**sq., 432, 452sq., 455
 ~ grammaticalisation. cf. re-grammatical-
 isation
 sectorial language 478
 secularisation 437, 457
 secular literature **489**, 497
 segmentation of language domains **163**, 165,
 phonetic ~ 200
 self-referential 7, **10**
 self-reflection 25
 semantic borrowing **369**, 376, ~ cohesiveness
 367, ~ conflict 302, ~ content 285, ~
 delimitation 386, ~ family 331sq., ~ frame
 321, ~ group 278, ~ interference 339, ~ level
 329, ~ prototype 329, ~ reduction 248, ~
 relationship 311, 321, 330, ~ restriction 368, ~
 specificity 352, 354, ~ subject 290, ~ unit 366,
 ~ unity 226, ~ value 131, ~ weakening 169, ~
 weakness 345
 cf. change, connotation, field, innovation,
 intensification, memory, motivation, network,
 role
 semantics 34, 165, 385, cognitive ~ **328**sq.,
 grammatical ~ **164**, 278, historical ~ **330**, 342,
 345, 386, lexical ~ 164, 166, **317**, structural ~
 38, **328**
 semasiological dictionary 465, ~ method 342
 semasiology (historical ~) 347
 seme, sememe 328
 semi-learned **222**, 360
 semiology 8
 semi-oral writing 498
 semiotic ~ pentagon (Raible) 320, ~ square (Blank)
 321, ~ trapezium (Heger) 320, ~ triangle
 (Ogden/Richards) 319, enhanced ~ pentagon
 (Glessgen) 322, 324
 semiotics **8**, 318
 semi-phonetic orthography 462
 semi-uncial 500
 semivowel **182**, 198
 sender 6
 sensory phenomena 278
 sentence 163, 168, 257, **258–260**, 285, ~ head 288,
 ~ melody 178, ~ modality 269, ~ stress 197, ~
 structure **258**, 285, 366, 513,
 assertive ~ 149, 198, **258**sq., 270, complex
 ~ **259**, 285, 307, compound ~, exclamatory
 ~ 259, hypothetical conditional ~ 282,
 imperative ~ 198, injunctive ~ 149, **259**,
 interrogative ~ 149, 198, 259, non-assertive ~
 270, pro-~ 259, simple ~ 285
 separation, geographical ~ **17**, 20, 108, word ~ 200,
 357
 Serbo-Croat 74
 sermon 140, 146, 158, 429, **490**
 set phrase. cf. phraseologism
 Seville 457
 sex (biological ~) 129, **234**, 238, 316
 Seychelles 68, 106
 SGML (Standard General Markup Languages) 517
 shared association 327, ~ knowledge 110, 151, 311,
 326
 shift (topic ~) 296
 shortening, phonetic ~ 15, word ~ 223, 361
 sibilant 184, 215, 218
 Sicilian 64, 75, 116, 444, 450, 458

- Sicilian system (vowels) 206sq.
 Sicily 77, 406, 416, 433sq., 455
 sigmatic plural 97, 215, 236
 signature 437
signifiant / signifié (Saussure), signified, signifier 9sq., 317–320, 324, 326, 328, 330sq.
 sign(s) 8, 26, ~ language 10, **43**, 300, ~ system 8, 10
 grammatical ~ 8, graphic ~ 201, iconic ~ 8, 17,
 indexical ~, lexical ~ 8, linguistic ~ 8, 25, 318,
 321, 397, 517, symbolic ~ **8**, 17
 cf. semiotics/semiology
 similarity 331sq., **337**, 339
 simple future 254, 267, ~ past 268sq., ~ perfect 268, ~ sentence 285, ~ word/form 250, 359
 simplification of consonant clusters 214, **218**, ~ of diphthongs 254, ~ of relative subodination 304
s-impurum **213**, 222
 simultaneity 267
 single name 381
 singular 228
singularia tantum 234
 singularity 249
 situational context 110, 157, ~ specificity 157, ~ variety 129
 situation (communicative ~) 109, **131**, 140, 144, 151
 slang 136
 slavery 454
 Slavonic invasion 88, 418, ~ languages 61, 104, **372**, 415, 418, ~ peoples 415, 418
 Old (Church) ~ 88, 112, 418, 470, South ~ 372
 Slovenian 59
 social class 457, ~ distance 246sq., ~ function (of language) **6**, ~ group 7, 55, 108, 132, 135, ~ mobility 477, 481, ~ network 439, ~ organisation 402sq., ~ position 7
Société de linguistique romane (SLR) 530
 socio-cultural change 456, ~ group 6, 18, 108, ~ variation 482
 socio-historical context 344
 sociolect 18, **55**, 133, 136, 159, 307
 sociolectal variation 131, ~ variety 159
 sociolinguistics **40**, 132
 sociological context 470
 sociopolitical organisation 403
 soft palate 181sq., 184
 sonority 188, 195, 198
 sound **175**, 180, 201, ~-image (image acoustique) 9, ~ laws ('Lautgesetze') 32, 178, 386, ~ loss 214
 archetypal ~ 194, nasal/oral ~ 186, transitional ~ 194, voiced ~ 180
 cf. change (phonetic ~)
 source(s) concept 337, ~ criticism 521, 525, ~ language 374, 376
 individual ~ 496, intentional ~ 494sq.,
 medieval ~ 417, 419, oral ~ **137**, 473, 524,
 serial ~ 496, unintentional ~ 494, written ~ 3, **137**, 411, 423, 430, 451, 516, 524
 SOV (subject-object-verb) 288–290, 300
 space 159, 336, communicative ~ **58**, 60, 92, 110, 204, 459, 474, 477, geographical ~ **121**, 157, 402, stem ~ **226**, 233, 252, 350, 360, variational ~ 110
 Spain 82, 411, 433sq., 438, 442, 453, 529
 Spanish 59–61, 67, 72, **81**–84, 89sq., 124, 127sq., 135, 171, 190, 251, 254, 269, 297, 301, 303, 306, 372, 393, 412, 417, 425, 433sq., 439, 450, 453, 459, 469sq., 475, 512, 528
 consonant/vowel system 190sq.
 American ~ 57, 125, 454, Andalusian ~ 84, 108, **124**, Canarian ~, Castilian ~ 108, Judeo~ 83, **125**, 455, Mexican ~ 108, Peninsular ~ 57, 268, 410, 454, 475
 familiar ~ 475, popular ~ 135, standard ~ 432, standardisation of ~ 483, written ~ 439
 español coloquial, español literario, español popular, estándar, habla coloquial 135sq., 475
 spatial contiguity 335
 speaker 245, **250**, ~ awareness 56, 309, 312, 358, 363, ~ number 67, 401, ~ point of view 268, native ~ 154
 specialisation, semantic ~ 332, **333**, 343
 specialised knowledge 21, 459–461, **472**, 491–494, 504, ~ prose **491**, 493, 497, ~ terminology 20, 107, 312, 327, 403, 513
 specificity 262, 358, 364, semantic ~ 352, 354, situational ~ 157
 spectrogram 175
 speech **21**, 137, 139–142, 469, ~ act 10, 41, **150**, 278, ~ act person **293**, 297, ~ community 11, 29, 151, 326
 linearity of ~ **12**, 13, part of ~ 26, **225**, 311, 324
 spelling, archaic/learned ~ 222
 spirant 184
 spirantisation 193, 194
 spontaneous expression 472
Sprachhandeln 150
Sprechakt. cf. illocutionary/speech act

- square, capitals 500
 semiotic ~ (Blank) 321
stabilitas loci 125
 stability 358, **367**, 437, geographical ~ 125, 422,
 linguistic ~ 422
 standard French 55, 127, 307, ~ Galician 89,
 ~ Italian 55, 57, 64, 74, 296sq., 307, **450**,
 455, **462**, ~ Ladin 81, ~ Latin 409, **428**, ~
 Portuguese 432, ~ Romansh 79, **481**, ~
 Spanish 432
 ~ language/variety 58, 107, 112, 125–129,
 134, 143, **160**, 315, 317, 387, 398, 402, 404sq.,
 407, 413, 450, 458, 460, 466, 468sq., 472, 474,
 478–480, 497, 503, non-~ 133sq., 398, 411, 471,
 sub-/supra-~ 133, 135
 standardisation 20, 21, 57 (Stewart), 69, 77, 89, 92,
 126, 170, 202, 304, 307, 389, 400, 402, 444, 446,
 451, 461, 466, 468, 474sq., 481sq., 497, ~ of
 Italian 483, ~ of Latin 413, ~ of Portuguese,
 ~ of Romanian, ~ of Spanish 483, pre-~ 398,
 446, 504
 state 479, centralised ~ 407, 438, 462, nation ~
 448, 456, 463
 statement **258**, 285
 stative verb 278
 stem **226**sq., ~ allomorphy **227**, 233, ~ alternation
 226, **231**, 241, 243, 252sq., ~ space **226**, 233,
 252, 350, 360, ~ variation **252**sq., 350
 perfect/present ~ 284, stressed ~ 253,
 thematic ~ 230, 231, unstressed ~ 253, word
 ~ **231**, 350
 stemma **508**sq., 511
 stop 183, voiceless ~ 198
 stream, dorsal/ventral ~ 167
 stress 195–**197**, Latin ~ 199, sentence ~ 197, word
 ~ 197, 199
 stressed possessive 247, ~ pronoun 244, **245**, 295,
 ~ stem 253, ~ vowel 207–209
 structural borrowing **376**, ~ semantics 38, **328**
 structuralism **38**sq., 331, American ~ 22, **39**, British
 ~ 39, European ~ 22
 structure(s), language ~ (*structures de langue*)
 (Flydal) 128
 style **129**, 159, 367, high/low/middle ~ 132
 subject 239, **260**, 273, 277, 289, 290–296, 300, 302,
 ~ case **237**, 239sq., 286, ~ clitic 244–**245**, 296,
 ~ conjugation 283, 295, ~ expression 244,
 273, 295sq., 308, ~ marked 240, ~ omission
 295sq., ~ pronoun 296
 agent ~ 289, clitic ~ 251, **295**, double ~,
 full ~, nominal ~ 295sq., patient ~ 289,
 295, pronominal ~ 290, prototypical ~ 300,
 semantic ~ 290, syntactic ~ 295, unmarked ~
 299, zero marked ~ 227
 subject-object-verb (SOV) 288–290, 300
 subject-verb-object (SVO) 103, 288–290, 300
 subjunctive **269**sq., 302–304, past ~, past imperfect
 ~, past perfect ~, pluperfect ~, present ~
 280–282
 subordinate clause 249, 269, 286, **302**–304, ~
 relative clause 264
 subordinating conjunction 302, 304
 subordination 140, 285, **302**, relative ~ 304
 substandard 133
substratum **370**, 408–410, ~ hypothesis **409**, 419
 succession 334, **336**
 suffix 226, 262, 348, **350**, 355, adverb ~ 243,
 derivational ~ 356, plural ~ 236
 suffixal delocutive 362
 suffixation 286, **350**–352
 superlative (absolute/relative ~) 242
superstratum **371**, 419–421, ~ hypothesis **410**, 419,
 Germanic ~ **38**, 533
 supine 284
 suppletion 230–**232**, 242, 252sq.
 support material 496, **498**, 500
 supraregional form 447
 suprasegmental phonology 195
 supra-standard **133**, 135
 surface structure 358
surmiran, *sursilvan*, *sutsilvan* (Romansh) 78sq.
 SVO (subject-verb-object) 103, 289sq., 300
 Swiss German 79
 Switzerland 59, 68, **69**, 73, 90, 415, 453, 458, 530
 syllabic isochrony 195, ~ writing system 200
 syllable 178, **195**, 197, ~ inversion 363, ~ nucleus, ~
 onset 195, ~ pitch 197
 closed ~ **196**, 207sq., free ~, ideal ~ 195, open
 ~ **195**, 207sq.
 symbol **8**
 symbolic sign **8**, 17
 synchronic 23, ~ linguistics **19**, 39, ~ turn 40, ~
 variation 107
 synchrony **19**, 23
 syncope **194**, 213, 411
 syncretism 230, **232**sq., 236, 251
 synonym 317
 synonymy 311, 328sq.
 synoptic edition 510

syntactic behaviour 379, ~ borrowing **377**, 443, ~
 clitic 295, ~ context 224, ~ diversity 493, ~
 structure 285, ~ subject 295, ~ variant 107
 cf. complexity, role
 syntactocentrism 40, 42, **165**
 syntagmatic association 366, ~ axis **13**, 169,
 ~ borrowing 377, ~ compound 357, **358**,
 366, 368, ~ context **311**, 321, 324, 366. cf.
 phraseologism
 syntagm (free ~) 366
 syntax 39, 114, **163**, 167sq., **257**, 287, 308, 443,
 centrifugal/centripetal ~ 288, complex ~ 107,
 functional ~ 257
 synthetic form(s) 281sq., ~ future 280, 411, ~
 language 233, ~ passive 284
 Syria 406

T

tablet, clay ~, marble ~, stone ~, wax ~ 496, 498sq.
 taboo 339, 344
 tagging, morphological ~ 286, part of speech ~ 521
 TAM (tense, aspect, mood) **43**, 267, 275, 285
 target 228, ~ concept 337, ~ language 374, 376
 taxematic determination 286
 taxonomy, taxonomic change, ~ network 330–337
 teaching (language ~) 442, 529, 533
 technical language **144**, 147, 221, 378, 478, ~ text(s)
 440, 442, **491**, ~ vocabulary 347
 technology (information ~) 487, 515
 teeth (upper ~) 181
 television 70
 temporal contiguity 335, ~ deixis 271, ~ marker/
 marking 148
 temporality **267**sq., 293, reinforcement of ~ 308
 tenor **129**, 198
 tense 43, 228, **250**, 266sq., 270, 285, compound
 ~ 283
 terminal phase 274sq.
 terminology, scientific ~ 327, 371, **387**, 442, 460,
 specialised ~ 20, 107, 312, 327, 403, 513
territoires d'outre mer (TOM) **62**, 68, 476
 territorial fragmentation 413, ~ language 92
 territorialisation of variation 400
 tertiary dialect 126
 test, linguistic ~ 154, memorisation ~ 312, psycho-
 linguistic ~ 166, 318
 text(s) 153, **168**, 478, ~ linguistics 147, ~ loss 506, ~
 structure 149, ~ type 316
 textual criticism 512, ~ data 524, ~ genre 12, 132,
 143–156, 285, 304, 309, 347, 397, 401, 433sq.,

439–442, 447, 458sq., 472, 478, 487sq., 493,
 513–516, ~ heritage 506, ~ tradition 56, 285,
 403, 445, 490, 512, ~ transmission 506
 thematisation 301
 thematic stem 230sq., ~ vowel 227
 theme 275, 290, **300**
 theme-rheme structure 285, **300**
 theological writing, theology 491sq.
 theta role 292
 Thirty Years' War 457
 thought, (meta-)linguistic ~ 24–27, 401, **404**sq.,
 423, 439, 445, 447, 461, 464, 467, 479, 483, 536
Ticinese 64
 Ticino 69, 74, 415
 tilde 201
 time 170, 267, **293**, 336
 Toledo 463, kingdom of ~ 419, 420
 tongue 181, 185sq. (~ position)
 tonic vowel 206
 topic 148, 198, 289, 300, ~comment structure 198,
 ~ shift 296
 topicalisation **148**sq., 244, 289
 toponym **312**, 370, 379sq., 383–385, 408, 427, ~
 dictionary 384, denominal ~ 379, **384**
 totalitarianism 533
 Toulouse (kingdom of ~) 419
 Tours (Council of ~) **429**, 490
 trade 421, 436
 tradition, discourse ~ 109, 137, **144**–147, 151–157,
 285, 309, 403, 478, 488, 503sq., 513, literary ~
 478, textual ~ 56, 285, 403, 445, 490, 512
 transcription 507, 524, phonetic ~ **182**sq., 524,
 prosodic ~ 197
 transfer 102–105
 transitional sound 194, 218
 transitive verb, transitivity **276**sq.
 translation 440, 445sq., 490, 504, 508, Bible ~ 460,
 470
 transmission 495, material ~ 496, 507, textual ~
 506
 transparency 244, 271, 283, 312, 330, 349, 351, 364,
365, 368, 525
trapezium, semiotic ~ (Heger) 320, vowel ~ 187
 travel account/literature 459, 493, 515
Tre corone 450, 463, 467
 Trent (Council of ~) 382, **491**
 Trentino 81
 triangle (semiotic ~) (Ogden/Richards) 319
 triglossia 429
 trill 177, 180, **185**, 189–192, 194, 220

triphthong **185**, 210sq.
 trivalent verb 275
trobairitz 507
troubadour 442, ~ poetry 31sq., 71, 434, 437, 439, 489, 513
 truncation 361
 truth 258, 269
 Tunisia 60, 62, 111, 413, 415
 Turkism 103
 turn-taking 141
 Tuscan 64, 75, 77, 91, 116, 171, 444, **450**, 458, 463, 467sq.
 Tuscany 444
 typological approach 23, **45**, 106, 163, 388, ~ difference 232
 Tyrol (South ~) 59, 80sq.

U

Udine 81
 umbrella language **57**, 60, 78, 92, 110, 458sq.
 Umbrian 64, 406, Osco~ 371, 408
 unaccomplished. cf. imperfective
 unaccusative verb **277**, 295
 uncial, semi~ 500
 unconscious (collective ~) 344
 uncountable 294, 329
 unergative verb 277, **278**
 un-grammaticality 110
 ungrammaticality 141, 152
 Unicode 517
unidad del idioma 480
 unilateral marking 287
 unintelligibility 14
 unintentional source 494
 unit, bound ~ 227, distinctive ~ 176, free ~ 227, semantic ~ 366
 multi-word ~ cf. phraseologism
 United States 62, 81, **475sq.**, 528
 unity, formal ~ 226, 357, lexical ~ 358, linguistic ~ **480**, semantic ~ 226
 universality 249
 university 403sq., 437sq., 479, 502, 534
 Unix command 521
 unmarked case 241, 286, ~ construction 148, 300, ~ first object 297, ~ position 295, ~ subject 299
 unstressed determiner 244, ~ pronoun **245**, 295, ~ stem 253, ~ vowel 190sq., 194, 199, 207, **212**, 411
 urban area 137, 404, 481, ~ language/variety 365, 481

urbanisation 403, **477**, 481
 Uruguay 61
 usage 358, ~ context 349, 357, ~ norm (Coseriu) **11**, 152, 366, ~ restriction 369
 use 322, ~ context 346, language ~ 461
 usual context 368
 utterance 10, **11**, 102, 146, 150, 152, 158, 326, 523, ~ act 149, ~ context 267, ~ modality 259, graphic/ phonic ~ 140
uvula 181sq., 185
 uvular consonant **182**, 189, 191, 220

V

Val Badia, ~ Bregaglia 79, ~ d'Aosta 60, 64, 73sq., ~ d'Aran 72, ~ de Villé 70, ~ di Fassa 79, ~ d'Orbey 70, ~ Gardena, ~ Mesolcina, ~ Poschiavo 79
 Valencia 85, 432, 528
 Valencian 59sq., 67, 84sq., 125, 444
 valency 259, 267, 274, **275**, 291, 294, 311, 321, 366, ~ dictionary 276, ~ frame 259, 266, **275sq.**, 299, 311, 366, reduction of ~ 272, 295
 bivalent/monovalent/trivalent **275**, 291, 301
Valläder (Romansh) 78, 79
 value judgement 338, connotative ~, semantic ~, variational ~ 131
 Vandals 410
 variability **15**, 19, 45, 109, 287
 variable, variance 107, diatopic ~ 411
 variant **107sq.**, articulatory/combinatory ~ 177sq., free ~ 178sq., 207, 276, lexical ~ 107, phonetic/phonological ~ 107, 223, positional ~ **177**, 202, syntactic ~ 107
 variation 11, **15**, 107, 445, allophonic ~ 179, 192, chain of ~ **130**, 137, diachronic ~ **23**, 107, dialectal ~ 419, 432, free ~ **177**, graphic ~ 202, individual ~ 15, 107, lexical ~ 126, lexically determined ~ 230, 231, morphomic ~ 252, 253, morphosyntactic ~ 443, phonetic ~ 15, 114, 192, phonological ~ 178, **192**, socio-cultural ~ 482, sociolectal ~ 131, stem ~ 253, 350, stem space ~ 252, synchronic ~ 107, territorialisation of ~ 400
 cf. diaphasic, diastatic, diasystematic, diatopic, regional
 variational approach 4, ~ borrowing 130, ~ context 397, ~ dimension 157, ~ marker/marking 157, ~ space 110, ~ theory 128, ~ value 131
 cf. connotation, linguistics

- variety **56**, 107, 110, 156, 397, ~ according to use/
 user (Halliday) 129, ~ diaphasic 132, ~ of
 expansion 432
 acrolectal ~ 468, *banlieue* ~ 137, colloquial
 ~ 413, 472, 474, 524, dialectal ~ 62, **447**, elite
 ~ 307, familiar ~ **135**, 307, 317, high ~ (HV)
 132, informal ~ 133–**136**, 143, low ~ (LV) **132**,
 non-standard ~ 134, prestige ~ 378, 433,
 476, professional ~ 137, semi-artificial ~ 445,
 situational ~ 129, sociolectal ~ 159, urban ~
 365, written ~ 482
 cf. diaphasic, diastratic, diasystematic,
 diatopic, popular, regional
- Vaudois 71, 456
 Vegliot 208
 velar consonant **182**, 189–191, 220, ~ glide/~ vowel
 182
 velarisation **182**, 193, 214, 218
 velum 186
 Venetan **77**, 415, 425, 444, 450
 Venetian 57, 61, 64, 75, 218, 456, 458, 468, ~ rule
 433
 Veneto 77
 Venezuela 61
 Venice 74, 80, 455, 457
 verb 225, **249**, ~ marker/marking 43, ~ paradigm
 250, ~ position 301
 attributive ~ 277, copular ~ 277, deadjectival
 ~ 356, denominal ~ **354**, 356, deponent ~
 228, 272, **284**, ditransitive ~ 277, dynamic
 ~ 278, epistemic ~ **278**, 302sq., existential
 ~ **278**, 301, full ~ 298, high-frequency ~ 253,
 intransitive ~ 276sq., irregular ~ 253, lexical
 ~ 298, main ~ 270sq., 298, modal ~ 270,
 278, monotransitive ~ 277, non conjugated ~
 284, periphrastic ~ 287, presentative ~ 301,
 pronominal ~ 277, stative ~ 278, transitive ~
 276sq., unaccusative ~ **277**, 295, unergative
 ~ **277**sq.
 cf. auxiliary, phrase
- verbal adjective 284, ~ base 362, ~ compound
 357, ~ delocutive 362, ~ event 266sq., 274sq.,
 ~ inflection 228, 233, ~ modality 298, ~
 periphrasis 270sq.
- verbalisation 324–327
 verb-object-subject (VOS) 290
 verb-object (VO) 289, 296
verlan **363**, 386
 vernacular 428, 430, 459, 489, 490, ~ element 352,
 428, ~ Latin 423, written ~ 439
- verse romance 489
verso 499
 vertical change 332, ~ communication **413**, 430, ~
 dimension 145
 Villers-Cotterêts (Ordinance of ~) 466
 violation (of linguistic norm) 13
 virtual 322, the ~ 320
 Visigothic script 500
 Visigoths 410, 419sq.
- vocabulary 142, **310**, 315, 347, 513, ~ enrichment
 460, ~ loss 345, ~ processing 348, ~ size 21,
 314
 legal ~ 494, passive ~ 314, Romance ~ 388,
 scientific/technical ~ 347
- vocalisation 218
 vocative 239, 241
 voice **228**, 250, 266, 270, **272**, active/middle ~ 272,
 passive ~ 266, 270, **272**, 411
 voiced consonant 180, **185**, ~ sound 180
 voiceless consonant 180, **185**, ~ stop 198
 voicing 97sq., **194**, 214sq.
- volgari* 442
volgarizzamenti 440
 volume (phonic ~) 196, 223
 VOS (verb-object-subject) 290
 VO (verb-object) 289, 296
 vowel 180, **182**, 186, ~ back **182**, ~ backness
186sq., ~ height **186**sq., 205, ~ length
 195–**197**, 205, 207, 411, ~ openness 183,
186sq., ~ quality **180**, 205sq., ~ quantity
 195, ~ roundedness 186, ~ trapezium 187, ~
 unstressed 207
 back ~ 186sq., central ~ **182**, 186sq.,
 epenthetic ~ 213, extreme ~ 191, 208, **210**,
 front ~ **182**, 186, nasal ~ 183, **187**, 191, 198,
 208, oral ~ 198, palatal ~ 182, post-tonic
 ~ 213, pretonic ~ 66, stressed ~ 207–209,
 thematic ~ 227, tonic ~ 206, unstressed ~
 190sq., 194, 199, 207, **212**, 411, velar ~ 182
- vowel system 189sq., 205, 207sq., archaic/Sardinian
 ~, Balkan ~, Italic ~ **206**, Portuguese ~ 191
- vox media* 272
 VSO (verb-subject-object) 288, 290
 Vulgar Latin 173
- W**
- Waldensians 456
 Wallachia 456
 Wallonia 58, 415
 Walloon 70, 120, 297, 444

war 435sq., 448, 457, Hundred Years' ~ 433, 435, 452, Thirty Years' ~ 457
 watermark 500
 Wernicke's aphasia, ~ area 166
 witness 507, manuscript ~ 33, 507, 509–512, metalinguistic ~ 411
 women (emancipation of ~) 477
 woodblock printing 496
 word 225, ~ family 339, 348, **385**, ~ loss **345**, 364, ~ meaning **348**sq., ~ number 313, ~ position 286, ~ separation 200, 357, ~ shortening 361, ~ stem **231**, 350, ~ stress 197, 199, ~ structure 224, complex ~ 311, 363, graphic ~ 357, lexicalised ~, occasional ~ 312, *portmanteau* ~ 362, simple ~ 359
 cf. constituent order, function (word)
 Word-and-Paradigm 233
 word class 349, 351, ~ marker/marking 310, change of ~ 313, 348. cf. part of speech
 word-final consonant 217, 299, 411, ~ position 212
 word formation 163sq., 224, 330, **347**–349, 361–365, 377, ~ meaning **348**, 365
Wörter und Sachen 35, **120**, 342, 386

writer (professional ~) 443, 502
 writing **21**, 25, 137, 139, 142sq., 437–439, 469, 478, **500**, ~ centre 503
 non-professional ~ 503, personal ~ 497, private ~ 459, 506, Romance ~ 425, semi-oral ~ 498
 writing system, cuneiforms, ideographic ~, Phoenician ~, phonographic ~, syllabic ~ 200sq. cf. also alphabet(ic)
 written conception **139**sq., 143, ~ culture **401**, 404, 413, 423, 433, 439sq., 459, 478sq., ~ source 3, **137**, 411, 423, 430, 451, 516, 524, ~ variety 482

Y

Yanito 104
 Young Grammarians. cf. Neogrammarians
 youth culture/language 137, 143, 481

Z

zero argument 291, ~ article 262, ~ derivation 355, ~ indefinite plural 227, ~ marker/marking 43, 286, **295**, ~ morpheme 227, ~ pronoun 245

Name index

- Abeillé, Anne 305
 Accarisio, Alberto 203
 Acedo-Matellán, Víctor 279
 Adams, James Noel 408, 412
 Alberti, Leon Battista 462, 464
 Albrecht, Sabine 455
 Alcover, Antoni 393
 Aldrete, Bernardo de 22, 26, 30, 386, 464, 466
 Alembert le Rond, Jean d' 390, 465
 Alessio, Giovanni 392
 Alfonso the Wise 442sq., 446, 450
 Alonso Pedraz, Martín 393
 Alunno, Francesco 465
 Andreose, Alvisé 48sq., 68
 Antoine, Gérald 33, 399, 473
 Apollinaire, Sidoine 302
 Aprile, Marcello 392
 Apuleius 411
 Aquinas, Thomas 26
 Aquitaine, Eleanor of 433
 Argenter, Joan 85
 Aristotle 25, 259
 Armstrong, Nigel 194
 Arnauld, Antoine 27, 462, 466
 Aronoff, Mark 225, 227, 230, 233, 252
 Arveiller, Raymond 391
 Ascoli, Graziadio Isaia 33, 73, 80
 Ashdowne, Richard 247
 Auerbach, Erich 37
 Auger, Julie 296
 Ausonius 411
 Avila, Teresa de 491
 Bachmann, Iris 106
 Baglioni, Daniele 434, 502
 Baker, Craig 508
 Baldi, Benedetta 180
 Baldinger, Kurt 390–392, 398, 530
 Bally, Charles 38, 269
 Bal, Willy 51
 Balzac, Honoré de 506
 Bambini, Valentina 44
 Banniard, Michel 179, 413sq., 430
 Barbato, Marcello 29, 50, 67, 121, 217, 434
 Bartoli, Matteo 35, 121
 Basile, Giovanbattista 456
 Bastardas, Maria-Reina 50, 381
 Battaglia, Salvatore 392
 Battisti, Carlo 80, 194, 392
 Bauer, Laurie 349
 Baum, Richard 174
 Bayle, Pierre 465
 Bec, Pierre 49, 67–91
 Bédier, Joseph 33, 508, 510
 Belli, Guiseppe Gioacchino 456
 Bellini, Bernardo 34, 392
 Bello, Andrés 34
 Beltech, Eugen 194
 Beltrami, Pietro 50
 Bembo, Pietro 463, 466, 468, 497
 Bentley, Delia 279
 Berger, Günter 505
 Berkenbusch, Gabriele 464
 Bernand, Carmen 459
 Bernissan, Fabrice 69sq.
 Bernstein, Basil 41
 Bertinetto, Pier Marco 199, 268
 Bertolotti, Virginia 303
 Berzoli de Gubbio, Pietro 446
 Bickerton, Derek 133
 Bierbach, Mechthild 389, 445, 462, 465
 Biffi, Marco 194
 Billy, Pierre Henri 381
 Blanche Benveniste, Claire 141
 Blank, Andreas 321, 322, 326, 331, 335, 338, 345
 Blasco Ferrer, Eduardo 78
 Bloch, Oscar 391
 Bloomfield, Leonard 22, 39
 Blumenthal, Peter 276
 Boas, Charles 22, 39
 Boccaccio, Giovanni 315, 442, 450, 504sq., 512
 Boethius 26
 Boileau, Nicolas 461
 Bollée, Annegret 106
 Bolocan, Cauti Gheorghe 88
 Bonet, Eulàlia 180
 Bon, Jean le 465
 Bopp, Franz 30, 31, 178
 Bosque, Ignacio 306
 Bossong, Georg 49, 61, 67sq., 82, 89, 230, 254, 257sq., 262, 286, 289, 296, 298, 302, 445sq.
 Bossuat, Robert 488
 Boullón Agrelo, Ana Isabel 381
 Branca, Vittore 512
 Braselmann, Petra 349

- Bréal, Michel 34, 330
 Briquet, Charles Moïse 500
 Brucart Josep Maria 306
 Bruguera, Jordi 310, 349
 Brunel, Clovis 439
 Brunet, Étienne 314sq., 515
 Brunot, Ferdinand 37, 138, 398sq.
 Buchi, Éva 349, 382, 388, 391, 460
 Bühler, Karl 6, 149
 Burgio, Eugenio 50
 Buridant, Claude 305sq., 349, 514
 Busse, Winfried 276
 Bybee, Joan 266
 Cabré Castellví, Maria Teresa 349
 Cabredo Hofherr, Patricia 279
 Calvin, John 96, 460
 Canal, Paolo 44
 Cano Aguilar, Rafael 194
 Capet, Hugh 448
 Cappellaro, Chiara 225, 245
 Cappelli, Adriano 502
 Cardinaletti, Anna 229
 Carles, Hélène 36, 70, 126, 310, 317, 412, 426, 430, 452, 506
 Carlier, Anne 262
 Carlos III 453
 Carrilho, Ernestina 87
 Carstens, Henry 488
 Casalicchio, Jan 81
 Casares, Julio 328
 Casas Gómez, Miguel 329
 Castiglione, Baldassare 462
 Castro, Amerigo 438, 455
 Catullus 204
 Ceașescu, Nicolae 87, 533
 Céline, Louis-Ferdinand 290
 Cennamo, Michela 273
 Cerquiglini, Bernard 399
 Cervantes, Miguel de 497
 Chambon, Jean-Pierre 36, 72, 127, 171, 173, 179, 380, 385, 414, 428, 430, 435, 499, 512
 Charlemagne 443
 Chauveau, Jean-Paul 114, 117, 349, 452
 Chitoran, Ioana 194
 Chomsky, Noam 11, 22, 41, 109, 151, 157, 165, 224, 257, 292
 Chrétien de Troyes 172, 440, 442, 489, 504
 Christine de Pizan 505
 Christmann, Hans 346, 527
 Chrysippus 25
 Ciconte, Francesco Maria 279
 Cifoletti, Guido 434
 Cioranescu, Alexandru 393
 Confais, Jean-Paul 269
 Constantin 505
 Contini, Michel 117
 Corder, Pit 105
 Corneille, Pierre 315
 Corominas, Joan 124, 386, 393, 514
 Corr, Alice 260
 Cortelazzo, Manlio 392
 Coseriu, Eugenio 10–12, 22, 39, 55, 109, 124–132, 137, 152sq., 157, 316, 320, 366
 Cotgrave, Randle 389, 465
 Courtois, Michèle 427
 Covarrubias, Sebastián de 465
 Cresti, Emanuela 138
 Croft, William 257
 Cruschina, Silvio 287, 301
 Cuervo, Rufino José 34, 276, 393
 Cunha, Antônio Geraldo da 393
 Curell Aguilà, Clara 374
 Curtius, Ernst Robert 37
 D'Achille, Paolo 304
 Dahmen, Wolfgang 88, 418, 470, 502
 D'Alessandro, Roberta 296
 Da Milano, Federica 247
 Damourette, Jacques 36, 305
 Dante Alighieri 22, 26, 315, 446, 450, 506, 512
 Dardel, Robert de 409
 Darmesteter, Arsène 390
 Darms, Georges 470
 Darwin, Charles 32
 Dauzat, Albert 36, 385, 391
 de la Cruz, San Juan 491
 De Mauro, Tullio 473
 Demonte, Violeta 306, 514
 De Planta, Robert 394
 Descartes, René 460
 Detges, Ulrich 329
 Diamond, Jared 454
 Diderot, Denis 390, 465
 Diefenbach, Lorenz 31, 48
 Dietrich, Wolf 82
 Diez, Friedrich 31sq., 35sq., 178, 386, 391, 465, 487, 514, 527
 Di Giacomo, Salvatore 456
 Dik, Simon 22, 292
 Diori, Hamani 476
 Dittmann, Jürgen 168

- Donatus 26, 445
 Dragomirescu, Adina 241, 267
 Du Bellay, Joachim 461
 Dubois, Cyril 197
 Dubois, Jacques (Sylvius) 462
 Dubost, Pierre 276
 Dufter, Andreas 157
 Dülmen, Richard van 457
 Dumniștrăcel, Stelian 418
 Duval, Frédéric 493, 512
 Dworkin, Steven 310, 388
 Eberenz, Rolf 171
 Eckkrammer, Eva Martha 84
 Edmont, Edmond 35, 114, 117, 345
 Eichenhofer, Wolfgang 194
 Elcock, William 49
 Engler, Rudolf 38
 Ernst, Gerhard 46, 310, 442, 450, 459, 471
 Esher, Louise 71, 225, 256
 Esnault, Gaston 36
 Estienne, Henri 22, 462
 Estienne, Robert 26, 389, 462, 465sq.
 Étaples, Lefèvre d' 96, 460
 Eufe, Rembert 499
 Fabra, Pompeu 85, 481
 Fábregas, Antonio 355
 Fanciullo, Franco 77
 Faraoni, Vincenzo 237
 Faré, Paolo 35, 393
 Felixberger, Josef 420
 Ferguson, Charles 102, 132
 Filippo, Edoardo de 456
 Fillmore, Charles 22, 257, 292
 Finbow, Thomas 201
 Firth, John 39
 Fischer, Iancu 418
 Fishman, Joshua 101
 Fitch, William Tecumseh 165
 Fleischman, Suzanne 281
 Floricic, Franck 256, 349
 Flydal, Leiv 128sq., 137, 157
 Foerster, Wendelin 527
 Forner, Werner 145
 Fortin, Antonio 352
 Forza, Francesca 357
 Fradin, Bernard 224sq.
 Francard, Michel 126
 Frank, Barbara 201, 488
 Frei, Henri 36
 Frederick II of Hohenstaufen 433
 Fryba, Anne-Marguerite 446
 Furetière, Antoine 389, 465, 467
 Gabelentz, Georg von der 37, 45, 398
 Gamillscheg, Ernst 391, 419, 533
 García de Diego, Vicente 35, 393
 Gardani, Francesco 47, 105, 369
 Gargallo, Gil 50
 Garriga Escribano, Cecilio 147
 Gartner, Theodor 80
 Gasse-Grandjean, Marie-Josée 427
 Gauchat, Louis 394
 Gauger, Hans-Martin 50, 534
 Geckeler, Horst 82, 310, 349
 Geeraerts, Dirk 329
 Geisler, Hans 194
 Gelli, Giovan Battista 463
 Génicot, Léopold 488
 Georgescu, Simona 17, 338, 386
 Georgescu, Theodor 338
 Gerner, Dominique 442, 468, 479
 Gévaudan, Paul 269
 Giambullari, Pier Francesco 463
 Gili Gaya, Samuel 393
 Gilliéron, Jules 33, 35, 114sq., 117, 122, 345, 530
 Giolito, Gabriel 497
 Giolitto, Marco 531
 Giraud, Cedric 427
 Giurgea, Ion 301
 Giusti, Giuliana 229, 261
 Glessgen, Martin 36, 44, 46, 48, 70, 72, 108, 126, 153,
 157sq., 165, 225, 310–312, 322sq., 331, 349,
 364, 389, 392, 440, 443sq., 449, 452–454, 472,
 503, 506–509, 513, 520, 531
 Godard, Danièle 305
 Godefroy, Frédéric 390, 514
 Goebel, Hans 75, 117–120
 Goethe, Johann Wolfgang 468
 Goldoni, Carlo 456
 González Seoane, Ernesto 86
 Goosse, André 305
 Goscinny, René 290
 Gossen, Carl Theodor 444
 Gougenheim, Georges 378
 Gouvert, Xavier 381
 Greenberg, Joseph 288
 Grégoire, Abbé 479
 Grélois, Emmanuel 385
 Greub, Yann 72, 430, 499
 Grevisse, Maurice 305, 360
 Grice, Herbert Paul 41, 150

- Grimm, Jacob & Wilhelm 31sq., 178, 487
 Grimm, Reinhold 511
 Gröber, Gustav 34, 91, 488
 Gross, Gaston 276
 Grossmann, Maria 349, 365
 Grünert, Matthias 310
 Gruzinski, Serge 459
 Gsell, Otto 180
 Guéricos, Rosário Farâni Mansur 393
 Guillaume, Gustave 36, 165, 489
 Guillot-Barbance, Celine 512
 Gutenberg, Johannes 460, 500
 Haerle, Phillip 346
 Haller, Herrmann 74
 Halliday, Michael A.K 22, 39, 129, 157, 292
 Hallig, Rudolf 328, 391
 Harris, Martin 49
 Hartmann, Jörg 488
 Hatzfeld, Adolphe 34, 390
 Hauser, Marc 165
 Hausmann, Frank Rutger 533
 Heger, Klaus 93, 292, 320
 Heidinger, Steffen 279
 Heinimann, Sabine 81
 Heinimann, Siegfried 96
 Heisenberg, Werner 404
 Held, Gudrun 150
 Heraclitus 25
 Herman, Jozsef 412, 426
 Hernanz, María Lluïsa 306
 Hérouard 471
 Hickok, Gregory 166sq.
 Hillen, Wolfgang 51
 Hilmer, Hermann 17
 Hilty, Gerold 424
 Hinzelin, Marc-Olivier 234, 256
 Hjelmslev, Louis 11, 39, 319
 Hockett, Charles 22, 39
 Höfler, Manfred 374
 Hoinkes, Ulrich 101
 Holtus, Günter 46sq., 434
 Hreapcă, Doina 418
 Hubschmid, Johannes 530
 Hudson, Richard 111, 129, 156
 Huguet, Edmond 390
 Huizinga, Johan 437
 Humboldt, Wilhelm von 11, 17, 22, 29, 30
 Hummel, Martin 329
 Iliescu, Maria 260, 310
 Imbs, Paul 390
 Isidore of Seville 26, 386
 Jaberg, Karl 35, 116sq.
 Jackendoff, Ray 22
 Jacob, Daniel 101
 Jacobs, Bart 106
 Jakobson, Roman 38
 Jauss, Hans Robert 488
 Jones, Mari 134
 Jones, Sir William 30, 68, 178
 Joyce, James 506
 Jud, Jakob 35, 38, 116sq., 438
 Jungbluth, Konstanze 247
 Kabatek, Johannes 87, 103, 493, 521
 Kaehlbrandt, Roland 152
 Kafka, Franz 506
 Kay, Paul 22
 Keller, Rudi 20, 120
 Kemmler, Rolf 463
 Klapp, Otto 51
 Kleiber, Georges 329, 379
 Klein, Jean 388
 Kloss, Heinz 56, 57
 Koch, Peter 42, 48, 78, 129, 139–142, 158, 281, 322, 329, 425
 Köhler, Erich 488
 Kramer, Johannes 104, 346, 394, 409, 418, 429, 451
 Krefeld, Thomas 152
 Kristol, Andres 385
 Kuhn, Alwin 40
 Kürschner, Heike 488
 Labov, William 41
 Laca, Brenda 271, 303, 349
 Lachmann, Karl 510
 Lafage, Suzanne 453
 La Fauci, Nunzio 257, 261, 277sq.
 La Fontaine, Jean de 315
 Lamiroy, Béatrice 262
 Lancelot, Claude 27, 462, 466
 Lanher, Jean 113
 Larousse 390
 Lausberg, Heinrich 49
 Lausberg, Ulrich 49, 117, 206sq.
 Lavoie, Thomas 452
 Lazard, Gilbert 297
 Lebsanft, Franz 150, 153, 322, 447, 479sq.
 Ledgeway, Adam 47–49, 120, 225, 229, 231sq., 248, 260, 262, 267, 271, 273, 283, 287, 289, 305
 Leibniz, Gottfried Wilhelm 27
 Leonardi, Lino 50
 Le Roy Ladurie, Emmanuel 435, 437

- Levy, Emil 392
 Littré, Emile 34, 390
 Liver, Ricarda 260, 310, 349, 381
 Livescu, Michaela 260
 Llorach, Alarcos 39
 Lodge, R. Anthony 449, 457
 Lommatzsch, Erhard 33, 390
 Loporcaro, Michele 19, 47, 50, 120, 180, 199, 229,
 231, 238, 267, 279
 Louis the Pious 443
 Louis XIII 471
 Lubello, Sergio 77
 Lüdtke, Helmut 194, 281
 Lüdtke, Jens 85, 351
 Lupis, Antonio 338sq., 344, 389
 Luther, Martin 497
 Machado, José Pedro 393
 Machiavelli, Niccolò 463
 Magnus, Albertus 26
 Maiden, Martin 47, 49, 68, 209, 229, 232, 234–238,
 250, 252, 256, 260, 268, 270, 340
 Malherbe, François de 461
 Mangold, Max 181, 183, 192
 Manoliu, Maria 301
 Manuzio, Aldo 497
 Manzini, Maria Rita 114
 Manzoni, Alessandro 290, 506
 Maraschio, Nicoletta 194
 Marchello-Nizia, Christiane 155, 260, 305, 314
 Marco Datini, Francesco di 495
 Maria I 467
 Marie de France 440, 489
 Marín, Rafael 355
 Marotta, Giovanna 199
 Martin, Robert 390, 399
 Martineau, France 452
 Martinet, André 22, 38, 192
 Martínez González, Antonio 82
 Martins, Ana Maria 87
 Marzi, Claudia 44
 Marzo, Daniela 78
 Mateu, Jaume 279
 Matthews, Peter 225
 Maurer, Philippe 106
 Mayr, Ernst 16, 18
 Meigret, Louis 202sq., 462
 Meillet, Antoine 34
 Meinschaefer, Judith 199, 225
 Meisenburg, Trudel 194
 Melchior, Luca 81
 Melis, Federigo 495
 Ménage, Gilles 30, 386, 464sq.
 Menéndez Pidal, Ramón 37, 398
 Mensching, Guido 305, 446, 502
 Messner, Dieter 393
 Metzeltin, Michael 46, 532
 Meyer, Paul 33
 Meyer-Lübke, Wilhelm 35sq., 97, 306, 386, 391, 530
 Migne, Jacques Paul 427
 Minervini, Laura 50
 Miron, Paul 394
 Mistral, Frédéric 71, 392, 481
 Möhren, Frankwalt 390, 460
 Molière (Jean-Baptiste Poquelin) 27, 313, 315, 380
 Moll, Francesc de 393
 Moneglia, Massimo 138
 Mongin, Jean 115
 Monjour, Alf 310, 349
 Montaigne, Henri de 171, 512
 Morf, Heinrich 38, 437
 Morin, Yves Charles 194
 Morlet, Marie-Thérèse 385
 Morlicchio, Elda 392
 Morris, Charles 41, 165
 Moscati, Vincenzo 280
 Muhammad 416
 Müller, Bodo 92, 315, 393
 Munaro, Nicola 260
 Munteanu, Eugen 88
 Musil, Robert 506
 Mussafia, Adolf 33
 Muysken, Pier 106, 300
 Nebrija, Elio Antonio de 463, 464
 Nègre, Ernest 385
 Neruda, Pablo 290
 Nicolae, Alexandru 241, 267
 Nicot, Jean 465
 Nietzsche, Friedrich 32
 Noomen, Willem 510
 Nyrop, Kristoffer 36, 305
 Odoacer 413
 Oesterreicher, Wulf 48, 50, 129, 139, 140–142, 158,
 257, 263, 281, 459
 Ogden, Charles Kay 319
 Olivetan 460
 Olshki, Leonardo 38
 O'Neill, Paul 180, 225, 253
 Origo, Iris 495
 Orléans, Charles de 172
 Oudin, Antoine 465

- Oudin, César 465
 Palsgrave, Jehan 462
 Pană Dindelegan, Gabriela 267
 Păṇini 25
 Paris, Gaston 33
 Parry, Mair 120, 134, 248
 Pascual, José Antonio 393
 Pato, Enrique 231sq.
 Paul, Hermann 32sq., 178
 Paulikat, Frank 381
 Pei, Mario 49
 Peirce, Charles 8, 41
 Pellat, Jean-Christophe 305, 389, 445, 462
 Pellegrini, Giovan Battista 75sq., 394
 Pérez Saldanya, Manuel 260
 Perlmutter, David 277
 Perrault, Charles 505
 Perugi, Maurizio 445
 Pescarini, Diego 42, 245
 Petrarch, Francesco 442, 450, 463, 505
 Petrocchi, Giorgio 512
 Petrucci, Livio 425, 439
 Pfister, Max 117, 338sq., 344, 389, 392, 430, 434, 446, 510, 530
 Pharies, David 363
 Philip II Augustus 448
 Philippe de Mézières 505
 Pichon, Édouard 36, 305
 Picoche, Jacqueline 314
 Pillet, Alfred 488
 Pineda, Anna 279
 Pinker, Steven 314
 Pinto, Immacolata 349
 Pirrelli, Vito 44
 Plato 25
 Pöckl, Wolfgang 147, 176, 534
 Poeppel, David 166, 167
 Poirier, Claude 126
 Poletto, Cecilia 245, 280
 Pöll, Bernhard 176, 310, 349
 Polzin-Haumann, Claudia 70
 Pompidou, Georges 448
 Poplack, Shana 102
 Porta, Carlo 456
 Porta, Giuseppe 510
 Posner, Rebecca 49
 Pottier, Bernard 328
 Pountain, Christopher 68, 89, 145, 279
 Prati, Angelo 120
 Prifti, Elton 392
 Priscianus 26
 Probus, Valerius 26, 205
 Proust, Marcel 506
 Puşcariu, Sextil 394
 Pusch, Claus 138, 521
 Putzu, Ignazio 310
 Quadri, Bruno 120, 328
 Queffélec, Ambroise 310, 349
 Queneau, Raymond 290, 478
 Quint, Nicolas 106
 Rabelais, François 313, 315, 515
 Racine, Jean 315
 Radermacher, Ruth 390, 522
 Radtke, Edgar 150
 Rădulescu Sala, Marina 349
 Raible, Wolfgang 141, 147sq., 168, 201, 257sq., 286, 320, 321, 324, 488, 521
 Rainer, Franz 48, 152, 176, 238, 349, 352, 355, 356sq., 365, 384
 Ramat, Paolo 260
 Ramée (Pierre de la) cf. Ramus, Petrus
 Ramírez, Salvador 306
 Ramus, Petrus 460, 462
 Rask, Rasmus Kristian 31, 178
 Raynouard, François 32, 391, 487
 Recasens, Daniel 217
 Reclus, Onésime 476
 Regis, Riccardo 355
 Remacle, Louis 444
 Remberger, Eva-Maria 141, 301
 Remysen, Wim 452
 Renan, Ernest 32
 Renson, Jean 346
 Renzi, Lorenzo 48sq., 68, 260, 305sq., 514
 Repetti, Lori 217
 Reutner, Ursula 70
 Rey, Alain 391
 Rézeau, Pierre 126, 465, 497, 506
 Rheinfelder, Hans 346
 Ricca, Davide 260
 Richard I 432
 Richards, Ivor Armstrong 319
 Richelet, César Pierre 389, 465, 467
 Richelieu, Armand-Jean I. du Plessis 467
 Richter, Elise 179
 Ridruejo, Emilio 84
 Riegel, Martin 305
 Rieger, Angelica 507
 Ringger, Kurt 530
 Rio-Torto, Graça 349

- Rioul, René 305
 Roberts, Ian 245
 Roger II 433
 Rohlf, Gerhard 64, 65, 98–100, 114, 116, 236
 Romulus Augustulus 413
 Ronsard, Pierre de 497, 512
 Roques, Gilles 369
 Rossebastiano Bart, Alda 446
 Rostaing, Charles 385
 Rothwell, William 390
 Roudet, Léonce 330, 331
 Roulet, Eddy 105
 Rovere, Giovanni 276
 Ruhstaller, Stefan 381
 Sabatini, Francesco 428
 Sainéan, Lazare 36
 Saint Augustine 26, 408, 411
 Saint Jerome 408
 Saliceto, Guglielmo da 502
 Salviati, Francesco 463
 Salvi, Giampaolo 49, 290, 305
 Salvioni, Carlo 35, 80
 Sampson, Rodney 180
 Sánchez-Miret, Fernando 47
 Sapir, Edward 22, 39
 Saussure, Ferdinand de 9–13, 19, 22sq., 34, 38, 151, 157, 318sq., 530
 Savelsberg, Frank 446, 502
 Savj-Lopez, Paolo 426
 Savoia, Leonardo Maria 114, 180
 Scalise, Sergio 357
 Schaden, Gerhard 280, 283
 Schäfer-Priess, Barbara 463
 Schaller, Susan 43
 Scheuermeier, Paul 116, 117
 Schifano, Norma 287
 Schlegel, August Wilhelm von 31sq., 91, 233, 487
 Schlegel, Friedrich 31
 Schlemmer, Gerd 419
 Schlieben-Lange, Brigitte 453
 Schmid, Heinrich 79, 481
 Schmid, Stephan 180, 195
 Schmidt-Radefeldt, Jürgen 310, 349
 Schmitt, Christian 46, 221, 347
 Schøsler, Lene 108, 157, 158
 Schrödinger, Erwin 404
 Schuchardt, Hugo 36
 Schwarze, Christoph 305
 Schwarzwald, Ora 231sq.
 Schwegler, Armin 106
 Schweickard, Wolfgang 26, 46, 48, 70, 146, 310, 349, 369, 374, 380, 389, 392
 Searle, John 150
 Sechehayé, Albert 38
 Seidl, Christian 409, 412, 532
 Selig, Maria 140
 Selinker, Larry 105
 Senghor, Léopold 476
 Serianni, Luca 305, 399, 460
 Shakespeare, William 314
 Sheenan, Michelle 267
 Sibille, Jean 71
 Sidonius Apollinaris 411
 Siller-Runggaldier, Heidi 349
 Smith, Adam 20, 47, 225, 248, 271
 Smith, John Charles 134
 Söhrman, Ingmar 347
 Söll, Ludwig 139
 Sombart, Werner 457
 Sousa, Xulio 86
 Speroni, Sperone 461
 Spitzer, Leo 36, 37
 Sponde, Jean de 491
 Squartini, Mario 268
 Städtler, Thomas 390
 Stark, Elisabeth 141, 157, 305
 Stefenelli, Arnulf 347, 387, 388
 Stein, Achim 188, 260, 276
 Steiner, Linda 331
 Stempel, Wolf-Dieter 392
 Stewart, William 57
 Stilicho 413
 Stimm, Helmut 392
 Stone, Louise 390
 Straka, Georges 179
 Stussi, Alfredo 500
 Suchier, Hermann 449
 Sylvester I 505
 Tagliavini, Carlo 49, 65, 91, 420, 426, 460
 Tappolet, Ernst 35, 327, 346
 Terreaux, Louis 497
 Tesnière, Lucien 22, 38, 165, 257, 272, 275, 292, 296
 Theodosius 413
 Thibault, André 113, 126, 452
 Thomas, Antoine 33
 Thornton, Anna 234
 Thun, Harald 117, 125
 Tiktin, Hariton 394
 Tobler, Adolf 33, 390
 Tock, Benoît-Michel 427

- Togeby, Knud 39
 Tomescu, Domnița 381
 Tommaseo, Nicola 34, 392
 Tormes, Lazarillo de 289
 Torres Montes, Francisco 82
 Torres-Tamarit, Francesc 180
 Tortora, Christina 120, 245
 Tory, Geoffroy 462
 Trévoux 390
 Trier, Jost 38, 408
 Trifone, Pietro 399
 Trissino, Gian Giorgio 445, 462
 Trotter, David 443
 Trovato, Paolo 497
 Trubetskoy, Nikolai 38
 Trumper, John 137
 Tuaillon, Gaston 117
 Tuten, Donald 231, 232
 Ullmann, Stephen 330
 Vachon, Claire 171, 399
 Valdés, Juan de 464, 466
 Van Peteghem, Marleen 242
 Varchi, Benedetto 463
 Varvaro, Alberto 50
 Vasiliu, Laura 349
 Vaugelas, Claude Favre de 26sq., 461, 466–468
 Vendler, Zeno 278
 Viale, Matteo 463
 Videsott, Paul 81, 450
 Videsott, Ruth 81
 Vielliard, Françoise 488
 Vietti, Alessandro 194
 Villon, François 172
 Vincent, August 225, 229, 274, 385
 Vincent, Nigel 49, 271
 Völker, Harald 444, 453
 Voltaire 448
 Wagner, Max Leopold 116, 394
 Wall, Albert 87
 Wartburg, Walther von 36, 38, 58–62, 64, 67, 77sq.,
 87, 91, 96–100, 117, 120, 212, 215, 328, 343, 386,
 391, 419sq., 530, 533
 Weidenbusch, Waltraud 156
 Weinhold, Norbert 122
 Weinreich, Uriel 41, 129
 Weisgerber, Leo 38
 Wenker, Georg 116
 Wesch, Andreas 260
 Weyers, Christian 381
 Wheeler, Max 180
 Williams, Lynn 134
 William the Conqueror 432
 Windisch, Rudolf 50
 Winkelmann, Otto 117
 Winter-Froemel, Esme 329
 Wittgenstein, Ludwig 41, 258
 Wolf, Heinz-Jürgen 315, 378
 Wolf, Lothar 115, 122, 315
 Wüest, Jakob 38, 399
 Wunderli, Peter 530
 Zafiu, Rodica 145
 Zauner, Adolf 328, 346
 Zeno 25
 Ziltener, Werner 369
 Zinelli, Fabio 434, 512
 Zink, Gaston 179
 Zolli, Paolo 392
 Zribi-Hertz, Anne 106

Works cited

- Aalma* (glossary) 446
Aliscans 440
Amours (Ronsard) 497, 512
Appendix Probi 205, 235, 499
Augsburg Passion song 425
Auto de los reyes magos 490
Bible 96, 460, 470, 490sq., 497
Breviarium Alarici (Breviary of Alaric) 425
Cantilène de Sainte Eulalie (Canticle of Saint Eulalia) 172, 424
Cantico di Frate Sole 490
Catholicon 446
Champfleury (Tory) 462
Chanson de Guillaume 440
Chanson de Roland (Song of Roland) 172, 271, 289, 440, 489
Cligès 437, 440, 504
Corpus Iuris Civilis 492
Credo 490
Cronica de l'Anonimo Romano 510
Decameron 504sq.
De insigni obœdientia et fide uxoris (Petrarch) 505
Del origen y principio de la lengua castellana (Aldrete) 30, 464
De magistro (St. Augustine) 26
De vulgari eloquentia (Dante) 26, 445sq., 464
Dialogo delle lingue (Speroni) 461
Diálogos intorno a la lengua (Valdés) 464
Diccionario de Autoridades 393, 467
Diccionario (Nebrija) 464
Dictionaire françois-latin (Robert Estienne) 389
Dictionarie of the French and English tongues (Cotgrave) 389, 465
Dictionnaire de l'Académie française 389, 461, 465, 467
Dictionnaire françois, contenant les mots et les choses (Richelet) 389
Dictionnaire françois-latin (Robert Estienne) 26, 465
Dictionnaire italien-françois et françois-italien (Oudin) 465
Dictionnaire universel françois et latin (Trévoux) 390, 465
Dictionnaire universel (Furetière) 389
Divina Commedia (Divine Comedy) 506
Donats 445
Don Quixote 497
Encyclopédie française (Diderot/d'Alembert) 390, 465
Erec 440
Essai sur la Précellence du langage françois (Henri Estienne) 461
Essais (Montaigne) 171, 512
Etymologiae (Isidore of Seville) 26, 386
Etymologicon françois (Le Bon) 465
Fierabras 440
Formula di confessione umbra 491
Glosas Aemilianenses 425, 445
Glosas Silenses 425, 445
Glossario di Monza (Glossary of Monza) 424sq.
Glosses of Kassel 424, 445
Glosses of Reichenau 424, 445
Gramática de la lengua castellana (Nebrija) 463
Gramere (Pierre de la Ramée) 462
Grammaire générale et raisonnée (Arnauld/Lancelot) 27, 305, 462
Griseldis (Tale of Griselda) 505, 511
Hypomneses de Gallica lingua (Henri Estienne) 462
Indovinello veronese 424sq.
In linguam gallicam isagoge (Dubois) 462
La deffence et illustration de la langue françoise (Du Bellay) 461
La fabrica del mondo (Alunno) 465
Lai de l'Ombre 33, 508, 510
La Marquise de Salusses, ou La Patience de Griseldis (Perrault) 505
Lancelot 440
Laudes Creaturum 490
Laudes regiae 424
Legenda Aurea 490, 506
Leis Guillelmi 493
Lesclaircissement de la langue françoise (Palsgrave) 462
Lex Salica 424
Life of Saint Alexis 172, 491
Life of Saint Léger 491
Livre de la cité des Dames 505
Memoriale Bolognese 499
Ménagier de Paris (de Mézières) 505
Miroir des dames mariées (de Mezières) 505
Moamin 508, 513
Nodicia de kesos 425
Origines de la langue françoise (Ménage) 30, 465
Ortografia de la Real Academia española 203
Ovide moralisé 172
Pater noster 490

- Perceval* 440
Prose della volgar lingua (Bembo) 463, 497
Remarques sur la langue française (Vaugelas) 26, 461, 468
Renart le contrefait 172
Roman d'Alexandre, roman d'Énéas, roman de Renart, roman de Thebes, roman de Troie 440
Roman de la Rose (Romance of the Rose) 172
 Rule of Saint Benedict, ~ of the Cistercians 491
Séquence de Sainte Eulalie (Sequence of Saint Eulalia) 172, 424sq., 501
Serments de Strasbourg (Strasbourg Oaths) 172, 289, 424sq.
Sermoni Subalpini 490
Sermon sur Jonas (Sermon on Jonah) 424sq., 429
Tesoro de la lengua castellana o española (Covarrubias) 465
Thresor de la Langue Françoyse (Nicot) 465
Traicte de la grammaire française (Robert Estienne) 462
Traité d'anthroponymie française (Dauzat) 385
Traité touchant le commun usage de l'écriture Française (Meigret) 203
Trésor des deux langues espagnole et française (Oudin/Oudin) 465
Tręttę de la gramme française (Meigret) 462
Tristan and Iseult 504
Vocabolario della Crusca 465, 467
Vocabolario, Grammatica, et ortographia de la lingua volgare... (Accarisio) 203
Vocabolario universale italiano (Tramater) 392
Voyage de Charlemagne 440
Voyage de Saint Brendan 491
Vulgate 460
Yvain 440