THE METAFORMAL SYSTEM: COMPLETING THE THEORY OF LANGUAGE

Christopher Michael Langan

ABSTRACT: The standard theory of languages has two levels, one centering on the study, teaching, and application of natural languages, and the other on formal languages and formal systems as applied throughout the mathematical and empirical sciences, in analytic philosophy, and for computer programming, software engineering, artificial intelligence, and related technologies. On both of these levels, standard language theory is *dualistic*, defining languages in isolation from their domains of discourse and treating attributes in isolation from their objective instances while omitting important properties and functions ordinarily provided or executed by language users, automata, or physical systems on which they appear to supervene. This decoupling of languages from their universes, and from necessary linguistic functions such as *display, processing, interpretation,* and *communication*, has profound epistemological bearing, limiting scientific knowledge by precluding the linguistic formulation of any verifiable comprehensive description of reality. This paper proposes that in addition to the two existing levels of standard language theory involving natural and formal languages and systems, the theory of language be recognized to possess a third "metaformal" level on which languages and their universes are "wrapped" in a uniquely structured, totally self-contained metalanguage, the *Metaformal System*, which restores missing linguistic functionality while using a supertautological intelligibility criterion to generically couple languages with their universes on a fundamental level of shared structure and dynamics, thereby restoring the potential for a verifiable comprehensive and fully connected understanding of the reality we share.

KEYWORDS: Language Theory, Generative Grammar, Metaformal, Formal Language, Formal System, Metalanguage, CTMU

THE NATURE OF LANGUAGE

Language, which is widely regarded as an artificial construction of human beings, is conventionally defined in a variety of ways associated with various perspectives on philosophy, psychology, sociology, and pedagogy, often with an emphasis on culture and politics, and with various technical prerogatives having mostly to do with the study and employment of automata. This has led to an emphasis on linguistic variety and contextual specificity, with variations cancelling out rather than summing to a meaningful overall characterization. Thus, the "bare bones" definition of language omits any reference to content, application, or structure, reducing it to an abstract set of "strings" or linear sequences of sounds or symbols devoid of meaning.

All existing definitions of language – even those which delve deeply into its structure, production, genesis, and evolution, or have to do with its formalization or computational application - sell it short by restricting its definition and ignoring its ubiquity. In reality, language is not dependent on the vagaries of human culture and psychology, the preferences of academic specialists, or the capacity, functionality, or architecture of computational machinery. In fact, language is the most general and necessary abstract structure in mathematics, the protean abstract structure of which all other structures are instances by virtue of their effability and mathematical descriptions alone. Its generality is reflected in the fact that it can be equally well described as an *object, attribute, process,* or *operational medium*, and in fact as all of these in seamless combination.

From a technical standpoint, one of the most important features of language is that its study is necessarily self-referential. That is, any attempt to define, describe, analyze, or construct language uses language itself to do so. Self-referentiality is a powerful and underexplored feature of language, one which has puzzled and fascinated mankind for centuries. But at the juncture where language meets cognition, it is also a source of logical and epistemological complications. Due in part to these complications, language has met with many apparent failures, ranging from piles of self-contradictory and seemingly unproductive philosophical verbiage to the undefinability and undecidability problems of formal language-systems.

There are many who seem to believe that this is the end of the story, and that language-using creatures such as we will never understand or agree on the nature of the reality we inhabit. But fortunately for science, philosophy, and the intellectual destiny of mankind, complete surrender to epistemological relativism turns out to be premature.

THE STRATIFIED THEORY OF LANGUAGE

The standard theory of language recognizes two main classes of language, *natural* and *formal*, each occupying its own compartment. These compartments may be respectively described as:

(1) The mostly academic study of human language acquisition and usage and the structure and grammar of various natural languages. A *natural language* is any human language, whether spoken, written, or manually signed, that appears to have arisen spontaneously and evolved naturally, through repetition with variation, usually along with the native culture of a specific population. Natural languages are used for routine self-expression and communication, public speaking, literary production, governance, commerce, and other purposes to which ease and freedom of expression and understanding are important.

(2) The (mostly academic and industrial) study and application of formal languages and systems and computational languages. Formal languages are often just natural languages that have undergone some degree of formalization, usually with convenient simplification and technical standardization. They are used wherever clarity, precision, and lack of ambiguity are required. Applications include mathematics, the empirical sciences, computer science, engineering, and other fields where exact technical specifications must be expressed and communicated with maximum precision and minimal semantic ambiguity.

Just as it would be hard to overstate the importance of natural language to (e.g.) human psychology, social organization, and psychosocial evolution, it would be hard to overstate the importance of formal language to science and technology. The mathematical sciences overwhelmingly rely on formal systems, especially "proof systems", to achieve their results. The empirical sciences are equally reliant on formal systems, which are theoretically coupled with observation and experimentation as prescribed by the Scientific Method. As for computer science and the information technology that supports modern civilization, they clearly rely as heavily on language as they do on computers.

If we regard natural and formal languages as the first two levels of modern language theory, it is possible to identify a third. Whereas the first two levels are dualistic, taking languages in isolation from their universes, the third is monic and self-dual, structurally fusing language and universe into a single coherent identity on the highest possible level of discourse. Featuring an intrinsic definition of language which is analogous to intrinsic geometry in its self-containment and external independence, it consists entirely of identities, or coherent self-dual language-universe couplings, and the operators and operations which generate and act on them.

This logico-metaphysical level of language theory is designed to formulate and address fundamental metaphysical and metalogical questions regarding a comprehensive and therefore totally self-contained reality with respect to which nothing deeper or more extensive exists, regarding which no external references are possible, and in which literally everything comes down to the self-reference of a self-defined language-universe complex that has no background and to which nothing intelligible is external. This is the level on which language merges with universe, and on which signs and symbols merge with their associated objects, interpretations, and interpreters. Or in more pedestrian terms:

(3) The level on which language can be productively employed in the conclusive identification and analysis of reality in part and in whole, i.e., on which language and universe are generically coupled in terms of common structure and intelligibility. It is on this level that the true and ultimate nature of reality - that which we can know of it strictly on the basis of our ability to sense it and draw logical inferences about it - can be fruitfully explored.

Where the deceptive dualistic isolation of language from universe obscures their true relationship and precludes certainty, conclusive identification is impossible. But as physical reality can in fact be clearly identified through direct observation and logical deduction, without which neither science nor human experience could exist, and as deductive logic is a *formal language* inherent in cognition, this assumed linguistic incapacity is unfounded and illusory. Hence, the third level of language theory herein proposed already has an irrefutable basis.

Let us more closely examine these levels of linguistic theorization.

LEVEL 1: NATURAL LANGUAGES

A *natural language* is conventionally understood as the system and method of spoken or written expression and communication among human beings to whose various communities, nations, and cultures it is particular. Natural languages "evolve naturally" in real-world communities of language users, often in coupling with their cultures and conventions.

Natural languages consist of the structured and conventional use of *words*, i.e., meaningful elements of speech or writing that (where not used alone) are organized to form *sentences*. A sentence is a complete sequence of words typically containing a *subject* and a *predicate* which assigns some attribute, often involving an action, to the subject, thus constituting an attributive statement. A sentence includes a main clause which can stand alone but may have one or more subordinate clauses attached, and can take the form of a question, command, or exclamation without losing attributive functionality.

A given natural language $L = (\Sigma, \Gamma, S_\Sigma)$ can be specified in terms of (1) a finite *alphabet* $\Sigma = (s_1, s_2, \ldots, s_n)$ of uninterpreted *terminal symbols* s_i that can be linearly combined to form written words and expressions and are often associated with sounds that can be spoken in the same order, along with any transitory nonterminal symbols; (2) a *grammar* or set of structural and transformational rules Γ which determine which *strings* or sequences of alphabetic symbols and words are "well-formed", i.e., which strings are potentially meaningful expressions of the language, and (3) the set S_Σ of strings determined by Γ. Whereas expressions contain only terminal symbols of Σ, Γ may use any nonterminal symbols in Σ as means to its ends, inserting and then replacing them with other symbols in order to produce or "derive" terminal expressions as the final output of derivation trees orthogonal to the expressions themselves in the ordering of their derivational steps. In this case, the alphabet Σ may be divided into two sets N and T containing nonterminal and terminal symbols respectively: $\Sigma = (N, T)$.

There are two possible ways to construct a string or expression of L. One can write the expression as a linear sequence of terminal symbols and words by using a set of orthographic rules according to which one symbol or word follows another by concatenation or composition. Alternatively, one can rely on rules of production or generation consisting of syntactic forms and rules of substitution. Production works orthogonally to strings themselves, deriving a string by replacing nonterminal symbols with cumulative refinements along its entire length. Generative grammar begins with the universal start symbol of L, a compact way of representing L as a complete expressive potential, and progressively replaces nonterminal symbols with other nonterminal symbols and finally with the terminal symbols of the final expression.

The grammar of a natural language can be complex and unpredictable, and is typically discovered through intensive empirical investigation; until it is "formalized" by exhaustively enumerating its rules, the language generated by the grammar cannot be formalized, and the formal definition $L = (\Sigma, \Gamma, S_\Sigma)$ remains useless for reconstructing the language. The formalization criteria described and illustrated above are due to Zellig Harris, who first described generative grammar as currently understood (1951), and his student Noam Chomsky, who subsequently named and explicated the concept. (1956)

Natural language is fundamentally conceptual; its words represent concepts or mental abstractions. The words and signs attached to the concepts are more or less arbitrary; they can vary as long as the concepts themselves retain their integrity. However, there is some amount of controversy regarding the relationship between language structure and cognition. Theories of this relationship range from *universalism* (cognition determines language), through mutual independence, all the way to *strong linguistic determinism* (language determines cognition), with intermediate degrees of dependency considered.

Note that while this account of natural language refers to the representation of *concepts*, nowhere does it refer to the *objective content, meaning,* or *interpretation* of its expressions. Yet even though one of the primary functions of language is to represent objective facts about the real world - facts which naturally reflect its structure - the natural language concept is defined in complete abstraction from the external universe, which is commonly treated as though it were completely "independent" of language and cognition. This putative mutual independence of language and universe – the way $L = (\Sigma, \Gamma, S_\Sigma)$ omits the symbol U for "universe", thus isolating L from its domain of discourse - amounts to a pernicious form of dualism which abstractly, artificially, and wrongly separates the elements of an inseparable pair.

In short, the capacity of language to accurately represent the external universe is far too reliable and extensive for this unequivocal kind of dualism to hold. It is not that we are always successful in predicting factual observations from language alone; rather, it is *the very possibility of bringing language into conformance with reality* (and vice versa) that dualism cannot explain. If dualism were justified, and language and reality were truly separate and independent, then there would be no basis for bringing them together, and neither science nor any form of human experience would be possible. The above definition, $L = (\Sigma, \Gamma, S_\Sigma)$, is therefore misleading in the way that its parentheses, like impermeable walls, isolate human cognitive language from the universe it so effectively but improbably represents.

Some readers will recognize that this is simply a formal generalization of the oft-cited "unreasonable effectiveness of mathematics in the natural sciences" to nonmathematical languages and reality at large. It is also a statement to the effect that any language which truly supports the identification of its universe must everywhere couple to that universe through classical 2-valued logic, which provides the 1|0 distinction in terms of which U and its contents may be distinguished. This coupling is an untapped well that goes very deep indeed.

LEVEL 2: FORMAL LANGUAGES AND SYSTEMS

Any natural language can be formalized to the extent that its alphabet is known and its grammar has been formalized, i.e., fully determined by empirical investigation and exhaustively enumerated. That is, a formal language is a natural language $L = (\Sigma, \Gamma, S_\Sigma)$ for which the elements of the signature and the rules of grammar are exhaustively enumerated.

A *formal language* L usually comes implicitly packaged with a *metalanguage* L' which supports the description and analysis of its structure, functionality, and application. It also comes with predicate and propositional logic, which it requires for its intelligibility and functionality. While logic stands alone, being formulated independently and invariantly with respect to the other languages to which it is applied, it is itself a formal language conforming to the same general description. In addition, L must incorporate such abstract qualitative and quantitative concepts and relationships as may be necessary to specifically identify its content.

In contrast to the often-complex grammar of a natural language, the grammar of an artificial formal language like propositional logic (PL) is simple by design. Its properties are not empirically discovered but chosen for precise expression and analysis. In particular, PL is designed to analyze and describe the relationships among conjunctions, disjunctions, conditionals, and negations *(and, or, if-then, if and only if, not)* linking arbitrary sentences (sentential variables) to each other, i.e., to precisely model key properties of the logical operators and connectives naturally inherent in human cognition so that statements can be reliably and consistently formulated and reliably classified as *true* or *false*.

Any sufficiently expressive formal language L with an easily recognized (or at least computable) grammar can be incorporated in a *formal system* T which adjoins to L a *deductive apparatus*, or a set of axioms and rules of inference designed for the systematic investigation of some specific universe U (domain of discourse or field of knowledge) U. Data from U are plugged into T, and conclusions are deduced in T regarding U.

A formal system T can be defined as T = (L, A, I), where L is a formal language, A is a set of axioms, and I is a set of rules of inference. The deductive apparatus {A,I} can be viewed as an expanded version of the grammar Γof L, somewhat blurring the distinction between a formal system and a formal language. However, L and T can often be distinguished by two related facts: (1) the syntactic forms of L usually contain nonterminal symbols representing word-classes or "parts of speech", whereas the axioms of T usually do not; and (2) the production of expressions by the rules of Γ usually involves nonterminals, whereas the derivation of consequences from axioms using the rules of inference of T usually does not.

Predicate logic, including sentential or propositional logic, is implicit in any formal system T. This is because logic must distribute over every part and every element of T no matter how T is partitioned and quantized. It takes predicate logic to form attributions, and propositional logic to identify them from their logical complements; without logic, there can be no discernibility. Also implicit in any formal system T is a metalanguage defining and relating its components.

The second level of the standard theory of languages, studied mainly in mathematical logic and computer science, includes the study and application of these formal languages and systems, the means of proving theorems from their axioms, and their valid interpretations or *models* in their respective domains of discourse U.

THE LIMITATIONS OF FORMAL SYSTEMS

Despite their inadequacies, formal systems are powerful tools of (partial) understanding. However, a formal language omits its array, its processors, its universe, its models, the capacity to determine how it evolves (telesis), and even the capacity to perform its own linguistic operations such as *read, write,* and *erase,* let alone more advanced operations like the expression and communication of meaning. It is thus completely unable to stand on its own, instead relying on language users and sometimes on computers. It is therefore unable to model (share structure with) a perfectly self-contained universe. To do that, it must assimilate that on which it implicitly depends.

Note that formal systems are limited in the same way as language in general; just as for $L = (\Sigma, \Gamma, S_\Sigma)$, the formal system $T = (L, A, I)$ contains no symbol U representing the external universe. Again, the parentheses effectively wall off L and T, effectively isolating them and offering no assurance that T can be successfully interpreted in any given U. Thus, a given U need not preserve the structure of T, provide T with a model, or conform to T as semantic content; until it has been fully explored and put into exhaustive correspondence with T, the actual relationship between T and U is uncertain. In short, a formal system comes with what philosophers have called the *problem of induction*.

This model-theoretic formulation of the problem of induction implies that as long as theories of science, philosophy, and mathematics are regarded as formal systems (or something even less reliable), they can afford no certainty with respect to the real universe, and therefore cannot yield a verifiable understanding of reality. Indeed, the scientific investigation of reality is limited in just this way. Beholden as it is to the scientific method, the sciences have been limited to establishing tentative correspondences between formal systems and sets of physical observations (or subjective conceptualizations), and thus doomed to non-validation and epistemological inadequacy.

While it can be hard to establish that a given language has a model in a given universe, especially when the correspondence must be monitored and maintained "in real time" as U evolves, there are cases in which a model-theoretic correspondence can be taken for granted. For example, we can trivially define U on L; given the internal consistency of L, we can simply define U to be any universe that conforms to L. This is how mathematical theories and their underlying conceptual languages are often treated; their domains of discourse are simply defined as any universe U, i.e., any set of objects, relations, and functions, in which L can be modeled by virtue of common structure, i.e., by virtue of the nonempty structural intersect of L and U.

On the other hand, the structure of U must be known in order to define L on U, and U may not be completely knowable (e.g., in the empirical sciences, nothing about the observable universe can be taken for granted). In this case, it is natural to ask whether there exists any formal language L whose compatibility with any given U can be taken for granted.

Provided that U and its contents are intelligible or identifiable, the answer is yes. Any form of identification, conceptual, perceptual or otherwise, amounts to distinguishing something from its logical complement. This requires the formal language of logic, especially 2-valued propositional logic, which functions in effect as an "identification language" of U. By virtue of the very possibility of apprehending a given universe – by virtue of its *intelligibility* – it must everywhere incorporate the formal system "logic". Conversely, any supposed part or aspect of U devoid of logical structure cannot be identified, and is therefore not recognized as a part or aspect of U.

In short, logic evidently functions as a kind of "identity language" for U – something like the set of criteria through which a computer identifies input, or dually, an attribute that U displays in order to be identified. But it cannot do this alone, for whatever logic distinguishes from its logical complement must exhibit nonlogical properties – size, color, duration, and so on - of which it is an instance. So it appears that we must add these logical and nonlogical ingredients together to get a functional identity language L. But we still have a problem, as propositional logic can only identify full attributions. There must also be something capable of coupling the attributes in L to the objects in U, and while predicate logic comes with means of quantification, it has no way of producing or selecting such nonlogical attributes in order to couple them with the contents of U.

Throughout the empirical sciences (and most other sciences and the humanities as well), it is roundly denied that the physical universe and its internal content, or states, arise in coupling with L. Instead, it is held that physical reality evolves according to its own internal causal processes, mechanisms, and laws, despite the obvious fact that "laws" are formal (abstract, qualitative) entities. As usually conceived, causation is dualistic; it requires that reality evolve independently of any language, with causal functions that are metrically parameterized exclusively by the objects and independent medium on which they act. But this leads to a contradiction: it implicitly presupposes, once again, that U is already coupled with an intrinsic self-identity language that provides it with attributes and values from which to form the states which comprise the input and output of causal functions. In short, dualistic, language-independent physical causation implies that U is already inseparably coupled with its very own identity-language, which just happens to be the same as ours.

Finally, we realize that we might as well cut to the chase: Somehow, an intelligible universe, one that we can actually identify, must already be coupled with an identity-language. The simple fact is that where any degree of intelligibility is given, structure must already be shared between mind and external reality; mind must be equipped with a cognitive identity language L through which U can be recognized, and in a complementary way, U must have the capacity to display content recognizable to L. Unless these dual requirements are filled, the mind has no way to reach out and acquire external reality, and external reality has no way to send itself into the mind. And that, of course, means that the question is really just this: What are the implications of the fact that mind and reality already share structure which includes logic and various nonlogical attributes?

This is not the only issue in overcoming linguistic dualism. Standard dualistic languages lack a number of self-containment properties. They lack closure with respect to not just modeling and the acquisition of content (*reference*), but such functions as *read, write, production, processing, display, communication,* and the selection or execution of any constructive or restrictive process or mapping that might generate expressions or restrict the exhaustive linguistic potential represented by $\{S_\Sigma\}$, the exhaustive set of all grammatically well-formed strings. In order to determine and perform these operations, languages require ingredients that are not contained in their standard dualistic definitions. Dualistic language theory excludes everything that makes language telic or dynamic, waving it off to some language-user or programmed automaton, or supervening it on some random or determinate physical process that supposedly has no need of it.

THE METAFORMAL SYSTEM: AN INFORMAL INTRODUCTION

Due to limitations of time and space, this account is brief and non-justificative. We cannot hope to give anything but the most summary and superficial account of the Metaformal System in this paper.

First, it might be wondered why this paper is entitled "*The* Metaformal System" rather than "Metaformal Systems". This is because the only metaformal system known to exist is that which corresponds to reality as we know it, and as far as we know, there are no other realities to be considered (this system has been known for the last several decades as the *Cognitive Theoretic Model of the Universe* or CTMU; Langan, 2002). There are other essentially metaphysical constructs, including the Everett "Many Worlds" interpretation of quantum mechanics (1956) and the Tegmark "Mathematical Universe Hypothesis", which speculatively refer to other possible realities (2008). But such acts of reference connect these "other" possible realities to this one, implying that the resulting web of interconnections constitutes one overarching "ultimate reality" after all. It is this ultimate reality that the Metaformal System represents.

The Metaformal System is an "extended formal system" which is perfectly self-contained, i.e., which has all of the properties and carries all of the ingredients required for its existence and functionality, including those just mentioned (*read, write, reference, interpretation, production, processing, display, communication, potentiation, determination,* and so on). In short, all of these functions are executed by objects of M (syntactors, telors) using the grammar of M.

Even though the Metaformal System can hold an arbitrary number of formal systems in its framework, it can be represented in or described by a formal system of its own. This is how it is represented even here, on paper and on monitor screens controlled by patterns of bits in electronic computers. However, despite the fact that the Metaformal System can be *described by* a formal system, i.e., embedded in the formal system as descriptive content, it turns the tables, mapping into its own structure the very formal system which describes it along with any language-user reading the formal description. In relating itself to its own ingredients, M places itself in its own universe by a self-dual grammatical inclusion mapping called *involution*, and is thus a metalanguage of not only the formal systems in question, but itself.

The Metaformal System can be thought of as a kind of "closed wrapper" for formal systems. The wrapper is a supertautological metalanguage (Langan, 2002, 2017) in which formal systems are "located" or assigned a place in M as referential content on the basis of shared structure. Specifically, they are located in L_M, the formal or linguistic aspect of the Metaformal System M, the *identity-language* of U_M which contains all of the abstract properties which suffice to identify not only U_M, but all that can be identified in the ontic identity $M = L_M | U_M$ including all of its internal subidentities or discernible attribute|value couplings. Some of these identities and subidentities are *telors*, which are defined in terms of functional capacity and include human beings; M thus absorbs human language-users and the linguistic functions they perform, firmly embedding them in the metaformal structure of reality.

The **Metaformalist Program**, as it has previously been called (Langan, 2018), requires that language function as an *identity*, specifically the identity of a universe. That the universe we inhabit is intelligible to us means that our internal cognitive-perceptual language couples (interfaces or intersects) with it, allowing us to identify it by direct perception. It is in this sense that we function as "identification operators", using our cognitive language as an *identity language* to identify the universe and its contents including ourselves. Because identification operators use the syntax (formal structure) of an identity language to identify their operands, they are called *syntactic operators* or *syntactors*. Telors are structurally complex syntactors which can "factorize telesis" or actualize ontic potential, and have sufficient complexity to consciously generate internal representations of themselves and their relationships with the external environment.

The grammatical language|universe coupling effected by M is *self-dual* in the sense that language and universe are just different aspects of shared structure in which they everywhere intersect. This implies that they are the source and target, or (dually) the target and source, of a mutual *dualization* mapping (involution) which generically "models" one in the other. The coincidence of language, universe, and model makes the entire three-way coincidence not just self-dual, but *trialic*. (Triality was adumbrated by the American logician Charles Sanders Peirce (1976), whose semiotic approach to logic characterized signs in terms of three classes: *signifier, object,* and *interpretant.*) Because language is one of its coinciding aspects (no matter what its particulars may be and regardless of the degree to which they are specified in advance), this relationship has linguistic structure. Specifically, it is a new kind of "language" called an *intrinsic language*.

An intrinsic language can be understood as a trialic (three-aspect) self-identification language which does triple duty as language, universe, and model while functioning as a global primary syntactor, of which internally localized secondary and tertiary syntactors are *points of involution* (targets of the grammatical involution mapping). The primary syntactor, which is also the Primary Telor or universal start symbol of M-grammar, spans the entire involution mapping from L_M to U_M, while tertiary syntactors are confined to the U_M (terminal, physical) end and include fermionic (mutually exclusionary) pointlike particles. Secondary syntactors occupy the terminal region of the mapping to variable depth, where - in their dual role as secondary telors - they fill causal deficits due to the underdetermination of tertiary syntactors by the Primary Telor, which is invariant with respect to them and of which they are images by μ-morphic involution (the grammatical involution mapping is also called the *μ-morphism*).

At the trialic limit where the semiotic representation relationship between intension and extension contracts to form self-dual intension|extension identities as required by systemic self-duality, information "about" the state and evolution of the system on all scales equates to the evolution of the system itself; in evolving, the system is "informing itself" of its own evolution. Primary, secondary, and tertiary syntactic operators generate three levels of information. Primary information is associated with the primary syntactor and best described as *teleology*; its function is to provide for the requirements of self-identification on all scales down to that of tertiary syntactors. (Where generic utility is that which enhances identity or "selfhood", a utilitarian definition of teleology as the maximization of generic utility equates it to the Self-reinforcement of the identity of reality or Primary Telor.) (2003) Tertiary syntactors are the quantum objects of which secondary syntactors are composed and in which their extended phenotypes materially intersect. Tertiary information, which is about the terminal "surface" level of evolution and equates to it when attributes and values are merged as terminal identities by full L_M-U_M coupling, is the only level studied in physics and other empirical (terminal, surface-level) sciences. Secondary information fills causal deficits which owe to the underdetermination of tertiary states and interactions by primary information, this being the local freedom of the system and its inhabitants.

THE METAFORMAL SYSTEM, A BIT LESS INFORMALLY

In much the same way as a formal language, the Metaformal System can be formally expressed as:

$$M = L_{INT} = Ls|Lo = (\Sigma=\{N,T\}, \Gamma_{MU}, S_\Sigma)$$

M AS A WHOLE

Moving from left to right, "$M = L_{INT}$" says that the Metaformal System is an *intrinsic language*. Complete (ontic) self-containment is required of both M and L_{INT}, implying structural equivalence. It turns out that this intrinsic language doubles as a pair of complementary *semilanguages*, Ls and Lo, that are in fact dual to each other insofar as they comprise the intensional and extensional aspects of M.

Notice that "$M = (\Sigma=\{N,T\}, \Gamma_{MU}, S_\Sigma)$" is very close to the way in which formal languages are defined, with a signature, a grammar, and a set of linear "strings". The meaning is almost as simple: The fundamental objects of M are active signs called *telors* (N) and *syntactors* (T), and its grammar Γ_{MU} is a self-identification operation which, just as in a standard language, "identifies" strings by writing or generatively producing them. In other words, M evolves by "identifying itself".

The generative derivation and orthographic dimensions of self-identification are orthogonal, but not entirely independent; N and T must be continuously coupled so that telors can use Γ_{MU} to acquire terminal resources from T for grammatical transformation in N.

The metaformal intrinsic language M is an ingredient of its own syntax on all orders of reference. That is, M is self-referential and logically idempotent (Langan, 2017).

SIGNATURE

In the signature $\Sigma=\{N,T\}$, the subsets N and T of Σ are sets of nonterminal and terminal identity operators respectively; N = {telors} and T = {syntactors}. These two kinds of "active sign" are distinguished by functionality as well as location; because production ("derivation") occurs in N while concatenation ("write") occurs in T, telors can generatively produce entire strings in a spacelike or "nonlocal" fashion, while non-teloric (tertiary) syntactors, being restricted by the locality, timelike linearity, and continuity of T, cannot.

The signs in Σ are all identity operators which use M as an identity language to identify the world; hence, all are syntactors. There are three levels of syntactor: primary, secondary, and tertiary. The primary and tertiary levels respectively correspond to the universe as a whole, and the smallest and most elementary objects therein; the secondary level is intermediate.

N and T, which can be respectively likened to the internal and external and/or the mental and physical aspects of reality - and at the risk of considerable oversimplification, to a programmable automaton and its display screen - have some amount of overlap; every telor is a syntactor, but not every syntactor is a telor. Primary and secondary syntactors are telors, but tertiary syntactors are not.

Telors can exist in both N and T; tertiary syntactors can exist only in T. Because grammar operates mainly in N, only telors can fully exploit its power, generating spacelike derivation trees which output terminal strings in parallel, across time. But in order to meaningfully do so, they must have sufficient complexity in T to model their own relationships with the external environment.

GRAMMAR AND SEMIMODELS

Γ_{MU} is a self-dual distribution mapping, the *μ-morphism*, that is inwardly involutive and consists of spacelike metatemporal parse trees which produce terminal states and trajectories in T, but outwardly evolutive with respect to the timelike linear trajectories thus produced. μ-morphism is associated with a rescaling operation called *conspansion*, which occurs in a *conspansive manifold* which takes the form of a *syntactor bundle*, a self-dual analogue of the topological construct known as a "fiber bundle".

Conspansion is a coupling of spacelike dual processes, *inner expansion* or *d-ectomorphism*, a one-to-many mapping which outwardly distributes compact identities (states, points) to potentials, and *collapse* or *d-endomorphism*, a one-to-many dual mapping which inwardly distributes attributes to states or points. (The *d-* stands for "distributive".) The self-dual conspansion morphism corresponds to the *conspansive semimodel* of M, which models derivation and is associated with the nonterminal semilanguage Ls; it evolves discretely and in just the way that reality appears to evolve in the quantum-theoretic picture.

The terminal analogue and dual model of the conspansive semimodel, the *linear-ectomorphic* (or *l-ectomorphic) semimodel* of M associated with Lo, is based on dual (inward and outward) conspansive gradients and yields a very different picture of M, namely the classical one that we see and hear, in which terminal objects follow linear trajectories through "spacetime", which – together with its physical content - amounts to the surface structure of the conspansive medium. The superposition of these dual semimodels reflect the fact that reality evolves by self-dually "modeling itself" in two complementary ways, only one of which we actually see when we look outward, but in the other of which true Γ_{MU} dynamics reside.

SEMILANGUAGES

Ls and Lo are dual semilanguages into which L_{INT} is "factorized" by Γ_{MU}. Ls and Lo correspond to N and T respectively.

Ls and Lo correspond to the linguistic structure superimposed by M on the sets N and T respectively, providing the elements of N and T (telors and syntactors) with enough structure and organization to productively interact and fill their roles as identification operators. Respectively corresponding to the intensional and extensional aspects of M, they remain in superposition as Γ_{MU} locks them together in order to form the self-dual identities that comprise the terminal states of M.

Simplistically, Ls exists "inside" secondary telors of N, while Lo consists of the external states and trajectories or state-transitions of their constituent tertiary syntactors. As Ls transforms and is dually transformed by Lo through the action of Γ_{MU}, Ls is conspansively "involuted" in metatime, and Lo "evolves" linearly in time along dual conspansive gradients.

STRINGS

$S_\Sigma \supset T$, the set of "terminal strings" of M, consists of the external states and trajectories of syntactors. S_Σ, associated with the semilanguage Lo, occupies the "surface structure" of M.

THE RELATIONSHIP BETWEEN LEVELS 1, 2, AND 3

Level 1: Natural Languages

In a natural language, expressions are produced by writing or speaking linear sequences of words (or signing in a sequence of gestures). Empirical investigation of natural language grammars reveals that beneath their *surface structure,* they have *deep structure* which can be explicated in terms of word-classes ("nonterminal symbols") and *phrase markers* (parse trees, derivation trees). These phrase markers have derivational structure orthogonal to the linear expressions on which they terminate, suggesting that expressions are actually generated along two axes.

Natural language grammars are sometimes classified in terms of the least powerful acceptor, or language-recognizing automaton, that can recognize the languages they generate. As defined in the Chomsky Hierarchy, type 0 grammars generate unrestricted languages requiring a universal computer (Turing machine) with unlimited memory; type 1 grammars generate context-sensitive languages requiring a linear-bounded automaton with memory proportional to word length; type 2 grammars generate context-free languages requiring a pushdown automaton with a memory stack in which a fixed number of elements are available at any point; and type 3 grammars generate regular languages requiring a finite deterministic automaton with no memory. (Chomsky, 1956)

Note that where languages are defined on the grammars that generate them, and grammars are classified in terms of automata, languages themselves are implicitly classified in terms of automata. This language-automaton association has been in force since the mid-20th century, and it has only reinforced the unavoidable theoretical coupling of natural language with human cognition.

Level 2: Formal Languages and Systems

In a formal system T, the production grammar of the formal base language L refines nonterminal symbols to terminal symbols, progressively applying rules of substitution or transformation to a universal nonterminal "start symbol" representing the entirety or "identity" of L, terminating on fully formed expressions which contain nothing but terminal sentences consisting entirely of terminal symbols and words. To this, a formal system T adds a "terminal grammar" consisting of terminal forms called *axioms* which provide a general description of some universe or domain of discourse U, along with rules of inference.

The axioms – which are terminal with respect to grammatical production and constituent symbols, but initial with respect to the axiomatic description of their universe – comprise a boundary between the nonterminal and terminal realms. They come in two forms: one, often used in mathematics, continues derivation in the same metatemporal direction as production, while the other is typified by continuously iterative "laws of nature" or "laws of physics" which evaluate to numbers, accepting the values of prior states as input and yielding as output subsequent states which appear to arise "orthographically", in the temporal *write* direction of symbolic and lexical concatenation or composition. The axioms may be applied to known data or assumptions regarding U in order to derive consequences via rules of inference, usually in the hope and expectation of obtaining useful information – descriptions, predictions, or explanations - about U.

Because the formal system is merely a terminal extension of its formal language L, all of the above characterizations of language apply to it. Either no universe or model is explicitly included in the formal descriptions of L and T (including the case where U=L or U=T), or there is a particular application in mind, as when a formal language is structured for computer programming or control of a particular automated system. In either case, the association between language and automata persists, but all the more strongly due to the increased emphasis on mathematical logic and computational applications.

LEVEL 3: THE METAFORMAL SYSTEM

The third level of language theory is that of the Metaformal System. It resembles any other language in that its grammar brings to bear both rules of production and orthographic rules of symbolic concatenation and lexical composition along orthogonal *generation* and *write* axes which map to a pair of orthogonal axes in the conspansive medium of U_M. This medium, the *conspansive manifold*, evolves through a stratified, generative, self-dual rescaling operation associated with metaformal grammar Γ_{MU}, namely *conspansion*, which generates a logico-linguistic "boundary algebra" in which syntactic distinctions are mapped to closed static surfaces and hypersurfaces, a general concept associated with such logicians and mathematicians as Venn, Peirce, Frege, and most recently George Spencer-Brown (1969). It incorporates formal languages and systems, nonterminal forms and terminal strings, and is therefore linguistic in structure and operation. Nevertheless, it includes unique explanations for gravity, cosmic expansion, quantum phenomena, consciousness, and even spirituality. It is in every important way linguistic in structure and operation.

As for the trademark linkage of language and information processing, the Metaformal System M takes it to its logical conclusion by uniting language with the largest and most powerful "information processing system" conceivable, namely, reality at large. Over the last several years, there has been considerable buzz regarding the so-called "simulation hypothesis" or "simulation theory", which centers on the possibility that despite its propensity to fool us all, reality is an artificial simulation running on a vast cosmic computer, astral automaton, or some other kind of host. Some versions of this hypothesis seem reminiscent of science fiction movies; others involve ideas emerging from ongoing research into "virtual reality". But the most logically well-founded and therefore credible version of this idea is the unique logico-linguistic approach of the CTMU, with references as far back as the late 1980's. (1989) The CTMU shows that no matter how the issue is slanted, reality can indeed be categorized as a "self-simulation" with a nonterminal grammar and a terminal display space of which we are all the occupants, the spectators, and to a limited extent, the programmers.

Self-simulation aside, there is a certain logical inevitability to the proposed extension. Not only are fields like physics, cosmology, biology, consciousness, AI, and philosophy beating their heads against conceptual brick walls when they try to answer metaphysical questions for which standard theoretical languages are fundamentally inadequate – itself a powerful argument for rummaging around the language department for missing conceptual infrastructure – but the questions themselves are often garbled. If I might be permitted to employ the third level of language herein proposed, modern scientists and philosophers are telors in a state of total denial who - while blithely using their own technical languages to pose profound questions and seek deep answers – can often do nothing but look about them at the terminal domain $T \subset \Sigma$, peer at it with all their might, and fail to recognize even the linguistic surface structure of the associated M-semilanguage Lo! Little wonder that nothing ever gives.

But still, there are some very good, very logical reasons to think that any attempt to avoid the proposed extension would be futile. Let's look at it from a metalogical vantage. A metatheory is a theory about theories. Theories of science and mathematics are formal systems; thus, the capacity in which M incorporates formal languages and systems, including those employed throughout the mathematical and empirical sciences, is that of a metatheory. An axiom of a metatheory that deals with its object-theories and/or their axioms is called a **meta-axiom**, and likewise for any axiom referring autologically to itself.

The CTMU metaformal system can be brought to rest on a single "closure meta-axiom", the *Analytic Reality Closure* Principle (ARC), which is also called the "CTMU Identification (or Intelligibility) Axiom". It can be expressed as the CTMU "master-equation" (which, interestingly, can be regarded as a logical generalization of Einstein's Equation, with a medium on one side and its content on the other):

$$\mathbf{reality}_{\mathbf{INT}} =^{*} \mathbf{reality}_{\mathbf{EXT}}$$

This axiomatic equation, which for various reasons cannot be mistaken for a mere assumption, expresses the innate self-duality of M. Notice that in addition to being a global meta-axiom, it has on its left side an attribute or *descriptor*, and on its right side a collection of *operators* that are μ-morphic Self-images of a single *Global Self-Operator*. Thus, without too much blame, it might fairly be abbreviated as "the GOD Equation", where "GOD" is acronymic for "Global Operator-Descriptor".

This axiomatic equation translates as follows:

> "*The intension of reality* (the global property L_M = reality$_{INT}$) *is dually equivalent to the extension of reality* (the physical universe U_M = reality$_{EXT}$),"

where *dual equivalence* describes the situation in which dual entities coincide as aspects of a self-dual identity which "glues them together" using mutually distributed structure in which they everywhere intersect.

ARC can also be rendered as the "CTMU Reality Principle":

> "Reality contains all and only that which is real."

This can be reformulated as a self-contained proof by contradiction:

> "If there were anything *outside* reality on which reality were in any way dependent – e.g., for origination, causation, support, maintenance, or anything else - then the entire dependency relationship would be real and therefore *inside* reality. It follows that reality is self-contained with respect to physical and metaphysical functions like origination, causation, support, maintenance, etc."

This proof by contradiction immediately invalidates the premise of external relevance, implying the ontic closure of M = reality. It also means that the level of "reality" we are considering is the deepest level possible, namely the CTMU level. This is because "reality is ontically self-contained" doubles as a definition of reality under which it is idempotent under inclusion (or if one likes, "self-inclusive", with all the bells and whistles required to support this property; Langan, 2017).

The intelligibility criterion of the Metaformal System M overrides seemingly preclusive metalogical results such as the Tarski Undefinability Theorem and the Godel Undecidability Theorem. In a nutshell, arguments based on such results, which usually apply explicitly to standard formal systems, can only undermine their own intelligibility (and that of the formal systems to which they are attached) by undermining M. Similarly, insofar as intelligibility relies on classical 2-valued propositional logic, objections based on the existence of alternative logical systems, e.g. constructive logic, are without force. Other logics may be relevant within bounded contexts, but criticisms which rely on them can only compromise their own intelligibility by opposing M.

With any luck, this has brought the reader at least a little closer to a very important realization about the CTMU Metaformal System: as asserted in previous papers, it bids fair to be recognized as an integral part of the universal foundational language of science and philosophy, and thus an integral part of the scaffolding for the long-awaited bridge between science and theology, physics and metaphysics, and physical and mental reality. Indeed, it should now be easy to see why it would be very difficult indeed to erect any other kind of scaffolding for this purpose.

chris@ctmu.org

REFERENCES

Chomsky, N. (1956). *Three models for the description of language*. IRE Transactions on Information Theory (2): 113–124. Retrieved from https://ieeexplore.ieee.org/document/1056813/

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/616323.pdf

Everett, H. (1956). *Theory of the universal wavefunction*. (Doctoral dissertation). Princeton University. Retrieved from https://www-tc.pbs.org/wgbh/nova/manyworlds/pdf/dissertation.pdf

Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.

Langan, C. M. (1989) The Resolution of Newcomb's Paradox. Retrieved from http://megasociety.org/noesis/44/newcomb.htm

Langan, C. M. (2002). The Cognitive-Theoretic Model of the Universe: A New Kind of Reality Theory. *Progress in Complexity, Information, and Design*, *1*, 2-3. Retrieved from http://main.megafoundation.org/Langan_CTMU_092902.pdf

Langan, C. M. (2003). Cheating the millennium: The mounting explanatory debts of scientific naturalism. In W. A. Dembski (Ed.), *Uncommon dissent: Intellectuals who find Darwin unconvincing*, (pp. 233-262). Wilmington, Del.: ISI Books. Retrieved from https://www.scribd.com/document/24820585/Cheating-the-Millennium-The-Mounting-Explanatory-Debts-of-Scientific-Naturalism

Langan, C. M. (2017). An introduction to mathematical metaphysics. *Cosmos and History: The Journal of Natural and Social Philosophy*, *13*, 2, 313-330. Retrieved from https://cosmosandhistory.org/index.php/journal/article/viewFile/618/1040

Langan, C. M. (2018). Metareligion as the human singularity. *Cosmos and History: The Journal of Natural and Social Philosophy*, *14*, 1, 321-332. Retrieved from https://www.cosmosandhistory.org/index.php/journal/article/viewFile/694/1157

Peirce, C. S. (1976). *The new elements of mathematics.* Carolyn Eisele (Ed.). Atlantic Highlands, N.J.: Humanities Press.

Spencer-Brown, G. (1969). *Laws of form*. London: Allen & Unwin.

Tegmark, M. (2008). The mathematical universe. *Foundations of Physics*. *38* (2), 101–150. Retrieved from https://arxiv.org/pdf/0704.0646.pdf