

Beauty and the Blend

Implications of Cognitive Constraints and Word
Class Distribution in Lexical Blending

Daniel Kjellander

Department of English

Master Degree Project
English Linguistics

Spring 2015

Supervisor: Francesco Alessio-Ursini



Stockholms
universitet

Beauty and the Blend

Implications of Cognitive Constraints and Word Class Distribution in Lexical Blending

Daniel Kjellander

Abstract

This master thesis investigates underlying principles and usage-based aspects of lexical blending. In a corpus study examining the lexical items, or matrix words, *republican*, *liberal*, and *vegetarian* it was found that there were three cognitive constraints influencing their potential to form blends. Mapping of a prefixation schema onto the lexical item *republican* was shown to fuel blend formation. Neighborhood effects and morphological lexicalization, on the other hand, were observed to have a negative influence on the possibility to form blends from the terms *liberal* and *vegetarian*. Also, an examination of word class distribution in the matrix words and their blend and compound derivatives showed that the patterns of the blends were less consistent than the compounds. These findings point to an important duality in the derivative blends. While they operate in an interplay with regular morphological processes they are at the same time creatively elaborated in use. Therefore blends exhibit a fundamental aspect of language, which is phrased as the dynamic interrelatedness between socio-pragmatic motivation and schematization, or pattern-finding. This positions blending not in the margin, but at the centre of studies on language development.

Keywords

Blending, wordplay, word formation, morphology.

Contents

1. Introduction and research question	1
2. Theoretical framework	3
2.1 Overview of the field of research	3
2.1.1 Creativity and novelty	4
2.1.2 Wordplay	5
2.1.3 Fusion of form and designata	6
2.1.4 Syntagmatic and paradigmatic blends	7
2.1.5 Decoding	8
2.2 Blending and Morphology	9
2.2.1 Blending and other types of word formation	9
2.2.2. Implications of psycholinguistic processing	12
3. Method	13
4. Results	15
4.1 The data collection process	15
4.2 Structural analysis	18
4.2.1 Patterns of the matrix word "republican"	18
4.2.2 Patterns of the matrix word "liberal"	20
4.2.3 Patterns of the matrix word "vegetarian"	21
4.3 Analysis of word class distribution	22
4.3.1 Word class distribution of "republican"	22
4.3.2 Word class distribution of "liberal"	24
4.3.3 Word class distribution of "vegetarian"	25
4.4 Summary of results	26
4.4.1 Structure	26
4.4.2 Word class distribution	27
5. Discussion	27
5.1 Cognitive constraints	28
5.1.1 Positive constraints	28
5.1.2 Negative constraints	29
5.2 Implications of patterns of word class distribution	31
5.3 Summary of discussion	33

6. Conclusion	35
References.....	36
Appendix A	39
Appendix B	40

1. Introduction and research question

Judging from the literature, few areas of research seem to have troubled the taxonomical inclination of morphologists as deeply as the word formation process of blending (cf. Bauer, 2006). To make matters worse, it seems to enjoy a solid, and possibly increasing, popularity among users of English (e.g. Bryant, 1974; Cannon, 1986; Lehrer, 2007). Although researchers are in disagreement as to theoretical explanations, a common denominator of most accounts is its well attested complexity as regards predictability and structural constraints (e.g. Bauer, 1983; Cannon, 1986; Kelly, 1998; López Rúa, 2004; Gries, 2006). Gries (2006) describes blending as “one of the most creative word-formation processes - where creative is used in the sense of ‘defying characterization by means of hard-and-fast productive rules’” (p. 535). In addition, Cannon (1986) asks whether blends are “so slippery, ill-defined, and close to other word-forming categories that scholars are hesitant to describe them” (p. 736), which is mirrored in Kemmer’s (2003) claim that they have been “little more than a footnote in the study of morphology” (p.76).

Attempts at explaining blend formation, especially until the 1980s, have typically focused on classification and categorization (Cannon, 1986; Gries, 2006; Fandrych, 2008). These types of explanation form an important structural background, but they have often met considerable difficulties as regards taxonomy and terminology (e.g. Cannon, 1986; Bauer, 2006). However, in efforts to challenge the complexity of blends, a number of studies have shown that their structure is not really as elusive as it may seem (Lehrer, 1996; Kelly, 1998; Kemmer 2003; Plag, 2003; López Rúa, 2004; Gries, 2004, 2006; Fandrych, 2008; Beliaeva, 2014). The use of computer-based methodologies (e.g. Gries, 2004, 2006) in combination with broader interpretive frames including semantic (e.g. Lehrer, 1996), pragmatic (e.g. Fandrych, 2008) and cognitive perspectives (e.g. Kelly, 1998; Kemmer, 2003; Gries, 2006; Beliaeva, 2014) offer new possibilities to describe blending in more comprehensive and apt ways. In other words, although structural and morphological perspectives play an important part, semantic and pragmatic analyses are central in the development of this field of research.

One of the distinctive characters of blending is found in precisely this respect; although it is fully possible to attempt a taxonomic approach based on structure (e.g. Lopez Rúa, 2004), blends seem to have unique properties transcending the structural rules of word formation¹ (Fandrych, 2008). Therefore, it might be tempting to detach blending from morphology on the grounds that its functions and manifestations are too morphologically divergent (Dressler, 2000; Kemmer 2003). Doing this leads however to explanatory difficulties as regards borderline phenomena. For instance, the commonly occurring process of morphological lexicalization, instantiated in examples such as *burger* and *-(o)holic*, typically blurs taxonomical delineations (Bauer, 1983; Lehrer, 2007; Schmid, 2011; Beliaeva, 2014).

¹ “[T]he numerous patterns that they exhibit are too diverse to be generated within the traditional framework of generative rules, which must be either recast or else abandoned in favor of a more powerful kind of device that can accommodate these often unpredictable items.” (Cannon, 1986, p. 748)

Instead, an inclusive approach combining morphological aspects and blend specific patterns is needed to account for the complexity of blend formation. It follows from this that neither perspective can be excluded, but also that findings in blend research have the potential to contribute to morphological research in general (Dressler, 2000).

In research on blending, a common methodological procedure is to choose a particular set of blends, and then study various significant aspects of their nature (e.g. Cannon, 1986; Gries 2004; Fandrych, 2008; Beliaeva, 2014 to name a few). However, since blends emerge as a result of combining lexical items in intricate and type-specific ways, their lexical origins are also potential resources of information.

Therefore, this particular study aims at investigating the relation between three lexical items, or *matrix words*, and a set of blends derived from them. For reasons of space, the primary focus is derivatives headed by one of the matrix words. More specifically, it will be studied how the structure of these matrix words affect the potential for blend formation, and how other word formation processes in the data are related to the blends. This approach enables an analysis not only of the blends *per se*, but also of the linguistic environment in which they are manifested. Since there are indications that traditional morphology alone cannot fully account for blending (e.g. Cannon, 1986; Dressler, 2000), understanding such underlying strategies are important in the development towards an increased knowledge of this phenomenon.

Also, the chosen matrix words, *republican*, *liberal*, and *vegetarian*, occur both as adjectives and nouns, which means that it is possible to investigate their word class distribution. Assuming that this grammatical property is shared by their derivatives, a comparative analysis may reveal important aspects of usage-patterns of blends as well as other word formation types in the data. Therefore, a corpus analysis will be carried out with the object to find out whether the investigated items differ in this respect. Special attention will be given the blends in the data, but patterns in other possible word formations will also constitute an important background.

The database will be built on the corpus output of truncated query formulas such as **lican* rendering for instance *antirepublican*, *anglican*, and *teapublican*. Subsequently, instances of both regular morphological operations and creative formations can be analyzed and compared. Due to the attested structural diversity of blends (e.g. Cannon, 1986; Gries, 2004, 2006), the investigation does not claim that the corpus output will represent all possible derivatives of the matrix words. However, because of English headedness rules (Bauer, 1983, 2006) and the preponderant structure of final and initial omission of source word parts respectively (Cannon, 1986), this procedure has the capacity to produce significant data for the purposes of this investigation.

The thesis is organized in four main sections. Section 2 outlines the theoretical background including established assumptions and recent research. Methodological issues are discussed in section 3, and the results of the data collection are presented in section 4. Finally, section 5 discusses the empirical findings and account for theoretical implications of the data.

2. Theoretical framework

In order to establish the theoretical basis for the investigation, the following sections will outline main tendencies as regards established assumptions and recent findings in the field. Section 2.1 is concerned with a few central characteristics of blending, and a number of examples from the body of research will be presented and discussed in each of the five subsections. These presentations are neither exhaustive nor easily delimited, but they represent important tendencies necessary to describe the nature of blending.

In section 2.2 a brief discussion will relate blending to morphological theory. This is important mainly for two reasons. Primarily, the fact that blends deviate from regular morphological processes raises issues as to their status in morphological frameworks (Dressler, 2000; Kemmer, 2003). Nonetheless, empirical data, which will be presented below, reveals examples that testify to the close relation to the field of morphology. This aspect needs therefore to be addressed in order to account for the data. Secondly, due to the somewhat incongruous character of blends, linguists have taken on broad theoretical perspectives to explain empirical data. This means that novel approaches have added valuable knowledge that may have implications for morphological theory in general, if given proper attention (Dressler, 2000; Fandrych, 2008).

As regards terminology, the complex structure of blends makes it difficult to apply traditional morphological terminology to blending. Stable and productive bound morphemes such as *-tion* and *-s* are readily analyzable units within a morphological framework, whereas blending typically do not display such regular forms (e.g. Dressler, 2000; Bauer, 2006). Some researchers (e.g. Lehrer, 1996; Ronneberger-Sibold, 2006; Fandrych, 2008; Beliaeva, 2014) use therefore the metaphorical term *splinter* to denote structural parts of a blend's lexical origins. In the present investigation the term *segment* is however preferred due to its more figuratively neutral quality. Furthermore, the well established term *source word* is chosen to denote the lexical items from which a blend originates. These will be bracketed, italicized and combined with the symbol "+", rendering forms such as *smog* (*smoke* + *fog*). There are more detailed systems showing, for instance, which part of a source word is omitted, e.g. *smog* (< *smoke*+*fog*) (Bauer, 2006). For the purposes of the present study it is however considered sufficient to signal the lexical origin.

2.1 Overview of the field of research

Recent investigations of blending illustrate some important tendencies in contemporary research in the field of blend formation. Firstly, they demonstrate the seminal importance of access to powerful digital tools used in the collection and organization of data (e.g. Kelly, 1998; Gries, 2006). The use of electronic corpora is perhaps the most notable technological contribution (McEnery & Gabrielatos, 2006), and computer driven analysis and organization of data offer a wide range of methodological possibilities. For instance, in studies such as Kelly (1998) and Gries (2006) the digital technology enables the researcher to process historically unprecedented amounts of data, which enhances both accuracy and efficiency of the empirical material.

Secondly, doubts as to the viability of traditional morphological analyses of blend structure (e.g. Cannon, 1986; Kemmer 2003; Fandrych, 2008) have led to the emergence of new theoretical models that matches recent empirical findings in a better way. For instance, terms and notions such as *source word similarity* (Kelly, 1998), *recognition points* (Gries, 2006), and *extragrammatical morphological operation* (EMO) (Dressler, 2000) have been applied to serve as conceptual categories intended to describe specific properties of blends. On a more methodological level the deployment of schema-based approaches (Kemmer, 2003) and the application of socio-pragmatic models (Fandrych, 2008) also exemplify a theoretical expansion in relation to traditional morphology.

Thirdly, the recent research efforts mentioned above have resulted in findings that challenge previous classificatory attempts. In certain areas this has occasioned further disagreement, especially in taxonomic matters. For instance, Plag (2003) and Lehrer (2007) include *complex clippings* (e.g. *digicam* from *digital camera*) in the category of blends, whereas Beliaeva (2014) presents evidence that there should be a categorial distinction between blends and complex clippings. As regards blend types, Algeo (1977) and Dressler (2000) exclude *syntagmatic blends* (see section 2.1.4) from the category of blends proper, seeing them instead as contracted compounds. In contrast, Bauer (2006), Beliaeva (2014), and Lehrer (2007) include this type in their accounts on blending. The latter presents evidence showing that certain syntagmatic blends inhibit a compound-typical analysis such as *a kitchen towel is a kind of towel*. Instead, in a syntagmatic blend such as *meatitarian* subjects rejected the idea that *a meatitarian is a kind of vegetarian* (ibid.)

Finally a brief note will be made on the relation between blends and compounds. It is often assumed that blending is a subtype of compounding (e.g. Quirk et al., 1985). Acknowledging the complexity of this matter, the stance of the present investigation is however that blending is best understood as a discrete category. The main reasons for this are found in examining recent usage-based accounts. Not only are blends structurally different from compounds, but there is also strong evidence that semantic and pragmatic characteristics are equally important discriminating factors (e.g. Bauer, 2006; Ronneberger-Sibold, 2006; Beliaeva, 2014). This is not to say that the distinction is a straightforward one. The category boundaries are indeed fuzzy² as “classification seems to be forced upon us by the boundaries provided by reality” (Ungerer & Schmid, 2006, p. 8). The relation between blends and compounds will be discussed in more detail in section 2.2.1.

2.1.1 Creativity and novelty

Together with a range of other phenomena, blends seem generally to be associated with linguistic novelty and creativity in the English language (e.g. Algeo, 1977; Cannon, 1986; McMahon, 1994; Gries, 2006; Alm-Arvius, 2012). Some authors (e.g. Bryant,

² Ungerer & Schmid (2006) uses the term *fuzzy* to denote vague boundaries of *categories* (e.g. cup; see Labov, 1973), whereas unclear delineation of *entities* are named *vague* (e.g. mountain).

1974; Lehrer, 2007) even put forth the idea that an intensified development in society drives an increase in blend formation.

It is noteworthy that word formation processes such as compounding and abbreviation (including acronyms, initialisms and complex clippings) typically share this feature (cf. McMahon, 1994; Bauer 2006). For instance, Cannon (1989) claims that a major increase in the use of abbreviations took place at the time of World War II, although it is also known that this phenomenon dates back to Sumerian culture (ibid.). Furthermore, Cannon (1989) suggests that “they are moving ever more easily and quickly into general English” (p. 119). In this respect, there is thus a kinship between blending and processes of abbreviation.

As regards compounding, there seems however to be important differences. Although creativity and novelty is a recurring theme in the category of compounds, Bauer (2006) points to an underlying distinction in terms of processing. While compounds are described as instances of rule-governed *productivity*, blending and other “minor word formation types” (ibid.) are claimed to be instances of *creativity*. In this context *creativity* implies “the predominance of analogy and other processes which are not rule-governed” (ibid. p. 483).

This raises a question as to definitions; Kelly (1998) and Gries (2004, 2006) have shown patterns in blend formation that may be interpreted as rules, although these are different in nature compared to traditional morphological ones (cf. Cannon, 1986). For instance, Kelly (1998) shows that there is a strong phonological tendency towards “onset-rime over body-coda breaks” in blends (p. 586).

2.1.2 Wordplay

The phenomenon of wordplay is perhaps at the core of explanations of blending (e.g. Quirk et al., 1985; Fandrych, 2008; Schmid, 2011; Alm-Arvius, 2012; Fábregas & Scalise, 2012). The inherent punning qualities, or “witticisms” (Algeo, 1977), in wordplay are described in psycholinguistically oriented terms in Kelly (1998):

By constructing the blend so that the onset of word two sounds similar to the expected continuation of word one, the speaker postpones, however momentarily, the listeners' recognition that they have been sidetracked. The resulting blend might also sound smoother and, perhaps, more apt in meaning (Kelly, 1998, p. 587).

Kelly (1998) thus touches upon central motivations for this linguistic function, which is mirrored in Alm-Arvius' (2012) emphasis on Roman Jakobson's *poetic function* in various forms of wordplay.

In blending as well as in figurative language, idiom breaking, punning, and poetry, the poetic function is manifested as meaning that resides inside a language system (Jakobson, 1966; Alm-Arvius, 2012). Commonly, denotation is described as the relation between a “lexical sense [...] and the category of things out in the world that it represents” (Alm-Arvius, 2012, p. 9). In contrast, instances of poetic meaning refer to knowledge about the language system itself. In the recurring example *foolosophy* this is instantiated in a combination of phonological, orthographic and denotational

playfulness (cf. Fábregas & Scalise, 2012). Although perhaps not a necessary condition, the effect of wordplay, or poetic meaning, is nevertheless commonly considered an intrinsic aspect in accounts of blending (Ronneberger-Sibold, 2006). From this perspective, blends seem thus to be related to types of expression such as punning, tropes and rhetoric language. Therefore, the significance of wordplay in blending supports the idea that structurally oriented explanations cannot fully account for this word formation process.

2.1.3 Fusion of form and designata

The well-established assumption that blending implies the fusion of two (or more) designata (e.g. Schmid, 2011; Alm-Arvius, 2012) is exemplified in the recurring citation of Lewis Carroll's *Through the Looking Glass*³ (e.g. Bryant, 1974; Kemmer, 2003; Bauer, 2006). Carroll's term *portmanteau word* is sometimes even used synonymously to *blend* (e.g. Alm-Arvius, 2012), thus indicating metaphorically the inclusion of two specific meanings, or senses, into one linguistic sign⁴. It should be emphasized that the significance of this phenomenon in blending lies in its combinatory effect of simultaneous fusion of form and designata (López Rúa, 2004).

The fusion of form is commonly one of the main criteria in definitions of blends (e.g. Bauer, 1983; Cannon, 1986; Gries, 2004; Lehrer, 2007 to name a few). This implies that material from the lexical structure of the source words is omitted. Also, it is not uncommon for blends to share structural material. For instance, in *stoption* (*stop* + *option*) the letter combination *op* in both lexical sources coincides in the resulting blend. The established terms for these processes are *clipping* and *overlap* respectively.

There is widespread agreement as to a prototypical model of clipping, in which a final segment of the first source word and an initial segment of the second source word are omitted (e.g. Cannon, 1986; Bauer, 2006). There are, however, differing views on less prototypical types of clipping. For instance, the fact that Plag (2003) sees complex clippings as a type of blend entails that final clipping of both source words (which is a typical property of complex clippings) is considered a valid criterion for blends. In this respect, the stance of the present investigation is however based on Beliaeva's (2014) findings showing that there are differences in terms of usage between blends and complex clippings. Thus, forms displaying final clipping of both source words will not be included in this investigation.

Furthermore, overlap of the type *chunnel* (*channel* + *tunnel*) is sometimes considered a core feature of blends (e.g. Algeo, 1977). In more recent accounts, however, it is common to see this type of structural fusion as one out of a number of forms subsumed under the category of blend formation (e.g. Bauer, 2006; Gries, 2006; Lehrer, 2007; Beliaeva, 2014). Nevertheless, because of the close relation between overlap and

³ "You see it's like a portmanteau - there are two meanings packed up in one word" (Carroll, 1993, p. 207).

⁴ Algeo's (1977) suggestion that *portmanteau word* should denote only paradigmatic blends with overlap indicates that this term is less inclusive than *blend*. Therefore, *portmanteau word* is to some extent a hyponym rather than a synonym of the more inclusive term *blend* (cf. Beliaeva, 2014).

wordplay (see section 2.1.2) there are strong reasons to stress its importance in blending.

Research on the conceptual aspect of fusion ranges over a wide array of topics (for an overview see Geeraerts, 2006 and Ungerer & Schmid, 2006). In the context of the present investigation, Fauconnier & Turner's (2002) ideas stand out as particularly striking. Their term *conceptual blending* may be slightly confusing in that it includes all sorts of conceptual integration in a wide range of linguistic phenomena (Taylor, 2002). In contrast, the common understanding of lexical blending is restricted to the specific type of word formation process that is the topic of the present study. Moreover, accounts of lexical blending given before the 1990s have typically been focused on structural rather than conceptual matters (Gries, 2004, 2006).

In recent years attempts towards a coalescence of structural aspects and conceptually oriented perspectives have been made in studies such as Kelly (1996), Lehrer (1996), Kemmer (2003), Fandrych (2008), and Beliaeva (2014). Although their approaches differ as to methods and objectives, there seems to be agreement that a viable account of blending requires both a structural and a conceptual perspective (López Rúa, 2004; Ronneberger-Sibold, 2006).

2.1.4 Syntagmatic and paradigmatic blends

A distinction is often made between *paradigmatic* and *syntagmatic* blends (Cannon, 1986; Bauer, 2006; Beliaeva, 2014). Although there is some variation as to definitions and terminology, a fairly established understanding of these categories could be described in terms of a few characteristics.

Blends formed from contracted syntactic patterns are usually referred to as *syntagmatic* (Algeo, 1977; Dressler, 2000; Bauer, 2006). The perhaps most obvious criterion for this type of blend is its word order restriction. As is the case in compounding, the meaning of a syntagmatic blend such as *motel*⁵ (*motor* + *hotel*) depends on the understanding of *hotel* as its semantic head. It is therefore not possible to retain the semantic content if the order of the segments is changed to a hypothetical blend *hotor*. Such a blend would typically denote a type of *motor* rather than a type of *hotel* (example and rationale from Bauer, 2006). Thus, since syntagmatic blends are syntactically determined, the source word order also determines the order of the elements.

Furthermore, Algeo (1977) claims that syntagmatic blends lack a semantic relatedness seen in associative, or paradigmatic, blends of the portmanteau type suggested by Lewis Carroll. Therefore, the term *telescope word* is suggested to denote blends “formed by conflating two juxtaposed forms, rather like sliding the cylindrical parts of a telescope together” (p. 57).

As mentioned in section 2.1, the understanding of syntagmatic blends as mere contractions has been questioned by several researchers (e.g. Kemmer, 2003; Lehrer, 2007; Fandrych, 2008). In Lehrer's (2007) investigation a hyponymy test revealed that

⁵ Notably, the blend *motel* and the compound *motor hotel* appeared simultaneously in written form in 1925 (Barnhart & Steinmetz, 1988), which in this case hints at the close structural relation between the forms.

the subjects did not always accept a blend as a semantic subcategory of its original syntactic construction, or compound. They rejected, for instance, the statement “a skyscape is a kind of landscape” (p. 126). This is explained as a result of semantic transfer from the free morpheme *sky* to the blend *skyscape*, which produces a conflict of meaning between the blend and its assumed superordinate category (i.e. *landscape*).

While *syntagmatic blends* is a relatively coherent category, there is both terminological and descriptive variation as regards *paradigmatic blends*⁶ (Bauer, 2006). A preliminary definition of this category will however be phrased here as blends whose source words are in *paradigmatic relation* (ibid.) as in *geep* originating from the conjunctive phrase *goat and sheep* (Kelly, 1998; Beliaeva, 2014). Because of the absence of word order restriction, paradigmatic blends seem to exhibit a high degree of structural complexity and relatively low semantic transparency (e.g. Cannon, 1986). For instance, Kelly (1998) addresses this issue showing that the source words’ frequency, syllabic structure, and cognitive prominence influence the structure of *conjunctive*, or paradigmatic, blends.

The significance of paradigmatic blends is illustrated in the fact that some researchers consider the paradigmatic relation to be a more or less necessary condition in definitions of blend formation (e.g. Algeo, 1977; Cannon, 1986; Dressler, 2000). As touched upon above, however, there is disagreement on this topic. Defining criteria too narrowly concerning what type actually qualifies as a blend implies the risk of excluding data that in many senses have significant blend characteristics (cf. Lehrer, 2007).

2.1.5 Decoding

The successful comprehension of a blend obviously requires a reader or hearer to have access to adequate decoding strategies. Since blending commonly involves structures requiring active processing, semantic transparency is a crucial parameter (Ronneberger-Sibold, 2006). This can be rephrased as the possibility of identifying the correct source words (cf. Bauer, 1983). Compared to compounds, blends often display a relatively low level of transparency (Ronneberger-Sibold, 2006), which means that a certain amount of prior understanding is necessary. Besides factors such as lexical or encyclopaedic knowledge, context is a crucial factor in the comprehension of blends (Algeo, 1977; Lehrer, 1996, 2007; Ronneberger-Sibold, 2006). From this follows that a wide range of aspects come into play. For instance, speech community associations (Cannon, 1986) and sociocultural variables (Fandrych, 2008) also influence comprehension.

The tendency for certain blends, especially paradigmatic ones, to be relatively opaque also implies that they sometimes function as more or less arbitrary linguistic signs (Fandrych, 2008). Their status in terms of lexicalization is therefore important when analyzing decoding strategies (cf. Beliaeva, 2014) since the transparency of blends is highly gradual (Lehrer, 1996; Ronneberger-Sibold, 2006).

Furthermore, the significant preference for wordplay also underlines the importance of lexical status in the comprehension of blends. The fact that wordplay relates to

⁶ For instance, *conjunctive* and *associative* are sometimes used instead of *paradigmatic*.

intralinguistic meaning (and thus relies on previous lexical knowledge) can be seen in a blend such as *wedbsite* (*wedding* + *website*). A speaker of English with encyclopaedic knowledge about *wedding websites*⁷ typically has enough previous understanding to decode this blend without any additional clarification in context.

The decoding of blends can thus be described as an interplay between semantic transparency, lexicalization processes, contextual cues and pragmatic aspects. Indeed, Gries' (2006) depiction of blending quoted in the introduction does not seem to be an exaggeration as regards decoding. Although recent research has resolved many other issues in the field, much work is yet to be done as to comprehension aspects of blending.

2.2 Blending and Morphology

Due to the uncertain position of blending in morphological theory (cf. Dressler, 2000; Kemmer, 2003; Gries, 2006) the following sections will outline central theoretical issues necessary for the investigation. It will be shown that adding a processing perspective is necessary in order to explain the specific character of this phenomenon. This is supported by recent research illustrating the salience of cognitive and semantic aspects of blend formation (e.g. Lehrer, 1996, 2007; Kelly, 1998; Kemmer, 2003; Gries, 2004, 2006; Beliaeva, 2014).

2.2.1 Blending and other types of word formation

Given that blends are a type of word formation process, three complementing perspectives are chosen to encapsulate core aspects of their nature in relation to other word formation categories; *semantic dynamism*, *complex encoding* and *functional structure*. These concepts, and their respective empirical underpinnings, will be outlined below. It will be shown that they summarize recent research on a theoretical level, and that they contribute to categorial delimitations in this field of research.

Firstly, Lehrer (1996) shows the importance of semantic priming effects in the comprehension of novel blends claiming that the informants “were usually mystified until [they were given] the semantic domain” (p. 376). The study thus highlights the idea that *cognitive domains* (cf. Langacker, 2008; see also Ronneberger-Sibold, 2006), or semantic networks, play an important role in the lexical decoding of blends. The reliance on priming effects can be related to the significance of context and lexical status in that blends typically deploy a dynamic interplay between these decoding strategies (cf. Ronneberger-Sibold, 2006). In other words, the less information available in lexicalized knowledge and identification of source words, the more blends tend to depend on contextual resources such as semantic priming (Lehrer, 1996). This dynamic interplay is illustrated in figure 1. *Lexicalized knowledge* is distinguished from *identification of source words* in that the former does not rely on the transparency of the segments, although an etymological analysis may be possible (e.g. *smog*).

⁷ Note that the compound *wedding websites* is semantically transparent as opposed to the syntagmatic blend formed from it. See Ronneberger-Sibold (2006) for an in-depth discussion on the reasons for this difference.

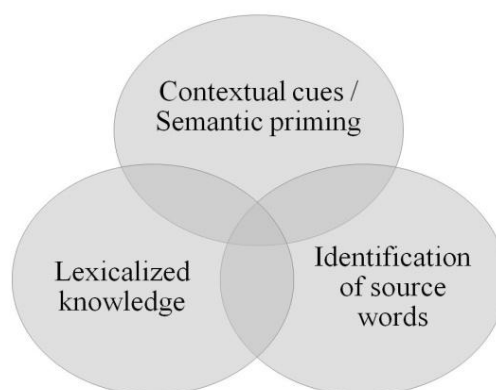


Figure 1. Interplay between decoding strategies

This property is often socio-pragmatically exploited in blends (Kelly, 1998; Ronneberger-Sibold, 2006; Fandrych, 2008). Fandrych (2008) exemplifies with the creation of “cataphoric suspense” observed in newspaper headlines (p. 76), and Kelly (1998) stresses the notion of “lexical tease” (p. 586) achieved by structural overlap and phonological similarity of source words. Both these examples point to the linguistic playfulness manifested in the interplay between different strategies of decoding. Therefore, the term semantic dynamism is chosen to summarize the flexible deployment of association networks in blends.

Secondly, the productivity rules of processes such as derivation or compounding do not seem to apply to blends (e.g. Cannon, 1986; Dressler, 2000; Bauer, 2006). Studies such as Lehrer (1996), Kelly (1998), Gries (2004, 2006), and Beliaeva, (2014) show instead that phonological, prosodic and cognitive factors underlie their complex patterns of encoding. Kemmer (2003) even questions whether blends should be considered a part of morphology at all (see also Dressler, 2000), in that they are “not really composed of morphemes in the sense of recurrent minimal meaningful parts” (Kemmer, 2003, p. 77). This has led some researchers to more or less discount blends from the domain of traditional morphology. To address this Dressler (2000) suggests the term “extragrammatical morphological operation” (EMO) to denote blending and other word formation processes that do not fit in a rule-based morphology. It is furthermore suggested that blends do not pertain to any particular language module, but are better seen as instances of “interaction of modules involved in processing language” (ibid., p. 4).

Thirdly, contraction of source words is considered a core feature in almost all accounts of blending (e.g. Algeo, 1977; Bauer, 1983; Lehrer, 1996; Kelly, 1998, Fabregas & Scalise, 2012; Beliaeva, 2014 to name a few). There is however disagreement as to what type of contraction is required, which have implications for taxonomical reasoning (cf. section 2.1.3).

The notions of *neo-classical compounds*⁸ and *morphological lexicalization* illustrate this clearly. Bauer (1983) exemplifies the former category with the lexical item *arcology*,

⁸ E.g. *biocide*, *electrophile* and *megacity* (examples from Bauer, 1983).

which is claimed to be a blend derived from *architectural ecology*. Although it is originally the result of a blending process, the lexical structure “may lead to [a] re-evaluation” (Ibid. p. 236) resulting in the view that it is a product of affixation. In cognitively oriented terminology this can be described as a usage-based influence of an image schema (Taylor, 2002; Tomasello, 2003; Langacker, 2008). I.e. the similarity between *arcology* and, for instance, *musicology* activates a generalized conception of how this type of lexical structure should be processed regardless of its original semantic structure.

Morphological lexicalization is even more indicative of the relation between compounding and blending. The famous example *cheeseburger* was likely processed originally as a blend (*cheese* + *hamburger*), whereas recent forms such as *teriyaki burger*, *veggie burger*, and *rice burger*⁹ seem to be the result of compounding (McMahon, 1994; Kemmer, 2003; Bauer, 2006; Lehrer, 2007; Beliaeva, 2014). Other notable examples are for instance *-scape* (e.g. *starscape*; Bauer, 2006), and *-erati* (e.g. *glitterati*; Kemmer, 2003). Cases such as these typically depend on in depth analyses in order to determine whether a segment is a lexical morpheme or an unanalyzable part of a monomorphemic lexical item. In addition, such investigations need to consider the significant speed of their diachronical development (Lehrer, 2007).

The relation between blends and complex clippings¹⁰ constitutes another example of the implications of differing patterns of contraction. Besides the differences as to clipping strategy (see section 2.1.3), Beliaeva (2014) shows that these categories are distinguished in terms of how they are processed. While complex clippings such as *digicam* (*digital* + *camera*) commonly occur in syntactic combination, blends such as *stoption* (*stop* + *option*) are less frequently so. Thus, complex clippings are seen as a type of contracted compound resulting from speech economy principles, while blends are “instances of creative word formation involving the formation of new notions in the process of conceptual integration” (Ibid. p. 51). Therefore, since they differ in both origin and function, a classificatory distinction is empirically motivated.

The cognitive significance of contraction is further discussed by Ronneberger-Sibold (2006), who stresses the fact that structural fusion, or contraction, also mirrors the fusion of designata in an iconic way. This is exemplified with the blend *smog*, which structurally mirrors “the indivisible result of a mutual penetration of smoke and fog” (p. 161). In contrast, the structure of compounds is instead transparent, non-contracted, and “rule-governed or lexically determined” (ibid. p. 161). Compounds are therefore considered less iconic than blends as regards form in relation to designata. (cf. Kemmer, 2003). Thus, contraction in blending represents more than apparent phonological, prosodic, or orthographic manifestations. The various ways in which it is elaborated exhibit the impact of structure as a functional property. I.e. the structure is creatively elaborated in use in the construal of new meaning.

⁹ http://en.wikipedia.org/wiki/List_of_hamburgers Retrieved 2015-02-04.

¹⁰ The term *clipping compound* is also used to denote this category.

In all, the terms *semantic dynamism*, *complex encoding* and *functional structure* summarize central properties of blending in relation to other word formation processes. In isolation, these aspects may well apply to other phenomena as well, but in combination they highlight the specific character of blends. This is precisely why a wide-ranging scope of explanatory models is necessary in this field.

2.2.2. Implications of psycholinguistic processing

Morphological theory commonly makes a distinction between the theoretical tenets *Item-and-Arrangement* (IA) and *Item-and-Process* (IP) (Fabregas & Scalise, 2012). In short, IA accounts claim that separately stored morphological units are computed into complex morphemes, or *words*, whereas IP based theories reject the notion of morphemes, and hold that words are stored separately in the lexicon (cf. Sereno & Jongman, 1997). In psycholinguistic terms, Harley (2008) calls this distinction *obligatory decomposition hypothesis* and *full-listing hypothesis* respectively. Strong versions of IA and IP pose, however, specific problems to blending since this phenomenon typically ranges from highly entrenched, or lexicalized, expressions to products of deliberate and conscious wordplay (e.g. Algeo 1977; Ronneberger-Sibold, 2006; Alm-Arvius, 2012). Moreover, the fact that blends often undergo a diachronic process of semantic *demotivation*¹¹ (McMahon, 1994; Schmid, 2001; Fandrych, 2008) further stresses the difficulty to ascribe general properties of lexicalization and morphological processing to blending.

Subsequently, since the IA vantage point is problematic in relation to lexicalized blends, and IP approaches face problems with the processing aspect of blending, another perspective is required (cf. Cannon, 1986). The *dual-pathway hypothesis* presented in Harley (2008) encapsulates both possibilities of production and is therefore more suitable to descriptions of blend formation. As regards regular morphological operations it is claimed that “[a]lthough it is uneconomical to list all inflected words, some frequent and common inflected words do have their own listing” (ibid, p. 191; see also Sereno & Jongman, 1997). The underlying rationale behind this claim is the principle of maximal efficiency of cognitive operations.

Indeed, from a psycholinguistic perspective blends constitute a striking piece of evidence as such in favor of the dual pathway hypothesis in that both lexicalization strategies (cf. Bauer, 1983) and deliberate online processing (cf. Alm-Arvius, 2012) are cognitively implemented. Furthermore, the fact that current morphological research usually assumes the dual pathway hypothesis (Plag, 2006; Hay, 2001) is indicative of the interrelation between regular morphological operations and blending. This is particularly noteworthy in relation to the expectation that research on blends may also have broader morphological implications (Dressler, 2000; Ronneberger-Sibold, 2006).

¹¹ Schmid (2011) describes semantic demotivation as a process whereby a complex lexeme gradually loses meaning that can be discerned in its constituent lexical parts. Indeed, this process is also common in other word formation processes (e.g. McMahon, 1994; Fabregas & Scalise, 2012).

3. Method

Investigating linguistic novelty with the help of corpora involves a set of methodological challenges. First of all, it is necessary to retrieve empirical material which status, at least to a certain degree, qualifies as novel. The difficulty arises from the fact that traditional corpora commonly generate linguistic material that is already fairly established, whereas the opportunities to study more recent phenomena remain few. Corpora collecting data from traditional written sources, such as the *Corpus of Contemporary American English* (COCA)¹² or the *British National Corpus* (BNC)¹³, are therefore limited in this respect.

To address this issue, the Internet domain is chosen to provide data for the study because of its dynamic character and its lower editorial thresholds (cf. Crystal, 2001, 2011). These aspects enable linguists to access large amounts of data that not only include novel phenomena, but also allow for an analysis of both formal and informal language use (ibid.).

The corpus that has been chosen as the empirical source is the *Corpus of Global Web-Based English*¹⁴ (GloWbE, pronounced /gloʊb/). This corpus, consisting of 1,9 billion words retrieved from 20 countries, includes both the COCA as well as the *Corpus of Historical American English* (COHA). For each token in the GloWbE, context as well as URLs¹⁵ are provided if requested, which simplifies a contextual analysis. This feature is particularly important in the domain of Internet media due to the transience of web material in general (Crystal, 2001); a specific URL may well be inactive at the time of retrieval, which becomes problematic if the context of an instance needs further examination. Figure 2 shows a typical example of the detailed presentation of a corpus hit including its source reference and an expanded context.

¹² <http://corpus.byu.edu/coca/> (Retrieved 2015-03-26)

¹³ <http://www.natcorp.ox.ac.uk/> (Retrieved 2015-03-26)

¹⁴ <http://corpus.byu.edu/GloWbE/> (Retrieved 2015-03-26)

¹⁵ *Uniform Resource Locator* (URL) is the technical term in the Internet domain for what is commonly referred to as a *hyperlink*, or simply *link*.

Source information:	
Country/genre	United States: General
Title	So What's Happening in the Petraeus Scandal? An Update News
Source	http://www.towleroad.com/2012/11/so-whats-happening-in-the-petraeus-scandal-an-update.html
Expanded context:	
am amazed, though, at the 20 - 30,000 emails from Gen. Allen to Mrs. Kelley. In my entire electronic life, I doubt I've sent a total of 20,000 emails. He's supposed to be commanding a war. When did he find time for that much email time with one person? # Posted by: jpeckjr Nov 13, 2012 9:52:07 AM # Judy Smith has a crap client list. # Posted by: John Nov 13, 2012 11:03:21 AM # That Eric Cantor knew of this on Oct. 27 pretty much blows holes in the Foxpublican theory that POTUS sat on this story until after the election. and why were they talking to EC and not, oh let's see, POTUS or Dianne Feinstein, Chair of the Senate Intelligence Committee? and how does one send 20-30,000 emails to one person? i smell a rat. # Posted by: deedrdo Nov 13, 2012 11:56:49 AM # " OMG men in the military chasing women! That is the second most important thing they do to amuse themselves. The first most important is killing	

Figure 2. Example of the GloWbE output design with expanded context.

Since collecting texts from the Internet positions the GloWbE in the domain of Internet language a few remarks concerning this specific linguistic domain are needed here. Crystal (2001) describes several issues related to the Internet technology and its users. The transience of the technology itself has already been touched upon, but usage patterns seem no less unstable. Rapid changes in both hardware and software development require users to “adapt their language to the demands of new contexts”, and give them “fresh opportunities to interact in novel ways” (p. 224). As a consequence, research conducted in this domain needs to take into account the considerable speed of language change.

Furthermore, Crystal (2001) argues that the traditional division between the spoken and written language does not easily apply to the Internet domain. It is claimed that common forms of Internet communication such as e-mail and chat groups “display several of the core properties of speech” (p. 29). In Crystal (2011) it is suggested that Internet language could be seen as “writing which has been pulled some way in the direction of speech” (p. 17). Seen from the perspective of the present investigation, this position is notable in relation to the interaction between phonology and orthography in blends (cf. Bauer, 2006). Indeed, this makes the Internet language domain especially appropriate for research on blending.

In the result sections the term *matrix word* has been chosen to represent the lexical items from which all the investigated data has been derived. E.g. *liberal* is called the matrix word of both the blend *lieberal* and the compound *neoliberal*. This term is to do with the methodological aspects of the investigation, whereas the term *source word* is an analytical category used in accounts of specific blends. The *source words* of the blend *lieberal* are thus *lie* and *liberal*. If there is a need to specify the relation between source words they are called *source word one* (SW1) and *source word two* (SW2) respectively.

Finally, the empirical material is divided into two distinct categories; “decontextualized items”, or *types*, found in the corpus, and “contextualized instances” of them, also

referred to as *tokens* (Taylor, 2002, p. 172). Applying these concepts enables an analysis of both structure and form in that the *types* reveal the characteristics of the investigated items' structural form in relation to their matrix word, whereas the *tokens* reveal pragmatic aspects of the *types*.

4. Results

4.1 The data collection process

Three lexical items (henceforth called *matrix words*) were investigated as to the structure of their derivatives and their word class distribution. The matrix words were *republican*, *liberal*, and *vegetarian*. These items share several features, but they are also significantly different in other aspects. As to similarities, their semantics and pragmatics are related in that they pertain to domains charged with ideological sentiment, which was seen in the data in terms of heated discussions and sometimes even outspoken utterances. This was particularly evident in the lexical items *republican* and *liberal*, which is illustrated in the corpus examples (1)-(3) below. Matrix words, blends, compounds, and other significant forms in the data are presented in italics in the text as well as in examples from the corpus queries.

- (1) All the more reason to hold your nose, vote for whichever *Republican* idiot manages to inherit this mess
- (2) Your ignorant remarks exemplify how idiotic *liberals* are and are honey, you are a *liberal*.
- (3) By the way, I am not advocating for any side. I am a *vegetarian* but I am not going to call anyone a murderer or force my own choices on you

The lexical status of the matrix words is also comparable in terms of etymology and function in that they are lexically well established words (in the sense of commonly used), and function both as adjectives and nouns. The differences are mainly related to their structure. Apart from the syllabic variation, the affixes *re-* and *-arian* present two examples of morpheme-like segments, which was shown to have implications for their blending potential.

The examination was performed in two stages, the first one consisting of a corpus search where the first syllable and the onset of the second one were removed from *vegetarian*, while omitting the first syllable was sufficient in *liberal*. Preliminary tests showed that the two first syllables and the first consonant of the third syllable's onset had to be removed from *republican* in order to obtain sufficient data. The resulting query forms were thus **etarian*, **lican*, and **beral*. These choices were based on the

intention to enable the corpus to generate a considerable number of possible formations, while omitting more material would have made the semantic scope too wide.

The issue of scope relates to the semantic transparency of the blends in the corpus output, which can be illustrated with an alternative query form **eral*. Literally thousands of lexical items in English end with the segment *-eral*¹⁶, which would not only be unmanageable in terms of manual analysis, but would also pose specific problems to blended structures. The sheer number of possible meanings of blends ending with *-eral* would make it more or less impossible to determine whether they are actually based on *liberal*, which is important in this investigation (cf. Ronneberger-Sibold, 2006; Lehrer, 1996).

The forms generated in the first stage resulted in three lists of lexical items; i.e. one list for each matrix word. These lists were then analyzed manually, which resulted in a database consisting of the categories *blends* and *compounds* for each matrix word. The compounds were chosen as a reference since they proved to be the predominant word formation process besides blending. Since one of the aims was to look for distributional patterns of blended constructions, the compounds were considered important as a reference. The well attested kinship between blends and compounds, as well as the immense body of research on the latter category, made them ideal for this purpose.

All selected types had either *republican*, *liberal*, or *vegetarian* functioning as their semantic head. As regards criteria for considering a lexical item a blend, clipping of segments in the beginning of SW2 was the primary determining feature, while no distinction was made between paradigmatic and syntagmatic blends. Also, wordplay and fusion of designata were considered significant but not necessarily restricting features.

The tokens as such were not examined in stage one; not least because of their vast number (>200k instances). Instead, the focus of attention was the types that the corpus output would generate. In the lists, 962 lexical items (types) based on one of the chosen segments *-lican*, *-beral*, or *-etarian* (411, 434, and 117 types respectively) were studied with the object to determine their word formation structure. This resulted in 94 types that would provide material for a general analysis as well as a base for the second step of the investigation. For further clarity, the schematic description in figure 3 summarizes the organization of the empirical material.

¹⁶ The corpus would accept 1500 types ending with *-eral*, while queries requesting for higher numbers consistently resulted in the failure note “your query timed out”.

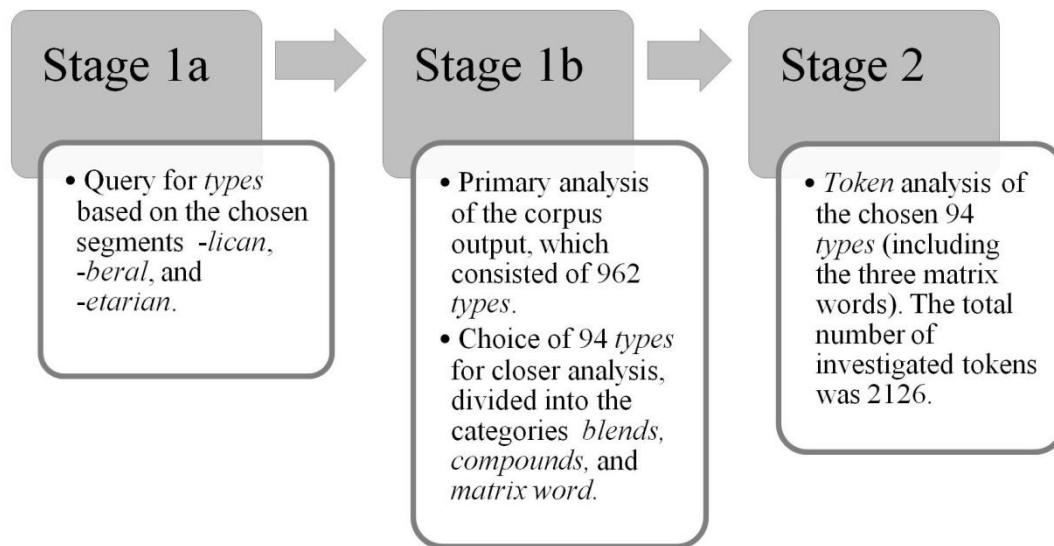


Figure 3. The data collection process.

Furthermore, it was necessary to scrutinize the types manually since the output from Internet sources is notoriously full of misspellings and meaningless forms (Crystal, 2001). Also, due to size limitations of the study not all compounds in the original corpus list were included in the selected 94 types. This means that the type proportion of compounded derivatives could not be calculated. As regards the blends, limiting the number of selected types was a matter of definition; wordplay, fusion of designata, and clipping were considered central aspects, while obvious printing mistakes and proper nouns were excluded. The proportion of blended types in relation to the total output is presented in table 1.

Table 1. Query output and proportion of blends.

<i>Query output</i>	<i>*lican</i>	<i>*beral</i>	<i>*etarian</i>	Total
All query entries (types)	411	434	117	962
Blends chosen for stage 2 (types)	27 (6,6%)	7 (1,6%)	9 (7,7%)	43 (4,5%)

After having sorted the types in the data, the second step consisted of token analysis of the 94 types including the three matrix words. In total, 2126 tokens were analyzed manually for word class distribution. The token frequency variation between the types in the data has several reasons. Obviously, the primary explanation is that there are common forms as well as rare ones occurring in the material. This is however not the only reason. When the generated strings were processed in the concordance software

AntConc¹⁷ to produce manageable lists, a considerable number of duplicates were revealed, which meant that the original number of hits was reduced in many cases.

In table 2 the categories of the data are presented together with the number of types and tokens. The rows *all chosen entries* display all types that were considered relevant for the investigation, whereas the rows *blends* and *compounds* represent the selected types of each category chosen according to the process described above. The *matrix word* row has been included for reference. Because of the varying number, the generated tokens have also been included to illustrate the proportions of the entire database.

Table 2. Types chosen for closer examination.

<i>Chosen types</i>	<i>*lican</i>	<i>*beral</i>	<i>*etarian</i>	Total
All chosen entries: types	49	26	19	94
All chosen entries: tokens	617	1062	447	2126
Blends: types	27	7	9	43
Blends: tokens	128	71	11	210
Compounds: types	22	19	10	51
Compounds: tokens	289	691	195	1175
Matrix word tokens	200	300	241	741

In some cases the string provided by GloWbE was not considered enough to determine the function of a specific instance. The solution to this was to make use of the possibility in the corpus of studying the instance in a wider context consisting of approximately 180 words. Where further investigation was needed the original HTML¹⁸ source was finally consulted if available.

4.2 Structural analysis

4.2.1 Patterns of the matrix word “republican”

Of the 411 types generated by the corpus search on the segment *-lican* only four were found in *Cambridge Advanced Learner’s Dictionary* (CALD), namely *republican*, *publican*, *Anglican*, and *pelican*. According to CALD, *publican* is mainly used in

¹⁷ Descriptions and program files for download are found at <http://www.laurenceanthony.net/software.html> (Retrieved 2015-03-25.)

¹⁸ According to *Cambridge Advanced Learner’s Dictionary* (CALD), *HyperText Markup Language* (HTML) is the term used to denote “a way of marking text so that it can be seen on the Internet”. In other words, web sites are encoded in HTML and then decoded by web browsers such as Internet Explorer, Firefox and Google Chrome so as to be accessible for Internet users.

Australian contexts denoting “a man who is in charge of a pub or bar”¹⁹. Also, 264 out of 411 were hapaxes, which gave a first indication of the creativity of these derivatives. The fact that a number of the retrieved types displayed non-traditional spelling conventions (e.g. *r3publican*) also attested to the notion of novelty and informality. A significant number of the items were web aliases, i.e. proper nouns, and were therefore considered of minor importance for the purposes of the investigation.

A preliminary analysis resulted in a list of 34 structural blends with the segment *-lican* as head. The context of each of the blends was controlled to ensure that this segment was derived from *republican*. Among these, seven were either web aliases or misspellings, and were thus excluded. The 27 remaining blends consisted mainly of seemingly incidental wordplay, whereas four of them had a token frequency of ≥ 10 . These items were *teapublican*, *rethuglican*, *demopublican* and *rubepublican*. Interestingly, *demopublican* displayed variation as to segment breakpoint in the lexical item *demepublican*. Due to this, the latter was treated as a type of its own rather than a variation of *demopublican*.

The character of the blends in the data varied considerably. There were instances with clipping of only SW2 such as *rubepublican* and *Foxpublican*, as well as intricate formations including elaborated clipping, overlap, and wordplay (e.g. *repooblican*, *Appublican*, *repiglican*, and *rape-ublican*). A close analysis also showed that there were examples of both *paradigmatic* and *syntagmatic* blends as well as instances where this character could not be determined out of context. For instance, *teapublican* could be seen as a syntagmatic blend derived from a syntagm such as *a tea-party movement republican* (cf. Algeo, 1977), whereas *reprobatelican* was considered paradigmatic (emerging from the conjunctive phrase *reprobate and republican*; cf. Kelly, 1998). However, the analysis of blends such as *talibpublican* and *demopublican* depends rather on word class designation. These instances allow for both syntagmatic and conjunctive rephrasing due to the grammatical ambiguity of their source words. Since both *a taliban republican* and *taliban and republican* are possible rephrasings, *talibpublican* can be seen as both syntagmatic and paradigmatic. Determining the type of a specific instance requires therefore an analysis of the context of each token.

An instance clearly displaying the creativity was seen in the blends *rape-ublican* and *rapepublican* (*rape* + *republican*). Besides probable differences as to aspiration and stress, these items can be described as phonological neighbors of *republican* (/reɪˈpʌblɪkən/ contrasted with /rɪˈpʌblɪkən/; cf. Sandra, 2009). *Rapepublican*’s structural similarity to a complex morpheme suggests however that the phonological realization implies gemination²⁰ of /p/ (Spencer, 1996), which makes this form phonologically less similar to *republican*. The orthography of *rape-ublican*, on the other hand, avoids the gemination effect but exhibits instead a less transparent spelling. There

¹⁹ Apparently, CALD seems to suggest that only men can hold this position, which is not entirely correct. According to Blocker et. al. (2003) there is a long tradition of women as pub managers, the first registered occurrence dating back to 1797. To the defense of CALD it should be noted that this lexical item has two entries in the dictionary, of which one of them defines the term simply as “the manager of a pub”.

²⁰ Yavaş (2006) describes gemination as a “prolonged closure” (p.59) often seen in homorganic consonants in complex morphemes such as *rat tail*.

seems thus to have been a balancing operation at work offering a choice between a phonological and an orthographic optimum.

Another significant pattern found in the data was the tendency to exploit the similarity to a complex morpheme such as *rebuild*. Although *republican* is not possible to decompose morphologically, the users seem nonetheless to have applied strategies that resemble this in many of the blends. No less than 13 of the 27 blends retained the segment *-publican*, which attests to the cognitive impact of the morphological rule allowing for the use of the prefix *re-*.

Overall, the creative elaboration of *republican* was apparent in the empirical material. Among the derived blends there was not only a seemingly arbitrary distribution of well-known type features. The data also showed that morphological affixation have influenced the blend structure of almost half of the investigated types.

4.2.2 Patterns of the matrix word "liberal"

The structural differences between *-beral* and *-lican* became clear at stage 1b of the data collection process. The percentage of blends among the 434 types of the first query was significantly lower, and elaboration of the matrix word's internal structure was also less common. Furthermore, in the list of corpus entries only *liberal* and *illiberal* were found in CALD, while premodifier entries in the dictionary such as *neo-*, *non-*, *anti-* and *ultra-* were common in the output (39 neo-classical compound types²¹). In other words, although *liberal* seemed to be productive as to neo-classical compounding, it produced few blends.

These results should however not be surprising, given the few syllables and the considerable number of lexical neighbors ending with *-eral* (cf. Lehrer, 2007). The latter was however apparent even with the more marked formula applied in the corpus query (i.e. **beral*). Thus, although the query form could be described as relatively marked, the closeness to lexical items such as *general*, *federal* and *literal* seems to have reduced the possibilities of effective wordplay. One reason for this is found in the fact that blends formed with *-eral* as head have a considerably low level of semantic transparency.

Seven blends out of nine possible lexical items were chosen in the primary analysis of the corpus output; *fiberal*, *gliberal*, *laberal*, *lie-beral*, *lieberal*, *diberal* and *liarberal*. Overlap and phonological similarity of source words (cf. Kelly, 1998) were common in these items.

An analysis of the token distribution of the blends showed that *lieberal*, *laberal*, and *lieberal* represented 60 out of 71 instances, which means that the other types were uncommon forms. Because of the low frequency both as to types and tokens the idea that the matrix word *liberal* is not optimal for blending was thus supported by the data. The fact that there were six hyphenated tokens of *lie-beral* also raises the question

²¹ There were also border-line instances such as *homoliberal* and *metro-liberal*, which would have increased the number even further.

whether some users have found it necessary to emphasize *lie* to avoid misunderstandings related to misspelling.

In all, the corpus data showed two distinct word formation tendencies related to this matrix word. Its blending potential seems on the one hand to be problematic from a quantitative perspective. On the other hand, the blends that were found in the data displayed a relatively high density of type-specific features described in sections 2.1.1-2.1.4.

4.2.3 Patterns of the matrix word “vegetarian”

The word formations in the data based on *vegetarian* differed significantly from both *liberal* and *republican*. Firstly, 35 of the 117 types had no relation to vegetarian food at all. For instance, the lexical item *proletarian* occurred both independently and in compounds such as *slum-proletarian* and *anti-proletarian*. Wordplay such as *proNetarian* and *frisbeetarian* were also instances unrelated to the semantic domain of *vegetarian*. Both these lexical items reveal however important aspects of word formation in this context, which will be presented below.

Furthermore, the analysis of stage one proved to be a complex task due to the fact that the investigated segment *-etarian* can be related to the established, albeit rather infrequent, bound morpheme *-arian* (Quirk et. al., 1985). Corpus hits such as *monetarian* and *proprietary* displayed creative use of this morpheme, and were analyzed as incidental derivations of *monetary* and *proprietary*, or *proprietor*, respectively. The blend *vegetetarian* (*vegetate* + *vegetarian*) was one of the few blends that met traditional criteria for defining blends such as clipping, wordplay, and fusion of designata.

However, in lexical items such as *grainetarian*, *dietarian*, *eggetarian*, and *flexetarian* the distinction between blending and regular morphological derivation became less obvious. On a structural level *dietarian* displayed a productive use of the morpheme *-arian*, while the other items retained the segment *-etarian* from *vegetarian*. Judging from the context, it was also clear that a semantic reference to *vegetarian* was implied, which is seen in example (4):

- (4) I returned from a high school trip to visit my cousin at UGA declaring I was a vegetarian. It didn't last very long as I am not a huge fan of vegetables. (I would be an awesome *grainetarian*.)

Therefore, since there were both structural and semantic clues indicating that blending strategies have been involved in the use of these terms, they were considered blends rather than complex morphemes. All the same, the impression was that they had much in common with segment development of the type *burger* and *-(o)holic* (cf. section 2.2.1). In many cases the productive use of *-etarian* was therefore not related to blending per se, but rather to compounding and derivation. Indeed, the data indicated that *vegetarian* and related derivatives seem to undergo a process of morphological lexicalization.

The example *proNetarian* demonstrated the use of creative and productive strategies related to the morpheme *-arian* in at least three ways; neoclassical compounding (*pro-*), derivation (*-arian*), and orthographic marking. The latter is seen in the use of the capital letter *N*, which emphasizes the reference to *Net* and its superordinate domain *Internet*. Furthermore, given the tendencies towards morphological lexicalization of *-etarian*, the capitalization of *N* may be seen as a strategy to avoid ambiguity with a hypothetical form such as *pronetarian* (*prone* + *etarian*). Example (5) also shows the importance of context in the process of decoding this lexical item. The combination of Internet metaphors such as *platform* and *cloud* typically primes the conceptual domain of *Internet*:

- (5) We urgently need platforms in the cloud and educational programs done at country and regional level to train our students to be *proNetarian*, but also to create and protect their knowhow in a virtual economy.

A similarly creative nonce-formation found in the data was *frisbeetarian*, where semantic cues were seen in the context of this instance as well. In example (6) the wordplay likely alludes to presbyterianism, at least in terms of lexical structure, exploiting the phonological similarity between the nonce-formation and its source counterpart (/frɪzbɪˈtɪriən/ and /prezbɪˈtɪriən/). The comic and slightly absurd effect is realized in the merging of the religious domain and an apparently unrelated concept with only one syllable exchanged:

- (6) Is this going to be a religious question? Because as you know we're all members of the *Frisbeetarian* Order. Just signed up since we've come to America. We believe that when you die your soul goes up on a roof and you can't get it down (general hilarity).

Considerable linguistic creativity is thus exhibited in both examples (5) and (6), and besides the blended character of the latter the productive potential of the morpheme *-arian* and the segment *-etarian* is illustrated clearly in these instances.

4.3 Analysis of word class distribution

4.3.1 Word class distribution of "republican"

Among the 200 tokens of the matrix word *republican* 83% were used as adjectives and 17% were used as nouns, while the proportions in the blend category were 76% and 24% respectively. Example (7) displays prototypical instances found in the data. Each noun phrase including the analyzed component has been underlined.

- (7) # after the last election I do not think the TEApublicans will be happy with a VP candidate yet again sooo they will push HARD for a TEApublican candidate to head the ticket (*Nom.* + *Adj.*)

In *demopublican*, as well as in the 25 instances that were either hapaxes or types generating only two tokens²², adjectival and nominal use was distributed similarly to the matrix word. *Rubepublican*, on the other hand, displayed only adjectival use. It is also worth noting that the spelling of *demopublican* and *rethuglican* showed variation, although sparingly, as to hyphen insertion. In all, the general tendency was that the word class distribution in the blends seemed to be preserved, or slightly leaning towards an increase in nominal use.

In the category of compounds the proportion of adjectival use was higher, except for *tea-republican*, where, interestingly enough, nominal use was more frequent. The neo-classical compounds *prorepublican* and *nonrepublican* (92 tokens in total) displayed no nominal use at all, and there were no more than six instances of nominal *anti-republican* among 114 adjectival ones. The label *(*)republican* in figure 4 has been chosen to represent 18 types with up to eight tokens each (mean value 3,3 tokens). In this subcategory the compound *arch-republican* (two instances) was only used as a noun, while most types were exclusively adjectival.

Figure 4 illustrates the distribution of each category processed according to the procedure described above. The tokens of each category have been included as numbers in brackets to stress the importance of considering this representation as a procedural clarification rather than a statistical illustration. Had these figures not been included, the table might mislead the reader to interpret infrequent categories as more statistically significant than they actually are. The category *(*)lican* represents the 25 blends with ≤ 2 tokens.

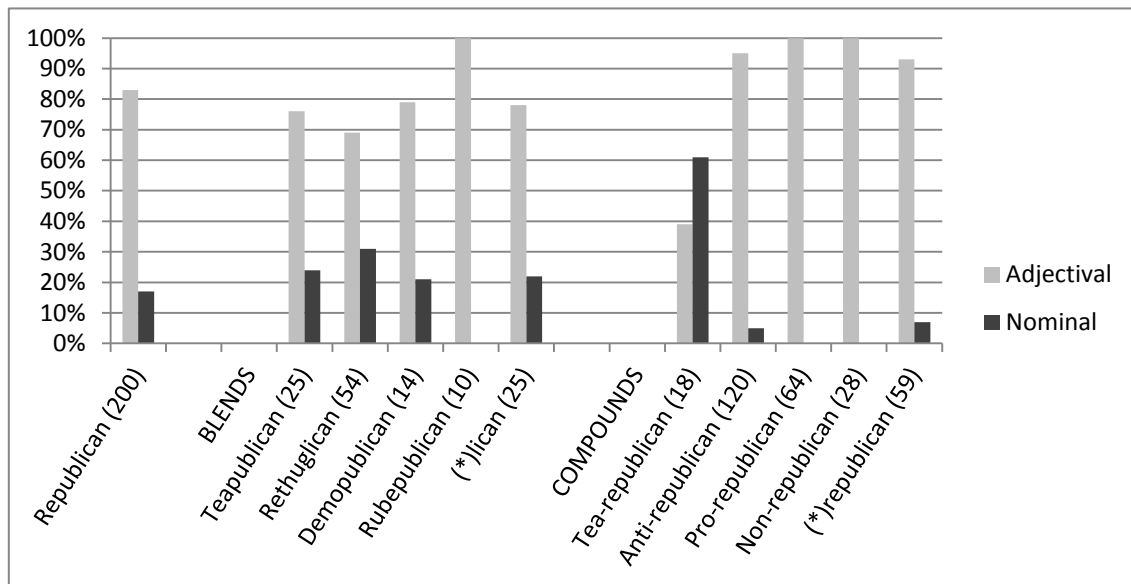


Figure 4. Adjectival and nominal distribution in blends vs. compounds based on the matrix word *republican*. The numbers of instances are put within brackets.

²² The full list is presented in Appendix A.

Despite some variation, the data showed that the 22 compounds chosen in the list of 94 types displayed slightly more adjectival use than the blends and the matrix word. In total, 76% of the blends were used as adjectives, while the adjectival use of the compounds amounted to 93%. This can be compared to the distribution of the matrix word *republican* (83% adjectival use). The statistical measures relevant for the investigation are illustrated in figure 5, where the categories *blends* and *compounds* have been compared to the matrix word *republican*.

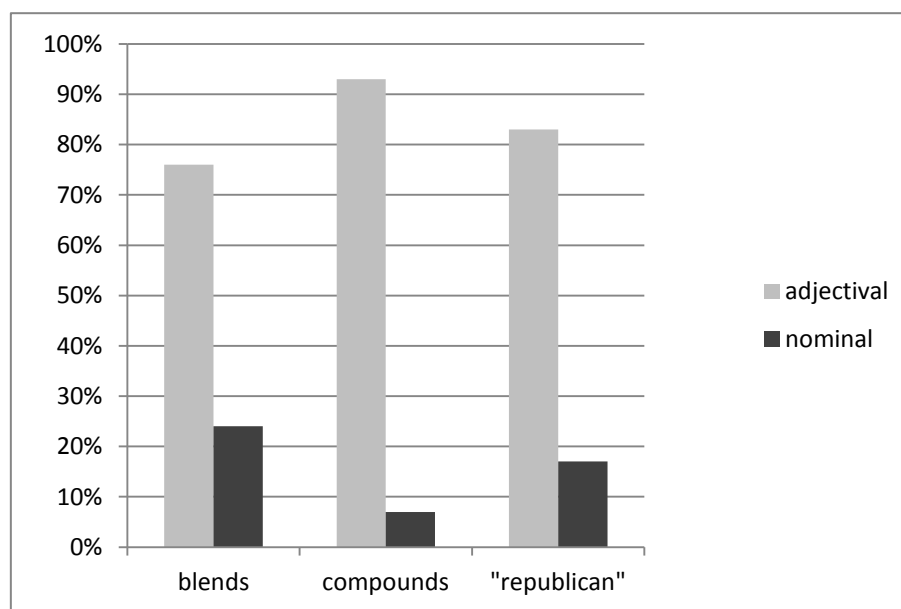


Figure 5. Adjectival and nominal distribution in blends vs. compounds based on the matrix word *republican*.

Notably, the type *tea-republican* represented 11 out of 21 instances of nominal use in the category of compounds as a whole, which suggests that *tea-republican* is not entirely representative in this respect. If *tea-republican* were to be excluded, the adjectival use would subsequently be even more favored in the category of compounds.

4.3.2 Word class distribution of "liberal"

The analysis of the 300 tokens of the matrix word showed that adjectival use was the common function of this lexical item (87%). In total, the blends displayed a slightly lower percentage in this respect (80%). The figures of the data analysis revealed however that *lieberal* together with its variant *lie-beral*, representing 47 of the 71 analyzed blend tokens, exhibited more or less the same pattern as *liberal* (85% adjectival use).

In figure 6 the proportions of adjectival and nominal use are displayed as regards blends and compounds compared with the matrix word *liberal*. Although the structural differences between the matrix words *liberal* and *republican* had other consequences (cf. sections 4.2.2 and 4.4.1), the figure illustrates the similarity as to word class distribution tendencies of their respective derivatives.

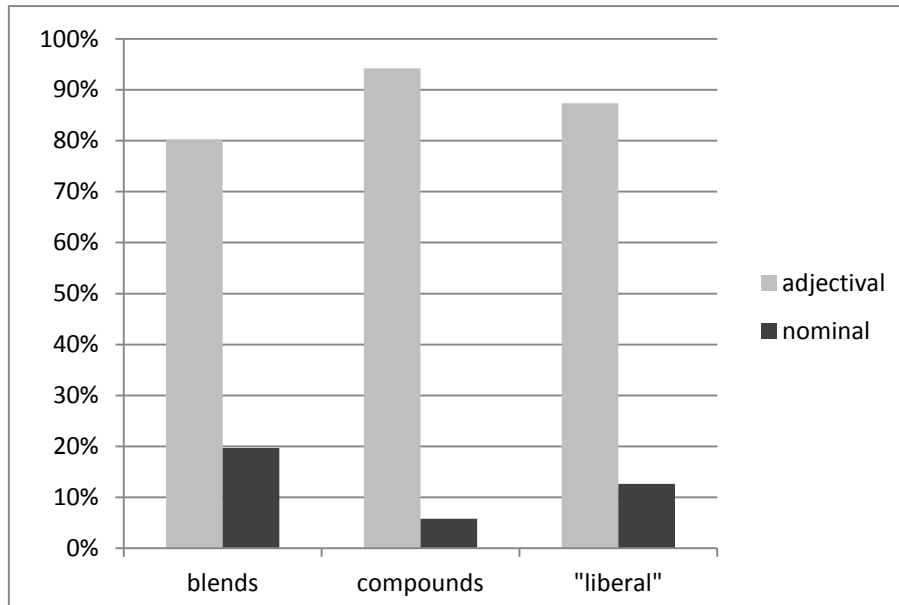


Figure 6. Adjectival and nominal distribution in blends vs. compounds based on the matrix word *liberal*.

Interestingly, the less frequent blend types (mean number of tokens per type <5) were used as adjectives in 71% of the instances. There were thus deviant patterns in *liberal* as well as in the matrix word *republican*. The general tendency was however that the distributional proportions of the matrix word were similar in the blends, although there was a slight indication of more nominal use.

4.3.3 Word class distribution of "vegetarian"

As regards word class distribution the low number of blended tokens did allow for few statistically based conclusions. It was nevertheless clear that the distribution of adjectival and nominal use of the matrix word differed from *republican* and *liberal* in that the proportions of adjectives and nouns were comparatively even. In the analyzed compounds the pattern was indicative of primarily adjectival use, which could also be seen in the other matrix words.

In order to underscore the classificatory uncertainty of the *blend* category, its bars in figure 7 below are represented with dotted lines and a slightly different shade. The low number of tokens in this category (11 instances) also means that the statistical significance is low.

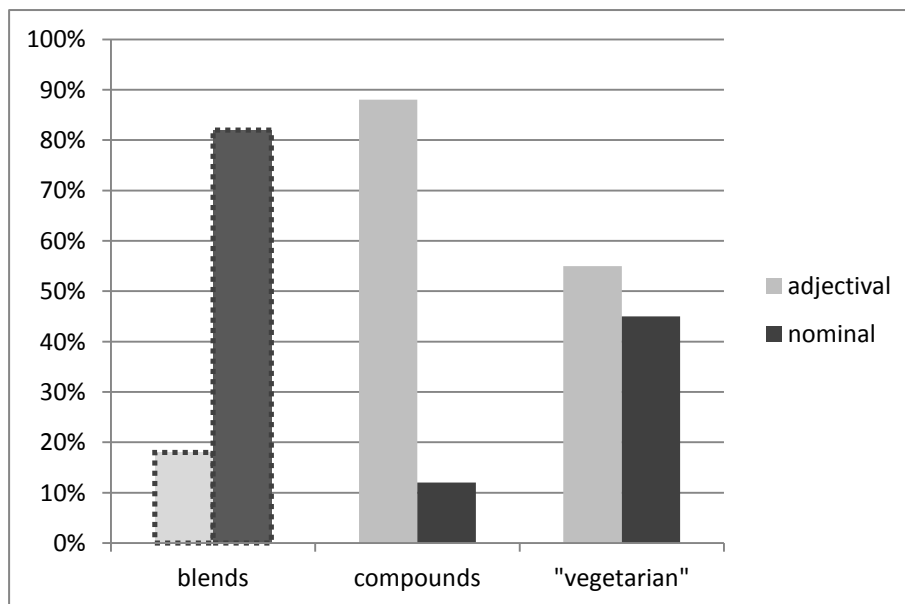


Figure 7. Adjectival and nominal distribution in blends vs. compounds based on the matrix word *vegetarian*.

Although the analysis of the *blend* category produced little information, the pattern of the derived compounds was more significant. The difference in nominal use between the matrix word and the compounds (45% and 12% respectively) is indicative of a drift in the semantic structure. While the semantic feature of denoting an animate entity is inherent in the nominal realizations of the matrix word, this property is reduced in the compounds where nominal use is less frequent.

4.4 Summary of results

4.4.1 Structure

Significant differences in word formation potential between the matrix words were found in the data. *Republican*, to begin with, generated a substantial amount of both blends and compounds. 23 out of 27 blends exploited the morpheme-like division between *re-* and *-publican*. *Liberal* seemed to be more restricted as regards creative word formation processes. The data indicated that this matrix word is less suitable for segmenting compared to the other matrix words. Instead, neo-classical compounding with the addition of a short premodifier was favored.

Also, *republican* and *liberal* differed in terms of the character of the blends they generated. Types based on *republican* displayed various forms of clipping of source words, both with and without overlap (e.g. *repukelican* and *teapublican* respectively). In items ending with *-lican* a significant degree of creativity as regards overlap and clipping typically corresponded with low frequency, the exception being *rethuglican*. In contrast, blends based on *liberal* such as *gliberal* (*glib* + *liberal*) exclusively exhibited overlap. The pattern of overlap in these blends was also fairly consistent. Except for the seemingly deliberate spoonerism *diberal lemocrats*, the *liberal* blends displayed overlap with two letters in five cases (e.g. *lieberal*), and three letters in one type (*gliberal*).

Vegetarian showed a certain potential for creative derivatives. There were a few instances of wordplay deploying phonological and orthographic strategies. Prototypical blends were however scarce, and the clipping and overlap patterns showed little variation. Instead, the segment *-etarian* showed signs of being used creatively and productively in a morpheme-like fashion. The generated context of the latter instances indicated the importance of semantic priming in that the segment in itself is relatively void of meaning. In other words, in applying food related terms users seemed to emphasize the intended conceptual domain if the term was not self-explanatory.

In all, the results show that the matrix words exhibited considerably differing structural patterns in the forming of blends and other creative word formations. There were clear indications that the structure activated a variety of processes, but also that certain lexical conditions seemed to favor other word formation processes over blending.

4.4.2 Word class distribution

The general tendency of the blends in the investigation was to adopt the distributional pattern of the matrix words, be it with slightly more nominal use. There was also considerable variation between the blend types. While blends with token frequency ≥ 10 were used similarly to the matrix words, there were also instances of deviation from this pattern. The compounds in the data, on the other hand, exhibited a consistent tendency towards adjectival use. Besides obvious structural differences, the data suggests therefore that the blends and compounds differ in terms of grammatical usage. This can be compared to Beliaeva's (2014) findings as to blends and complex clippings. Although these categories may seem similar, blends depend to a much larger extent on recognizability of segments. Complex clippings, on the other hand, already have an established semantic link in that the source words typically co-occur in context. The distinctions are thus found in examining the usage rather than their structure.

Furthermore, no systematic principle could be discerned in the data as to how a specific blend would function in context. This was particularly apparent in low frequency types. In the rather few more frequently occurring forms, however, the distribution resembled the matrix words although the nominal use was somewhat more common.

5. Discussion

As the data of this investigation has shown, there are good reasons to believe that structure as well as grammatical usage influence the process of blending. Section 5.1 will show the impact of constraints concerned with the cognitive underpinnings of word formation. In this part the conditions for blend formation will be discussed, and it will be shown how cognitive processing can explain some of the seemingly capricious characteristics of this phenomenon.

Section 5.2 focuses on how the investigated grammatical aspect, i.e. word class distribution, affects analyses of blends. It will be shown that the usage-patterns seen in

the data sometimes take precedence over traditional structural analyses of isolated instances. It will also be shown that this has implications for taxonomical reasoning.

5.1 Cognitive constraints

There are at least three cognitive aspects influencing the corpus output of the matrix words. One of them, effects of morphological *schema transfer* (cf. Kemmer, 2003), seems to have positive effects on the formation of blends and other creative word formation processes. The other two, *neighborhood effects* (cf. Lehrer, 2007) and *morphological lexicalization* (cf. Beliaeva, 2014) tend to have negative consequences for blending. These three aspects are subsumed under the notion of *cognitive constraints*²³.

In *liberal* the structural narrowness seems to inhibit some of the extra-morphological creativity, whereas the forms that actually emerge exhibit considerable prototypicality as to blending. The mono-morphemic *republican*, on the other hand, seems to deploy a morphological schema of a complex morpheme, thus fuelling the variety and quantity of creative and novel forms. The generated blends tend, however, to be less prototypical than *liberal* blends. The third matrix word *vegetarian* differs from the others in that morphological lexicalization processes seem to gain precedence over blending. This is seen in that the corpus output generates forms that display patterns resembling regular morphological operations while there are few prototypical blends. Also, in both *liberal* and *vegetarian* there are neighborhood effects influencing blend formation.

The cognitive constraints are discussed in two sections below. Since the data suggests that their effect on blends and other EMOs differ, they have been organized according to positive and negative influence. It will be argued that the constraints discerned in the data largely depend on cognitive processing. Put simply, blends are likely to occur only where it is the most efficient strategy to achieve a linguistic objective such as punning or fusion of forms. For instance, *republican*, *liberal*, and *flexitarian* all display various degrees of wordplay and contraction. However, as will be argued below, while the first two examples exploit the semantic dynamism of blending, *flexitarian* is rather a result of morphological lexicalization.

5.1.1 Positive constraints

Schema transfer effects

The similarity between the productive prefix *re-* and the segment *re-* in *republican* proved to have a significant impact on the corpus output. In terms of phonological and orthographic structure, the form of this matrix word resembles that of a complex morpheme, which enables users to exploit the productive patterns of the latter. There is thus a common denominator between *republican* and *rebuild* in that the structure is

²³ It should be noted that Jackendoff's (1983) use of the term *cognitive constraint* is related to semantic representation.

similar regardless of a morphological analysis. The patterns in the data can therefore be explained in terms of a transfer effect of a productive morphological schema.

The main point here is that users seem to apply a productive schema on a lexical item that should not, strictly speaking, be able to instantiate it. In order to understand the process behind this, it is important to stress that schemas are essentially a cognitive phenomenon. Langacker (2008) describes schematization as “the process of extracting the commonality inherent in multiple experiences to arrive at a conception representing a higher level of abstraction” (p. 17). A more minimalistic definition is found in Kemmer (2003) in that schemas are “simply what two or more forms or meanings have in common” (p.79). In this case the forms of *republican* and *rebuild* have a “commonality” in their possibility to segment *re-*, which seems to be more influential than the fact that they are morphologically different. In other words, it seems convenient from a user perspective to make use of a ready-made morphological schema when possible. Prescriptive rules of grammar are not likely to impede this in actual language use (cf. Tomasello, 2003).

The idea of schema transfer effects is further supported by findings in the domain of psycholinguistics. The research presented in Harley (2008) concerning the *dual-pathway hypothesis* attests to the notion that a language user applies the fastest and most cognitively economic way to process language (cf. Sereno & Jongman, 1997; Hay, 2001). Thus, the cognitive efficiency of the morphological process of prefixation facilitates the forming of blends in *republican*, whereas the same strategy applied to *liberal* rather results in neo-classical compounding. This is achieved in that a cognitive schema of prefixation (with instantiations such as *restart* or *rebuild*) is activated and mapped onto the ostensible instantiation *republican* (cf. Taylor, 2002; Kemmer, 2003). In other words, the phonological segmentation in *republican* benefits from the cognitive efficiency of prefixation.

In this respect *vegetarian* may superficially resemble *republican* in that the results displayed patterns of productive use of the segment *-etarian*. However, these results are better understood as a form of morphological lexicalization, which will be discussed below.

5.1.2 Negative constraints

Neighborhood effects

One of the parameters affecting the corpus output of *liberal* and *vegetarian* can be understood as an impact of lexical neighborhood²⁴ (cf. Lehrer, 2007). As shown in the results there is a profusion of English lexical items ending with *-eral*, and even the more marked segment *-beral* is problematic in this respect (cf. Beliaeva, 2014). Therefore, a supplementary corpus test was performed in order to examine this in more detail.

²⁴ For reasons of space, no distinction has been made here as to *neighborhood size* and *neighborhood frequency* (Sandra, 2009). Should a more in depth investigation be conducted into neighborhood effects on blend formation, this distinction is likely to be a necessary requirement.

Instead of omitting an initial segment the query formulas were reversed resulting in *lib**, *libe** and *liber**. This control test showed however that there were still strong neighborhood effects influencing the derivatives of *liberal*. *Vegetarian* was controlled in the same fashion applying a query form *vege**, which also confirmed the result of the original queries in that it displayed lexical neighbors such as *vegetate* and *vegetation*.

The suggestion of the present investigation is thus that lexical neighbors hamper successful blending. The main reason for this is that the ambiguity of the segments renders encoding and decoding of blends problematic (cf. Ronneberger-Sibold, 2006; see also section 2.1.5). Outside of context, it is for instance difficult to interpret whether a lexical item such as *zombietarian* relates to *vegetarian* or *proletarian* (or some other lexical referent). This type of ambiguity results therefore in forms with low semantic transparency.

The negative effect on blend formation can be described partly in terms of cognitive economy. This is supported by research on the related process of compounding. Libben et al. (2003) have shown that “compounds with opaque heads take longer to process” (p. 62), and Sandra (1990) even claims that opaque compounds are stored separately in the lexicon. Although there are differing views as to storage and decomposition of this type of item (cf. Harley, 2008; Ji et al., 2011), the findings point to the cognitive cost of processing low transparency in complex forms. Besides the extra processing effort on a lexical level, a low transparency complex form typically requires an elaborated context, which is also in conflict with efficiency principles.

The derivatives of *vegetarian* also highlight the importance of conceptual vicinity of the fused domains (cf. Lehrer, 1996). For instance, *grainetarian* needs few clues to be understood as a form of grain diet restriction, since *grain* and *vegetarian* could be described as relatively adjacent cognitive domains (cf. Langacker, 2008). The form *flexetarian* displays a lower degree of conceptual vicinity, which means that contextual clues and level of lexicalization play a more important role. I.e. assuming that *flex-* is derived from *flexible* offers few semantic clues that *-etarian* is segmented from *vegetarian*. *Grain*, on the other hand, activates associations that link *-etarian* to *vegetarian* in that it is more closely connected with the conceptual domain of *food* (Lehrer, 2007). Therefore, neighborhood effects making a segment less transparent may be compensated for by exploiting the conceptual vicinity of source words. This means, however, that the possibilities of forming effective blends become restricted to certain cognitive domains with this capacity.

In short, the observations of the data in combination with the findings outlined above explain why semantic neighborhood suppresses blend formation. The empirical material does not allow for in depth analyses of the degree of its impact, but there are nonetheless clear indications of its significance as a constraining factor.

Morphological lexicalization

The idea that some blend segments gradually develop morpheme-like properties is almost a truism in research on blending (e.g. Bauer, 1983; Lehrer, 2007; Beliaeva, 2014). The matrix word *vegetarian* shows clear signs of this process. The control corpus

test of *vege** revealed not only that neighborhood effects had an influence regardless of initial or final segmenting; the query also produced a considerable number of creative hapaxes. These were of the type *vegenaise*, *VegeBears*, *vegeburgers*, *vegemeat*, *vege-shoes*, and *vegemix*. There is thus a productive pattern in which *vege-* is given a semantic content related to *vegetarian*. The original corpus output is suggestive of this pattern as well, but due to neighborhood effects the combinatory devices are limited to lexical items belonging to the domain of *food*. Furthermore, the consistent breakpoint producing the segment *-etarian* means that it cannot be concluded that there is a schema transfer effect from the morpheme *-arian*.

In isolation there may be good reasons to interpret *flexetarian* and *vegenaise* as blends, but in terms of processing the picture is less clear (cf. Beliaeva, 2014). A comparison with the matrix word *republican* shows that the pattern of *vegetarian* derivatives displays a consistent lexical structure, whereas there is a significant variety in the *republican* data. The schema transfer effects of the latter promotes blending in that it scaffolds creative lexical elaboration, while at the same time it does not seem to impede it. The corpus output of *vegetarian*, on the other hand, shows instead that the breakpoint is apparently restricted to form either of the segments *-etarian* or *vege-*. And, more importantly, the consistent pattern of *vegetarian* indicates that the fixed forms have taken on productive properties allowing modification similar to compounding.

A lexical item such as *vege-shoes* seems therefore relatively transparent and reasonably sensible due to the semantic (and perhaps even morphological) detachment of *vege-* from the source *vegetarian*. This detachment implies a semantic drift (cf. Hay, 2001) altering the conceptual content of *vege-* to the extent that new associations of the segment allow for other lexical combinations. Indications of such a drift were also found in the analysis of the word class distribution of the matrix word *vegetarian* (see section 4.3.3). Were it not so, a conceptual fusion of the sources *vegetarian* and *shoes* would indeed seem innovative. In contrast, a blend such as *frankenpublican* (*frankenstein* + *republican*) “associates to both [...] the lexical items used, creating a semantic composite of rich associative potential” (Alm-Arvius, 2012, p. 209). In this case neither *franken-* nor *-publican* are productive segments, but activate instead the dynamic processes of blending (cf. section 2.2.1).

The suggestion is therefore that the powerful influence of morphological lexicalization to some extent blocks blending. This does not mean that *vegetarian* is less apt to be the base for creative derivatives. Instead, the data shows a number of instances of lexical innovation, although their formation resemble the productive pattern of instances such as *veggie burger* and *shopoholic* (see section 2.2.1).

5.2 Implications of patterns of word class distribution

The findings of the present investigation support the claim that structural properties are not sufficient as a base for category delineation of blends and closely related word formation processes (e.g. Lehrer, 1996; Fandrych, 2008; Beliaeva, 2014). Instead, the blends’ patterns of word class distribution indicate that there are usage-related differences separating them from the category of compounds. Not only was this seen in the strong and relatively consistent preference for adjectival use in the latter, but the

considerable variation in the blend category also attests to the qualitative difference between them. Furthermore, Bauer's (2006) processing distinction between blends and compounds is also mirrored in the data (cf. section 2.1.1). Although Bauer's (2006) term *creativity* (as opposed to *productivity*) is not defined in detail, the data shows that the usage-patterns rely on other processes than purely morphological ones (cf. Cannon, 1986).

A consequence of the variation in the blend's word class distribution is related to the distinction between paradigmatic and syntagmatic blends. This categorization depends to a large extent on grammatical aspects of the source words. The word order restriction in syntagmatic blends (Bauer, 2006) is one example of this in that the origin in a syntagmatic string restricts the word class of the source words. Thus, since the data showed that context is crucial in word class designation of a specific blend, it follows that paradigmaticity and syntagmaticity depend on context as well. Bauer's (2006) example *motel* is straightforward only because of its established lexical denotation; no context is required to rephrase it as *a motor hotel*, which means that it is easily defined as a syntagmatic blend.

However, blends are typically not as lexicalized as in Bauer's (2006) example. This is illustrated with a blend from the data. *Talibpublican* could be rephrased as *a taliban republican* (adjective + noun). English headedness rules predict that this interpretation denotes a specific type of *republican* (noun). Should the segments be reversed, English syntactic rules would require the word class of the source words to be changed in order to form an acceptable syntactic string. The semantics would then change accordingly to refer to a type of *taliban* (which necessarily takes on the function of nominal). This reading means that *talibpublican* is a syntagmatic blend of the type *motel*.

Were *talibpublican* instead to be seen as a contraction of a conjunctive phrase *taliban and republican* (noun + noun), an altered word order would not significantly affect the semantics of the blend²⁵. I.e. a hypothetical form *reputaliban* would essentially retain the meaning of the fused elements. This property is analogous to the character of the blend *smog* (*smoke* + *fog*) in that there is no logical impediment to a hypothetical reversed form *foke*²⁶ (Kelly, 1998). Moreover, the fusion of designata would also be the primary focus of attention mirrored iconically in the fusion of the forms (Ronneberger-Sibold, 2006). In this interpretation *talibpublican* is thus a paradigmatic blend.

The important point to be made here is that determining whether a blend is paradigmatic or syntagmatic relies heavily on how the users actually apply it. The structure in itself cannot be taken for granted as a defining feature. Therefore, excluding either category in accounts of blending, which is not an uncommon practice, may have repercussions for questions of validity if the context is not studied carefully.

²⁵ Indeed, it could be argued that headedness rules would apply here as well. The influence is however a matter of degree in this case, whereas in a syntagmatic interpretation they would completely alter the denotation.

²⁶ Kelly (1998) suggests that the first source word is more frequent, contains fewer syllables, and "denote more prototypical category members than the second word" (p. 579). The particular instance Kelly (1998) chooses as an illustration would, however, not likely appear as an authentic blend due to neighborhood effects (P. Shaw, personal communication, June 3, 2015).

In all, the analysis of word class distribution has shown two things. Firstly, a distinction can be made between blends and compounds in terms of grammatical usage. Adjectival use tends to be preferred in most compounded derivatives, whereas there is more variation in the use of blends. In addition, comparatively frequent blends display distributional patterns similar to the matrix words. Secondly, the categories *syntagmatic* and *paradigmatic* cannot be ascribed to decontextualized instances of blends. The reason for this is that these categories depend to a large extent on word class designation, which has proved to be unpredictable outside of context, especially as regards low frequency blends.

These aspects taken together suggest that analyzing the structure is not sufficient in analyses of blending. Neither are theoretical models reliable if a usage-based perspective is not taken into account. The reason for this is simple; blend formation is not only a matter of structural principles, it is also inherently usage-driven. This means that it is a dual process deploying both consolidated linguistic schemas (such as phonological, syntactical and morphological ones) and innovative strategies extending the inventory of available linguistic symbols. In this respect blends constitute therefore a manifest link between conventionalized structures of language and creative linguistic processes.

5.3 Summary of discussion

The analysis of the data shows that the three cognitive constraints have the possibility to fuel or hamper blending in intricate ways. The overall picture is that schema transfer effects in *republican* derivatives influence blending in a positive way, whereas neighborhood effects and morphological lexicalization observed in *liberal* and *vegetarian* obstruct blend formation. In cognitively oriented terms, the schema transfer effect can be described as a mapping from one cognitive domain to another (i.e. from the process of *prefixation* to *blend formation*). In contrast, neighborhood effects and morphological lexicalization function as competing processes. Morphological lexicalization seems to be particularly powerful in this respect, which is a consequence of the more efficient cognitive potential in regular morphological operations. Figure 8 is intended to illustrate how the cognitive constraints interact with the mechanisms of blend formation.

	Schema Transfer Effects	Neighborhood Effects	Morphological Lexicalization
<i>Semantic dynamism is...</i>	<i>enhanced.</i>	<i>impeded.</i>	?
<i>Complex encoding is...</i>	<i>enhanced.</i>	?	<i>impeded.</i>
<i>Functional structure is...</i>	<i>enhanced.</i>	<i>likely impeded.</i>	<i>impeded.</i>

Figure 8. The effect of cognitive constraints on blending.

The cognitive constraints observed in this study could therefore be described in terms of a hierarchical relation. Morphological lexicalization implies the full impact of the efficiency of regular morphological operations, which leaves little space for blend formation. Lexical neighborhood restrains blending to varying degrees, and schema transfer effects may even fuel blend formation. This relation is demonstrated in figure 9.

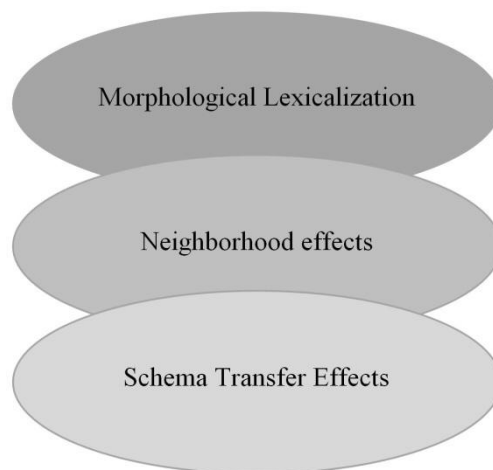


Figure 9. Impact on the potential for blending.

Furthermore, the data shows that blends are not necessarily more creatively used than other word formation processes, but they are instead elaborated applying a different spectrum of cognitive and grammatical mechanisms. The cognitive processes involved with blending tend however to be rather expensive in terms of cognitive economy. The corpus queries showed that the schema transfer effects incurred a small cost in that a comparatively high frequency of both types and tokens were found. The cost of forming blends with a lexical item in the process of morphological lexicalization were however considerable, which was seen in the poor corpus output. Lexical neighborhood effects also implied a cognitive effort, but, judging from the prototypical character of the blends, attractive properties such as wordplay and iconic fusion were considered worth the price in some cases. In similar future studies it should therefore be possible to calculate the cognitive cost balancing the frequency of blends with the density of type-specific features of the retrieved instances.

As regards word class distribution, the blends in the data displayed variation of usage that is not traceable to the structure of decontextualized instances. Therefore it cannot be taken for granted that structural properties of blends are sufficiently informative as to categorial delimitations and analyses of characteristics. Blending is thus a significantly pragmatic phenomenon, which explains many of the difficulties involved with structural analyses.

Blending has often been described as a marginal phenomenon in linguistic accounts. However, from a psycholinguistic perspective it is all but peripheral. Instead, the observations of this investigation is suggestive of Tomasello's (2003) findings that the basic mechanisms underpinning language is deeply rooted in socio-pragmatic motivation and "pattern-finding", or schematization (Tomasello, 2003, p. 4; see also

Taylor, 2002, and Langacker, 2008). Conventionalized schemas, or grammatical rules, governing morphological processes and designation of word class have been observed to influence blending in significant ways. At the same time, the profusion of innovative forms in the data and the pragmatic application in context attest to their impact as to linguistic creativity. Blending testifies therefore to the processes that are at the core of language development (cf. Cannon, 1986).

While children “(re-)construct the abstract grammatical dimensions” through years of exposure to instantiations of a language (Tomasello, 2003, p. 283), mature language users continue to explore this creative process in blending. The results of the investigation show that this happens on lexical, syntactical and cognitive levels. Blending can therefore be described as a manifestation of the very processes that drive human language; it is socio-pragmatically driven, it explores the boundaries of conventionalized patterns, and it eventually produces new (morphological) schemas. As such, it has therefore the capacity to reveal fundamental aspects of linguistic communication.

6. Conclusion

The object of this thesis has been to explore the underlying mechanisms of blending. In order to do so 94 derivatives of three matrix words and 2126 contextualized tokens have been studied in detail. The results have shown that there is a dynamic interplay between blending and other morphological operations; what linguistic strategy is chosen depends largely on cognitive efficiency. Also, the word class distribution of blends in the corpus mirrors their well attested structural complexity. Both these findings point to an important conclusion; blends are a pragmatic, or usage-driven, phenomenon exploiting the resources of morphological operations. This duality explains the dynamic semantics, the complex encoding, and the functional application of structure commonly observed in this word formation process.

The cognitive constraints and patterns of word class distribution found in this study point to central aspects of blend formation. However, due to the limited scope of the data, further studies examining a larger set of matrix words are needed to verify the generalizability of these assumptions. Also, a diachronic perspective highlighting the process of morphological lexicalization could have provided valuable information as to its impact of blending. In addition, the notion of segment markedness was another aspect that could not be included for reasons of space, which meant that neighborhood effects could only be estimated quantitatively. Besides these shortcomings, the findings of the investigation indicate that cognitive approaches to blending are a fruitful domain. Not only is this word formation process central to issues of language change, but studies in this field may also have implications for the understanding of the development of more rule-based processes.

References

- Algeo, John (1977). Blends, a structural and systemic view. *American Speech* 52, 47–64.
- Alm-Arvius, C. (2012). *Comprehensive Semantics. Lexicon, Grammar, Text Cognition & the World*. Unpublished manuscript. Department of English, Stockholm University.
- Barnhart, R.K. & Steinmetz, S. (1988). *Chambers Dictionary of Etymology*. London: Chambers Harrap Publishers Ltd.
- Bauer, L. (1983). *English Word-formation*. Cambridge: Cambridge University Press.
- Bauer, L. (2006). Compounds and Minor Word-formation Types. In B. Aarts & A. McMahon (eds.) *The Handbook of English Linguistics*. Oxford: Blackwell Publishing Ltd.
- Beliaeva, N. (2014). A study of English Blends: From structure to meaning and back again. *Word Structure*. Volume 7, Issue 1, Page 29-54.
- Blocker, J.S., Fahey, D.M., & Tyrell, I.R. (2003). *Alcohol and Temperance in Modern History: An International Encyclopedia, Volym 1*. Santa Barbara, California: ABC-CLIO Inc.
- Bryant, M. M. (1974). Blends Are Increasing. *American Speech*, (3/4). 163. *Cambridge Advanced Learner's Dictionary*. (3rd ed.) (2008). Cambridge: Cambridge University Press.
- Cannon, G. (1986). Blends in English Word Formation. *Linguistics* 24, 724-753
- Cannon, G. (1989). Abbreviations and Acronyms in English Word-Formation. *American Speech*, (2). 99-127.
- Carroll, L. (1993). *Alice in Wonderland and Through the Looking Glass*. Hertfordshire: Wordsworth Editions Limited.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, D. (2011). *Internet Linguistics. A student Guide*. Oxon, UK: Routledge.
- Dressler, W. U. (2000). Extragrammatical vs. marginal morphology. In U. Doleschal & A. M. Thornton (eds.) *Extragrammatical and marginal morphology*. Munich: Lincom Europa. 1–10.
- Fabregas, A. & Scalise, S. (2012). *Morphology: From Data to Theories*. Oxford: Oxford University Press.
- Fandrych, I. (2008). Pagad, Chillax and Jozi: A Multi-Level Approach to Acronyms, Blends, and Clippings. *Nawa: Journal Of Language & Communication*, 2(2), 71-88.
- Fauconnier, G. & Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Geeraerts, D. (Ed.). (2006). *Cognitive Linguistics: Basic Readings*. New York: Mouton de Gruyter.
- Gries, S. T. (2004). Shouldn't it be breakfunch? A quantitative analysis of blend structure in English. *Linguistics*, 42(3), 639-667.
- Gries, S. T. (2006). Cognitive determinants of subtractive word formation: A corpus-based perspective. *Cognitive Linguistics*, 17(4), 535-558.
- Harley, T. (2008). *The Psychology of Language. From Data to Theory*. 3rd ed. New York: Psychology Press.
- Hay, J. (2001) Lexical frequency in morphology: Is everything relative? *Linguistics*, 39 (4), 1041–70.
- Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge: MIT Press.
- Jakobson, R (1966). Closing statement: Linguistics and Poetics. In T. A. Sebeok (ed.) *Style in language*. Cambridge, Mass.: M.I.T. Press.

- Ji, H., Gagné, C.L. & Spalding T.L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, 65, 406–430.
- Kelly, M.H. (1998). To "brunch" or to "brench": some aspects of blend structure. *Linguistics*, 36(3), 579-590.
- Kemmer, S. (2003). Schemas and lexical blends. In H. Cuypkens, T. Berg, R. Dirven, & K.-U. Panther (eds.), *Motivation in language: From case grammar to cognitive linguistics. A Festschrift for Gunter Radden*. (pp. 69–97). Amsterdam and Philadelphia: Benjamins.
- Labov, W. (1973). The boundaries of words and their meaning. In C.-J. N- Bailey & R.W. Shuy (eds.). *New ways of Analyzing Variation in English*. Washington DC: Georgetown University Press, p. 340-373.
- Langacker, R.W. (2008). *Cognitive Grammar. A Basic Introduction*. New York: Oxford University Press.
- Lehrer, A. (1996). Identifying and Interpreting Blends: An Experimental Approach. *Cognitive Linguistics*, 7(4), 359-390.
- Lehrer, A. (2007). Blendalicious. In J. Munat (Ed.). *Lexical Creativity, Texts and Contexts*. (p. 115-133). Amsterdam, NLD: John Benjamins Publishing Company. Retrieved from <http://www.ebrary.com>
- Libben, G., Gibson, M., Yeo Bom Yoon, Y. B. & Sandrab, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, 84, 50-64.
- López Rúa, P. (2004). The Categorical Continuum of English Blends. *English Studies*, 85(1), 63-76.
- McEnery, T. & Gabrielatos, C. (2006). English Corpus Linguistics. In B. Aarts & A. McMahon (eds.) *The Handbook of English Linguistics*. Oxford: Blackwell Publishing Ltd.
- McMahon, A.M.S. (1994). *Understanding Language Change*. Cambridge: Cambridge University Press.
- Plag, I. (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
- Plag, I. (2006). Productivity. In B. Aarts & A. McMahon (eds.) *The Handbook of English Linguistics*. Oxford: Blackwell Publishing Ltd.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.
- Ronneberger-Sibold, E. (2006). Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica*, 40(1/2), 155-181.
- Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology*. 42A, 529-567.
- Sandra, D. (2009). Psycholinguistics. In D. Sandra, J.-O. Östman & J. Verschueren (eds.) *Cognition and Pragmatics*. Amsterdam: John Benjamins Publishing Company.
- Schmid, H.-J. (2011). *English morphology and word-formation: An introduction*. (2nd ed.) Berlin: Erich Schmidt Verlag GmbH & Co.
- Sereno, J.A. & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*. Vol.25, No.4. 425-437.
- Spencer, A. (1996). *Phonology*. Oxford: Blackwell Publishing Ltd.
- Taylor, J.R. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.

- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. New York: Harvard University Press.
- Ungerer, F. & Schmid, H.-J. (2006). *An Introduction to Cognitive Linguistics*. 2nd ed. Harlow, UK: Pearson Education Ltd.
- Yavaş, M. (2006). *Applied English Phonology*. Malden: Blackwell Publishing Ltd.

Appendix A

List of low frequency blends based on the matrix word *republican*.

Two instances:

repiglican
repukelican
repooblican

Hapaxes:

rapepublican
talipublican
demepublican.
foxpublican
frankenpublican
weepublican
talibpublican
reichpublican
appublican
regrumplican
repuklican
rumpbulican
repucklican
rebloodlican
reflublican
repoopblican
reprobatelican
democrublican
demolican
rape-ublican

Appendix B

Example of concordance list marked for adjectival and nominal use (A and N respectively).

1	N	box that way &; can fast forward when there is a	lieberal	on it, I know what they are going to say &; see n
2	A	n run by a Private company since the Jeff Kennett	lieberal	Party privatised certain parts of the Victorian p
3	A	us indignation. # Patriot: This is typical of the	lieberal	apologists. We of the so-called left totally cond
4	A	s pick up the tab. # But the best example of this	lieberal	hypocrisy was when Frazer &; Howard in 1980 set u
5	A	iticisms or scrutiny of Abbott is intolerable for	lieberal	party stooges. For them balance means you have to
6	A	ou must be joking. 2UE very often sounds like the	lieberal	Party mouthpiece vying with the other 2 something
7	A	asing private staff that run them. More prisons =	lieberal	governments. On the current matter of the NON SAL
8	A	ite really does smacks of delusional panic in the	lieberal	camp. # I wonder why his gay daughter is not sing
9	A	ing to find the Canadian patent/copyright to that	lieberal	method, but grudgingly realizing this is one thin
10	A	ely on and hence taint themselves with the old BS	lieberal	way of doing business, hence fall into the same t
11	A	ment!!! Yup! Kim Justine Il is the 21st centuries	lieberal	John Turner! We all remember how THAT worked out!
12	A	ed to leeching of the taxpayer!!!! No wonder the	lieberal	party is broke!!! # Not to dissuade your fantasy
13	A	e any sense. # At least it's a better system than	lieberal	and Dipper promoted government confiscation of mo
14	A	rong, I am joining with the NDP to help flush the	lieberal	party out of existence in BC. The lieberals had n
15	A	e we think is best able to extricate us from this	lieberal	sickness we are mired in. # From Wikipedia: " Lea
16	A	Why? # I'm more aligned to the NDP than anything	lieberal	. Not for anything they've done really, more likel
17	A	ime at least) Our two main parties, being NDP and	lieberal	, (sorry.. I can't help myself with the Lib spelli
18	A	Biased, definitely! so badly biased to all things	lieberal	and dipper that your news and most current events
19	A	just preaching their party line that as usual is	lieberal	good Conservative bad.Complaining to the CRTC wou
20	A	08-05 17:28 # Good move Jim kick out those BC	lieberal	(Conservatives) and put in some new Conservatives
21	A	e -- that dog don't hunt as they say just same ol	lieberal	fear and shmeat that most have grown tired off an
22	A	ook forward, when we can look back to 12 years of	lieberal	rule from 1993 to 2006. Really, it's not that lon
23	N	ill be bounced by the voters, in part because the	lieberal	's inability to know right from wrong &; think the
24	A	e in proximity. I was going to puncture some more	lieberal	balloons of bullshit today, like how the NDP turn
25	N	John Martin, decided he would rather run as a BC	lieberal	than sit on the sidelines whining about how corru
26	A	s that apparently didn't exist when so many other	lieberal	caucus or staff members had to hit the bricks, li
27	A	e one Paterson hired not any of the old socred &;	lieberal	politicians. # The lieberals have done some unfor
28	A	to accept money from Schreiber, but pales next to	lieberal	adscams etc -- only war room ass-kickers, that br
29	A	ing things, but we can not afford to have another	lieberal	government ruining this province! Eric is as corr
30	A	s very best to help her. Why! are the MSM and the	lieberal	party using this event a " Harper basting " Shame
31	N	m isn't called Brummistan for noting, either. So,	lieberal	, I think you dun thunk too much, Though this Rich
32	A	is the bedrock upon which the left stands on. The	lieberal	posters here display this in subtle but telling w
33	A	e school board is! Gee just look at the NEA 99.9%	lieberal	but their membership (aka the teachers are probab
34	A	rship (aka the teachers are probably about 55-60%	lieberal	) THUS not an accurate representation! Just look a
35	A	erica politicians, yes even republicans, are more	lieberal	than their constituency. Now just go to a field w
36	A	yone expect? LOOK if the next election is between	lieberal	closet democrat Romney or socialist Obama I WILL
37	N	to ID..... # Anonymous # Let me guess. You're a	lieberal	, right? Where in his statement did you see a thre
38	N	y this on anyone else;;; this is ALL on oBOYma,,	lieberal	's.. # It's even tougher when you subjected to lar
39	A	ngs even more, and it says, " I don't believe the	lieberal	media. ";) # I just now followed the links. Those
40	A	or at whom? # Randall # Me??? Send my kids to the	lieberal	ivory tower indoctrination centers? Better they b
41	A	ree posts in any comments thread without seeing "	lieberal	" or " Rethuglican " or " Obamination ". # But fo

Stockholms universitet
106 91 Stockholm
Telefon: 08-16 20 00
www.su.se



Stockholms
universitet