mathematics

MDPI

Article

Schrödinger's Cat Still Lingers on the Mat

Román Orús [1] and Juan Uriagereka [2,*]

[1] Quantum Department, Donostia International Physics Center, 20018 San Sebastian, Spain; roman.orus@dipc.org
[2] Linguistics Department, University of Maryland, College Park, MD 20742, USA
* Correspondence: juan@umd.edu

Abstract

The faculty of language can be modeled as an algebraic cycle between a *lexicon*—an Abelian Hilbert space $\mathcal{H}_2$ spanned by the diagonal matrices of a discrete group we call the Chomsky-Pauli group (G_{CP})—and a derivational *workspace*, realized as a Fock-lifted sector $\mathcal{H}_4$ in which syntactic chains arise as tensorized, separable superpositions. Results are three enumerable constraints that jointly identify the grammatically well-formed chains: (i) a rank-1 compressibility condition on chain contexts (satisfied by 96 of the possible matrix sums), (ii) a spectral condition on four-term chain aggregates (satisfied by 32 of those 96, each yielding spectral radius 2φ (twice the *golden ratio*)), and (iii) a hermiticity condition that restricts the observable subclass to interface-interpretable configurations. The spectral radius emerges as the algebraic fingerprint of X′-theoretic phrasal topology, connecting the lexical group-theoretic structure to a Fibonacci-like growth pattern of language phrase structure. The mathematical claims are finite and exhaustively checkable, and their linguistic interpretation constitutes a program of empirical hypotheses.

Keywords: language faculty; syntax; lexicon; Hilbert space; chain; superposition; Chomsky-Pauli group; Fibonacci topology; spectral radius

MSC: 37M99

Academic Editor: Alvaro Anton-Sancho

Received: 25 May 2026
Revised: 8 September 2026
Accepted: 22 October 2026
Published: 4 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction: The Paradigmatic-Syntagmatic-Paradigmatic Cycle

Since Pāṇini developed a rule-based description of Sanskrit [1], the faculty of language has been conceptualized as "a dictionary and a grammar" [2]. In Medieval [3] and 17th century Europe [4], grammarians were intrigued by a "universal grammar" for elements "of thought" discerned in all languages [5]. This aligns with the rationalist tradition emphasizing reason as a source of knowledge, arguing for innate ideas.

1.1. Conceptual Preamble

In the mid-20th century, Noam Chomsky synthesized those classical influences, responding to his structuralist predecessors, to propose that the human mind contains an innate language faculty [6] composed of: (i) a lexicon $\boldsymbol{L}$, understood as a list of items and their associated features, and (ii) a computational system $\boldsymbol{C_{HL}}$, specific to human language, of finite, explicit procedures to generate and interpret an infinite number of sentences [7]. This generative framework shifted the focus from descriptive grammar to the underlying knowledge that allows native speakers to produce and interpret sentences [8]. One, thus,

attempts to characterize how a human mind selects elements from $\boldsymbol{L}$ into $\boldsymbol{C_{HL}}$—a process long acknowledged in psycholinguistics [9], still scantily understood—somehow restoring these "words" back into $\boldsymbol{L}$ (without loss) after use.

The formal study of language and the empirical mapping of animal brains have advanced along distinct intellectual trajectories, although pioneers like C.R. Gallistel have emphasized the need to unify cognitive representations with neurobiological mechanisms [10]. In his foundations of structural linguistics [11], Ferdinand de Saussure conceptualized a dual framework for language: the *paradigmatic axis* of geometrically abstract options for association; and the *syntagmatic axis* of ultimately linear, "rule-bound", combinations. When viewed through Charles Sanders Peirce's logical distinction between timeless *types* and concrete *tokens* [12], this formal tension goes well beyond human speech, arguably revealing a universal requirement for a sufficiently complex animal mind.

Non-human animals, in point of fact, also present systems (of thought if not of shared information among conspecifics) exhibiting some of these tenets (in particular, a repository of quasi-permanent, and in that sense paradigmatic, concepts that novel, effectively syntagmatic thoughts take stock from and contribute to). So, we must calibrate any proposal about language with viable hypotheses about animal minds more generally [10].

In sum, the crux of the formal mental machinery we will be pursuing here implements long-term storage of types alongside a dynamic usage basin, as follows:

Paradigmatically stored types can be invoked for computation (1a)

Syntagmatic operation tokenizes stored types (1b)

Memorably invoked tokens may update type presuppositions (1c)

The abstract categorial types an animal needs to navigate reality arguably exist in vast, high-dimensional, paradigms (1a). In contrast, when an animal actively perceives or acts, as driven by attention within its cognitive workspace, said paradigms manage to get compressed into syntagmatic tokens for real-time computation (1b). But some of those, if informative and memorable to the animal, may return in an updated guise to their storage state (1c). As we see it, this Paradigmatic-Syntagmatic-Paradigmatic (PSP) architecture amounts to a dynamic cycle within an animal's circadian processes.

1.2. A Minimalist Program

While structural concepts along those lines were admittedly birthed to explain structural properties of human language, we see the PSP cycle as a foundational prerequisite for any form of computation in an animal's nervous system. To examine the structural relationship between distributed types and compressed tokens, we present our approach within the tenets of the Minimalist Program, fully understanding that this is not the only available theory (see [13] for recent perspective). Still, for reasons that will become apparent as we present our case, that framework is ideal for our approach.

The first explicit formulation of Minimalist syntax, via the operation *merge*, was Collins and Stabler 2016 [14]. Chomsky and associates [15] have argued how early versions of the program are descriptively too strong (allowing hard to justify mergers) and not sufficiently explanatory. They also formulate derivational workspaces as a natural way to constrain the computational system in terms of a revamped MERGE (caps intentional). Marcolli, Chomsky, and Berwick [16] (MCB) provide one specifically algebraic way of defining a workspace, in the process predicting some empirically ascertained limitations. Considerations along these lines pertain to what may be psychologically real or, ultimately, how neural dynamics manifest themselves through the discrete algebra any computational system presumes. In language, the goal of the PSP cycle is to select

paradigmatic types into tokenization processes that allow said types to assemble as binary sets, which are monotonically built (within a workspace) into novel thoughts/expressions—eventually to return to memory consolidation as types (with new nuances as relevant).

Recursion has been mentioned as the "Basic Property" separating human language from other forms of animal cognition [15]. We must challenge this below because it creates a theoretical impasse when confronted with evidence from biopsychology [10] and philosophy of mind [17]. If foraging, navigation, dwelling construction, or complex social interactions in many species are *systematic* and *productive* in strict computational terms, this is enough for relevant animals to hold predictable and creative implications in novel behaviors. Such animals may, thus, pass the tests to be in possession of a "language of thought" (LoT), in the sense of Fodor [18]. This LoT must have some syntax with enough structure to carry inferences, (equivalent to) a predicate calculus. This is the foundational reason to take it to require the ability to implement Peircean types and tokens.

Of course, if *merge* in the early sense [14] is the general structure-building process required for that basic LoT, an operation of that sort cannot, then, be considered a human-specific mutation. At the same time, it is hard to deny a saltation that emerged during the Paleolithic and led to human conditions beyond animal inferences: generalized quantification, exact numerosity, exploded vocabularies, elliptical communication, and incremental cultures that pass information beyond individual experience [19]. This is where an evolutionary shift is sensible: from something like *merge* to the narrower MERGE.

Minimalism has explored those two different versions of structure-building:

Merge: A basic operation combining (only) two separate items from some lexical array (extracted from lexicon $\boldsymbol{L}$) to build structure (e.g., merging arbitrary lexical items like *86* and *47* to yield {86, 47}) (2a)

MERGE: Presuming resource strictures within a Workspace $\boldsymbol{WS}$, a mapping from $\boldsymbol{WS_n}$ to $\boldsymbol{WS_{n+1}}$ to assemble items from $\boldsymbol{L}$ into restricted structures (e.g., {{86, 47}, WS_i), where $\boldsymbol{WS_i}$ is an argument of the operation). (2b)

Unlike its *merge* predecessor [14], MERGE as in ([15,16]) is inherently constrained by minimalist information properties that hardly arise in logical terms but are essential to the specifics of an animal mind—as empirically established by experimental research. One may then argue that formal properties of human syntax emerge from resource limitations of an architectural sort [16]. In other words, general *merge* may be presumed for many species, while MERGE may, instead, be seen as a uniquely human achievement.

While that move still leaves the issue open as to what may have caused our evolutionary transition [19], it moves the onus from general recursion conditions to narrower ones associated to the MERGE mechanism instead. Reflecting on the latter process will be our focus, presuming that it may have evolved from more elementary conditions that are also, of course, quite relevant to the phenomenon as presently understood.

1.3. Formal Preliminaries

Assume a space of possibilities where signals and concepts "live", in which nearby points represent similar conditions (like sounds or meanings) and distance represents how different two realizations happen to be. The question is what determines where an object ends up in this space—the intuition being that mental types realized into tokens tend toward stable configurations for those types, much as physical systems dissipate energy toward equilibrium. Mathematically, the landscape can be represented by a function $H(\psi)$, whose input is a possible (e.g., linguistic) state, and its output a number meant to address how stable that state is. The output of this function $H(\psi)$ assigns a score.

Concepts can be taken to move in different directions within their landscape (some making them more localized, others more periodic, etc.). We can mathematically represent each such "direction" by an operator. If we feed the operator a "stop" consonant like [t], it yields a large value: [t] is very concentrated (into a virtual turbulence burst); feed the operator, instead, a sustained, diffuse, "open" vowel like [a] and a small value arises. Such a localization operator must also recognize how other answers can be messier: when confronting a diphthong like [ay], it must decompose into components. The operator's stable answers are called its *spectrum*; an *eigenstate* (own, intrinsic, state) is one for which a variable defining the state being measured has a determinate fixed *eigenvalue*. A diphthong [ay] is not itself an eigenstate, as it must be decomposed into eigenstates.

Next, consider matrices that summarize limiting sectors. Suppose we presuppose two independent questions-to-measure, each with two stable answers (generating four possibilities). In the case of linguistic concepts, whether phonological or more elaborate ones as discussed below, it is profitable to consider questions-to-measure that are orthogonal to one another, like "localization" or "continuancy". We take these to be mutually extraneous because one focuses on sharpness while the other does, instead, on sustainability: like rests and notes in a musical score. As an intuitive anchor (formalized in Section 2), we note that Euler's formula to translate trigonometry into algebra states that $e^{i\pi} = -1$—so we may take unit ± 1 to address the "sharpness" angle, while $\pm i$ can be taken to express, instead, its "sustainability" counterpart. This, then, yields four logical possibilities:

$$C_1 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}, -C_1 = \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}, C_2 = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, -C_2 = \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix} \quad (3)$$

Square matrices of that sort are the simplest kind. We give the initial C to said matrices as a mnemonic of the fact that the array in (3) reproduces Chomsky's foundational categorial distinctions [20]—although he used the features $\pm N$ $\pm V$ instead, arguing for conceptual orthogonality between such substantive features. We are instantiating that intuition in terms of the equivalences $\pm N = \pm 1$ and $\pm V = \pm i$, presuming vectorial interpretation and mapping the ensuing vectors into the matrix diagonals in (3). This move will allow us to construct a Hilbert space, whether its substantive consequence is phonological (for consonants, vowels, diphthongs), lexical (for all the "words" in lexicon $\boldsymbol{L}$), or of any kind. In a nutshell, the framework takes linguistic states to form a vector space, assuming its features are eigenvalues of corresponding operators, and that these are 2 × 2 complex matrices as in (3), where eigenstates carry a definite algebraic answer.

1.4. Comparison with Feature Theory

In classical feature theory [21], a phoneme or lexical item is described by answering a list of questions: Is this [+continuant], [−sonorant], [+nasal], [+plosive], etc.? But then conditions of "feature geometry" (as first systematically examined in [22], and see [23]) become challenging: Why are features like [+continuant] universal and foundational, while [+nasal] or [+plosive] are sporadic and ancillary? Our framework reverses the focus: Algebraic (structural) questions have features as (stable) answers. An eigenstate is a linguistic answer to that question that is intrinsic to the system, while most states (like diphthongs and other "intermediate" situations) occupy a region where multiple questions interact, with different measurements revealing various aspects of the ensuing structure. Mutatis mutandis, similar conditions arise in the larger context of paradigms within the entire lexicon $\boldsymbol{L}$, where matrix instantiations as in (3) correspond to lexical decisions:

Chomsky matrices in lexico-conceptual fashion:

$$C_1 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}_{Noun}, -C_1 = \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}_{Verb}, C_2 = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}_{Adjective}, -C_2 = \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}_{Adposition} \quad (4)$$

Mathematically, one such "landscape" is a rule assigning a score to every possible linguistic state, a Hamiltonian H:

$$\psi \mapsto H(\psi). \tag{5}$$

This Hamiltonian tells us which states (phonetic, semantic, pragmatic, etc.) are naturally stable, which unstable, and how much it takes to move from one state to another—by way of operators that correspond to structural measurements (like those considered for "localization" or "continuancy"). When H is expressed as a weighted sum of operators as in (6), Equation (5) is understood as the expectation value $\langle\psi|H|\psi\rangle$: the score assigned to state $|\psi\rangle$ by the Hamiltonian, which stems from those measurements:

$$H = \alpha L + \beta K + \cdots \tag{6}$$

where L measures localization, K measures continuancy, etc.—and the coefficients α, β indicate how strongly each property contributes to the Hamiltonian H, via weighted sums. For H asking simultaneously *How localized* and *How continuant?*, the eigenstates already have definite answers, like valleys in the landscape, which is why the system naturally settles into them. But the spectral object need not be stable, so H must identify such "in-between-valleys" as a superposition (where each ψ_i is an eigenstate):

$$\psi = c_1\psi_1 + c_2\psi_2 + \cdots \tag{7}$$

The matrices in (3) (with a lexical implementation as in (4), a corresponding application to phonology as relevant—or beyond as needed) are an organized way of encoding the questions asked to specify which observables define the landscape. A matrix is a rule that transforms a state before we inspect it in a particular way, specifying which aspect of the state is being examined. Thus, matrices are the formal embodiment of linguistic, mental, or ultimately physical questions that define a perspective. For a linguistic or mental state $|\psi\rangle$, containing all the relevant information, questions can be asked whether the state is localized at some edge, it is continuant, periodic, and so on—each such question corresponding to applying a different operator, which determines whatever distinction becomes focused in the measurement. If the square matrix is diagonal, like those above and many below, it has nonzero entries only along its diagonal, meaning it transforms each axis of the space independently (without mixing them). The state interrogated along the two fundamental directions of the Chomsky matrices implicitly codes a geometry by way of the algebraic object. Below we also consider some anti-diagonal matrices.

Now suppose we apply one of the Chomskys in (4) to vector (7), so that, for instance:

$$C_1|\psi\rangle = -i|\psi\rangle, \tag{8}$$

If $|\psi\rangle$ is itself an eigenstate of C_1, then (8) means that, under such an operation (from that perspective), the result "rotates". Each of the matrices applied to $|\psi\rangle$ yields different, but comparable, results, like seeing Rodin's *Thinker* from diverse angles: the statue does not change, only viewpoints do. Features are observations rather than "recipe ingredients". If $|\psi\rangle$ is not an eigenstate of (say) C_1, applying that matrix C_1, then, does not return a single stable answer, but the state decomposes into several eigenstates. That decomposition tells us how much of each limiting category is present. An "intermediate" category like a diphthong, for instance, decomposes into eigenstates defining its particular distinction.

In sum, the system introduces "feature geometry" (with some foundational, while others are states arising from algebraic combination) by taking features per se to be not primitive but *measurement outcomes*. The foundations are only stable landscape sectors, each represented by one of the fundamental matrices, $\pm C_1, \pm C_2$ (which occupy the four corners of a coordinate system). Natural language (and we believe animal thought more generally) goes beyond limiting cases, with items shifting states as used in different contexts. The Hamiltonian acts as path-governor: some routes are smooth, others require climbing over ridges, etc., predicting not only which states are stable but, also, which transitions thereof are easy, difficult, or effectively impossible. In this way, the same

mathematical object explains both categories and processes, thus upgrading the Hamiltonian from a "scoring function" to something with explanatory power.

2. Matrix Materials and Methods for Local Linguistic Considerations

Many current approaches to the human lexicon code token meanings in vectorial fashion, based on the distributional semantics idea that "a word is characterized by the company it keeps" [24]. Although J. R. Firth is credited with this phrasing, the underlying concept that *noscitur a sociis* ("it is known by its associates") is traditional. Wittgenstein made it his as "the meaning of a word is its use in the language" [25], and Zelig Harris's distributional analysis pushed the view that words appearing in the same context tend to have similar meanings [26]. The earliest precursors to modern word vectors emerged from information retrieval research, with vectors derived from co-occurrence counts within a text corpus [27], the resulting universe then being a high-dimensional vector space. In a special issue of the *Journal of Mathematical Psychology* on quantum cognition, in 2012, Bruza and Busemeyer proposed using the mathematical formalism of a Hilbert space as a framework to model the mental lexicon (in an attempt to understand word association and recall [28]). Our proposal too adopts a Hilbert space framework, less as a claim about quantum processes in the brain than as a mathematical tool for representing the structure of the lexicon—an interpretive stance we return to in Sections 4 and 5.

2.1. The Lexicon L and Lexical Arrays (A_L) Thereof: The First Two Axioms

As noted, the Minimalist Program presumes that some $\boldsymbol{C_{HL}}$ manipulates a multiset of token items, selected from the types in $\boldsymbol{L}$ into a Lexical Array $\boldsymbol{A_L}$, from which combinatorial operations can be invoked [7]. Even that "selection" is nuanced [29], as concepts in $\boldsymbol{L}$ coexist among several thousand for any speaker, ready to be summoned at will during the waking hours; and yet as they feature in a particular syntactic derivation, said paradigmatic entities manage to activate other so-called functional categories (of which only a handful exist). At that point, the system begins to act in syntagmatic fashion: with categories occupying definite roles and linearly specified slots in a spaciotemporal dimension. By the end of a sleep cycle, the syntagmatic frenzy dissipates back to $\boldsymbol{L}$, restoring the dormant harmony—albeit with concepts often getting informationally enriched.

Current research (summarized in [15]) suggests that the process just reviewed must invoke an attention-driven Workspace $\boldsymbol{WS}$, wherein the operation MERGE (of categories from $\boldsymbol{L}$ and (already assembled) phrases in $\boldsymbol{WS}$ can be stated [16]). This is to understand the nature of the presumed constructive aspects, but also, ultimately, the converse *disassembling* (without loss or depletion), which leads to an algebraic space of operation. One must, then, ask whether all these theoretical notions ($\boldsymbol{L}$, $\boldsymbol{A_L}$, $\boldsymbol{WS}$, MERGE), which are certainly meant as psychologically real, are also neurophysiologically factual.

From such a perspective, a lexicon in someone's mind can hardly be a static dictionary; it must be a dynamic, distributed network that links different kinds of information about concepts (with lexical import or beyond, as relevant [29]). From what can be understood already, when users engage a "word", they activate multiple areas of a highly multimodal system (including phonological, semantic, and syntactic information [30]). Although there is no clear understanding, yet, about how paradigmatic items in a human lexicon are stored or selected, similar sounding or lexically related words are somehow connected, as experienced in "tip-of-the-tongue" and priming effects, where processing one item speeds up the recognition of a related one [31]. The strength and nature of these connections is obviously shaped by experience, as the brain continuously updates the lexicon—not just with new words acquired through a lifetime, but also in terms of revised presupposed nuances in "old" words [32]. In general, it is quite hard to entirely forget words that were attained at some point, particularly if at an early age [33].

https://doi.org/10.3390/math14214037

Our proposal models lexicon $\boldsymbol{L}$ as an Abelian Hilbert space spanned by the 16 diagonal elements of a group we call the Chomsky-Pauli group $\boldsymbol{G_{CP}}$, as it includes the Pauli group of 2 × 2 matrices extending to others with "mixed" (real and imaginary) entries, the Chomsky matrices discussed in (4) in Section 1.3. The Pauli group is a well-known finite, discrete, group consisting of the four Paulis (including the identity $\boldsymbol{I}$) and their products with the phase factors ±1 and $\pm i$ [34]. These matrices generate the Lie algebra of the special unitary group **SU(2)** [35], per the idea that relates a geometric, potentially complex, object (Lie group manifold) to a simpler, linear algebraic object (Lie algebra vector space).

A Lie group is also a smooth manifold (like the set of all rotations in 3D space, the special orthogonal group **SO(3)**) [36]. It is a group because one can compose rotations, and a manifold because the rotations vary continuously with parameters (like rotation angle). Every group has an identity (for **SO(3)**, "no rotation"). The space around this identity is its *tangent space*, and if at the identity of the Lie group manifold, it is a flat vector space capturing all possible infinitesimal directions to move in (for **SO(3)**, such rotations around the x, y, and z axes form the basis vectors for this tangent space). The tangent space at the identity, combined with an operation (the *Lie bracket*) that measures group commutativity failure at the infinitesimal level, forms a (linear) Lie algebra [35] (Section 3.1).

In quantum theory, a vector in Hilbert space represents the state of a system whose symmetries are captured by a group of unitary operators acting on that space [37]. The Lie group represents said symmetries and is itself captured by a group of unitary operators acting on the Hilbert space (for every group element g, there is a corresponding unitary operator $U(g)$). The Lie algebra for the infinitesimal symmetries is represented by a set of Hermitian operators also acting on the Hilbert space: the generators of the transformations [35]. The exponential map connects the Lie algebra and the Lie group. For any X in Lie algebra, the corresponding Lie group element is given by e^{tX}, where t is a continuous parameter corresponding to the time-evolution operator $U(t)$: $U(t)=e^{-iHt}$, for H the Hamiltonian (an element of the Lie algebra) and $U(t)$ an element of the Lie group of unitary time-evolution operators [37]. The Paulis multiplied by i form a basis for this Lie algebra. By exponentiating a linear combination of the Lie algebra generators, any element of the continuous Lie group **SU(2)** is produced. Thus, the finite, discrete, Pauli group and its generating matrices tie to the continuous, infinite Lie group **SU(2)** [38].

While such considerations may be deemed impossible in relation to classical systems, let alone (very hot) brains, we take no position here on whether animal brains implement genuinely quantum processes—our framework is consistent with either interpretation and we defer that question to the Discussion section, Section 5.

Recall from Section 1.3 how Chomsky took lexical categories ("noun", "verb", etc.) to decompose into syntactic features $\pm\boldsymbol{N}$, $\pm\boldsymbol{V}$, which we are translating into *the entries sustaining the Pauli group*: $\pm\boldsymbol{N} = \pm\boldsymbol{1}$ and $\pm\boldsymbol{V} = \pm\boldsymbol{i}$. Presuming a vectorial interpretation as discussed and mapping the ensuing vectors into 2 × 2 matrix diagonals, we can represent grammatical categories and their interactions within what amounts to a lexical Hilbert space, with consequences for symmetries reviewed below. Thus, the two initial axioms:

Axiom 1 (fundamental assumption): $\boldsymbol{L}$ attributes are N = 1 and V = i. (9)

Axiom 2 (Chomsky matrices): Lexical categories are equivalent to the diagonal matrices $Noun = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$, $Verb = \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}$, $Adjective = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$, $Preposition = \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}$. (10)

The algebra of the Chomskys (and their mirror images, horizontally and vertically, totaling sixteen "mixed" matrices), when combined with the Paulis, forms a larger group of matrices: the $\boldsymbol{G_{CP}}$. This is a discrete algebraic structure whose elements are not just inspired by, but also properly include, all elements from the sixteen-matrix Pauli group.

It is as easy to construct a Hilbert space $\boldsymbol{\mathcal{H}_2}$ from the thirty-two 2 × 2 matrices of the $\boldsymbol{G_{CP}}$ as it is to do it from the Pauli group [34]: one considers the vector space those matrices span. In themselves, they need not be vectors of the Hilbert space in this context (they are operators acting on a more fundamental set of vectors). In quantum mechanics, the Paulis act on a 2-D complex Hilbert space $\mathbb{C}^2$, which represents possible qubit states [37]. Vectors of that space are two-component complex vectors; e.g., to represent spin-up (↑) and spin-down (↓) states of a particle, in Pauli's formulation. The inner product, to allow for measuring lengths and angles within the space, is defined as $\langle\psi|\phi\rangle$, with the core Paulis as linear operators acting on these state vectors. The approach for our $\boldsymbol{G_{CP}}$ is analogous, although the interpretation of $\boldsymbol{\mathcal{H}_2}$ and its basis vectors is extended for our "mixed" states (with real and complex entries). The ensuing 2D Hilbert space is still isomorphic to $\mathbb{C}^2$ (see Appendix A.1), although here, instead of "spin-up/down", the basis states represent possible values of syntactically relevant constructs, as discussed below.

2.2. Selection and Labeling: The Next Three Axioms

The matrices of the G_{CP} act as operators that perform actions on linguistic features, bundled into items in $\boldsymbol{L}$ as selected into $\boldsymbol{A}_{\text{L}}$. For example, matrix $\left[\begin{smallmatrix}1 & 0\\ 0 & -i\end{smallmatrix}\right]$ (representing nouns in $\boldsymbol{L}$, per Axiom 2) may act on itself via matrix multiplication (a self-scaling), per:

Axiom 3: Labeling in $\boldsymbol{A}_{\text{L}}$ arises from matrix multiplication (11)

In projecting from $\boldsymbol{L}$ to $\boldsymbol{WS}$, two conditions arise: (i) *Selecting* into the unassembled portion of $\boldsymbol{WS}$ that is being thought of as $\boldsymbol{A}_{\text{L}}$ (the array of paradigmatic types summoned from $\boldsymbol{L}$ for a given derivation); and (ii) Suitably *assembling* those into structured tokens that modify $\boldsymbol{WS}$ proper (recall (2b) in 1.1). While the assembling condition results in tokenized structures, the selection condition just invokes a given lexical type for use in the derivation. The labeling process that Axiom 3 instantiates ensures that the tokenized syntagmatic structure addresses its proper paradigmatic type.

By postulate, our labeling is multiplicative, involving a scaling of the argument (a matrix interpreted as a vector) by a matrix, then, taken as a linear operator. Thus, within $\boldsymbol{A}_{\text{L}}$ a matrix like $\left[\begin{smallmatrix}-1 & 0\\ 0 & i\end{smallmatrix}\right]$ (representing verbs in $\boldsymbol{L}$ per Axiom 2) may act on, say, matrix $\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]$, via matrix multiplication to yield the state $\left[\begin{smallmatrix}-1 & 0\\ 0 & -i\end{smallmatrix}\right]$; this, of course, needs to be interpreted, which we return to. Formally, it could equally be that a matrix like $\left[\begin{smallmatrix}1 & 0\\ 0 & -i\end{smallmatrix}\right]$ (representing nouns in $\boldsymbol{L}$ per Axiom 2) acts as a linear operator *on itself*, then understood as a vector, an instance of self-multiplication or squaring $\left[\begin{smallmatrix}1 & 0\\ 0 & -i\end{smallmatrix}\right]^2$, thus yielding the state $\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]$. Even if such initial, selection-driven operations were limited to the four Chomskys, four Pauli matrices (Pauli's σ_z, its negative, and the identity I and its negative $-I$, following the schema $\left[\begin{smallmatrix}\pm1 & 0\\ 0 & \pm1\end{smallmatrix}\right]$) need to be invoked among the results. These, in turn, can also serve as arguments of further multiplications, yielding 64 possible "selective" combinations.

So, all these matrix multiplications are sequentially applied within a single system (the 2 × 2 Hilbert space $\boldsymbol{\mathcal{H}_2}$), resulting in 2 × 2 matrices within the same space $\boldsymbol{\mathcal{H}_2}$ (corresponding to the group's closure property: applying an operation to an element within yields another element of the same group). An algorithmic foundation for the linguistic system can, then, be constructed from these multiplications in $\boldsymbol{A}_{\text{L}}$: by postulating a universal *Start* (or Σ) stage. Such an initial stage is dictated by a combination of formal and substantive conditions. Note, first, that within $\boldsymbol{WS}$, the minimalist MERGE operation is generally taken to combine term α from $\boldsymbol{L}$, selected into $\boldsymbol{A}_{\text{L}}$, with an *already assembled* phrase K [15]. Obviously, this is impossible in the very first state in any derivational thread starting from selecting two (bare) items from $\boldsymbol{L}$: there is no "already assembled phrase" as the derivation starts—a sort of "Eve" puzzle for any genesis! A formal way to address this is by invoking a valid self-MERGER for labeling, which amounts to a matrix squaring.

An elegant way to express this idea, then, amounts to presuming that:

Axiom 4: Projection from $\boldsymbol{A}_{\text{L}}$ into ***WS*** is anti-symmetrical. (12)

Again, we need both (i) selection and (ii) assembling within ***WS***, which is what creates the "Eve" puzzle: in the steady state of MERGE, a new selected item (normally called a "head") MERGES with an already assembled item (normally called a "projection"); but as it all starts, the only way a selected head can relate to anything assembled is if it self-scales—resulting in a self-assembly. This is consistent with taking the process of relating heads in $\boldsymbol{A}_{\text{L}}$ to projections suitable for ***WS*** as *anti-symmetrical*, which Axiom 4 stipulates. (A binary relation R on a set X has the anti-symmetry property if, for any two elements a, b in X, whenever a is related to b and b to a, then $a = b$.) Formally, the only operation available at the start of a derivation that takes an element of $\boldsymbol{A}_{\text{L}}$ as both operator and operand is matrix squaring, which is why Axiom 4 takes projection to be anti-symmetrical, for the initial step to be a self-application: by effectively "letting Eve self-replicate", at which point we can think of her clone as a projection that, subsequently, a newly selected Adam can asymmetrically relate to as relevant to conditions of distinctness.

But note how, by squaring all the Chomskys (corresponding, by Axiom 3, to all four major "parts of speech"), we obtain the very same projection: Pauli's σ_z matrix.

$$\text{For nouns: } \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{13a}$$

$$\text{For verbs: } \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{13b}$$

$$\text{For adjectives: } \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{13c}$$

$$\text{For prepositions: } \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{13d}$$

If this first algorithmic step is necessary whenever a syntactic derivational thread starts in ***WS*** (to project that structured assembly from the selected, yet unstructured, $\boldsymbol{A}_{\text{L}}$ by the squaring self-scaling), a *neutralized projection* arises. This is formally fine, but also a nightmare for an information system that must be semiotic among conspecifics—as that situation would be the equivalent of Paul Revere's famous dictum stupidly asserting "One if by Land, two *if by Land!*", which is totally non-informative. Identically, were all four parts of speech to project to the same matrix, the labeling system would be informationally vacuous: all categories would be indistinguishable at the projection level, defeating the purpose of the lexical array. To break the impasse, we thus postulate:

Axiom 5 (Anchoring): A derivational thread in ***WS*** start at σ_z construed as NP. (14)

There are two components in this Anchoring Postulate. Syntactically, it stipulates that the algorithmic derivation *starts* (at Σ) in Pauli's σ_z; semantically, that this (trivial) projection is interpreted as a noun phrase only. That means restricting the first step in ***WS*** derivational threads (projecting structure from $\boldsymbol{A}_{\text{L}}$) to (13a) to the exclusion of other options in (13). This amounts to a conventional anchoring of the derivation in nominal structure, consistent with the empirical observation across languages that nominal phrases constitute the most elementary notions and the universal building blocks of predication. The choice might follow from cognitive reasons—but we set that aside in this context.

2.3. The Jarret Graph

By combining, in such terms, the Chomskys with Pauli's σ_z, the Identity I, and their polar counterparts, we obtain the algorithmic pattern in Figure 1 to organize selection conditions in $\boldsymbol{A}_{\text{L}}$, where colors are a visual aid to separate "projections" (NP in **red**, PP in **blue**, VP in **green**, AP in **yellow**). The graph presumes the further labeling axiom in (15).

Axiom 6 (Labeling) *PRELIMINARY:* Phrase labels are matrix determinants. (15)

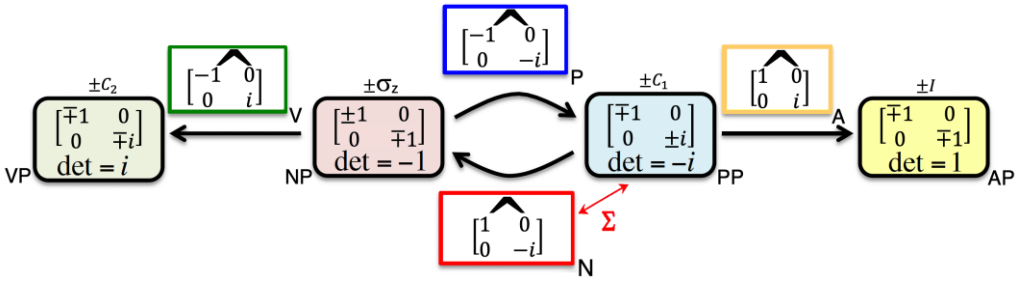


Figure 1. Graph for "thematic" relations in A_L (proposed by Michael Jarret); Hermitians boxed.

This amounts to linking the effectively Markovian dependencies [1], of the sort possible from selected versions of types in $\boldsymbol{L}$ (by the algebraic structure emerging from the axioms), to tokenized phrasal dependencies that emerge in $\boldsymbol{WS}$ under conditions discussed below. Note how, then, (i) the linear order of dependents is irrelevant (e.g., "adpositions" may surface as post- or pre-positions, as they do in various languages), and also (ii) "twin" projections exist for each lexical item, although what unifies them is always the matrix determinant (product of the eigenvalues, here diagonal items): $-\mathbf{1}$ for NP, $-\boldsymbol{i}$ for PP, $\boldsymbol{i}$ for VP, and $\mathbf{1}$ for AP). Observe how, when seen as linear operators, the four matrices (the Chomskys) are presented with a $\widehat{hat}$ (as $\hat{N}$, $\hat{P}$, $\hat{V}$, $\hat{A}$ operators, respectively).

The way to interpret the graph in Figure 1 is as summarizing the basic operations of, in effect, a kind of *categorial grammar* [39,40]. This is a first pass modeling of "thematic relations" for simple dependencies (of the sort describing how nouns or adjectives take PP complements, while verbs or prepositions take NP complements instead).

$$\text{For } \hat{N}\text{: } \begin{bmatrix} 1 & 0 \\ 0 & -\iota \end{bmatrix} \cdot \widehat{\begin{bmatrix} \pm 1 & 0 \\ 0 & \mp \iota \end{bmatrix}}_{-\iota = PP} = \begin{bmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{bmatrix}_{-1} = \text{NP} \tag{16a}$$

$$\text{For } \hat{V}\text{: } \widehat{\begin{bmatrix} -1 & 0 \\ 0 & \iota \end{bmatrix}} \cdot \begin{bmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{bmatrix}_{-1 = NP} = \begin{bmatrix} \mp 1 & 0 \\ 0 & \mp i \end{bmatrix}_{i} = \text{VP} \tag{16b}$$

$$\text{For } \hat{A}\text{: } \widehat{\begin{bmatrix} 1 & 0 \\ 0 & \iota \end{bmatrix}} \cdot \begin{bmatrix} \mp 1 & 0 \\ 0 & \pm i \end{bmatrix}_{-\iota = PP} = \begin{bmatrix} \mp 1 & 0 \\ 0 & \mp 1 \end{bmatrix}_{1} = \text{AP} \tag{16c}$$

$$\text{For } \hat{P}\text{: } \widehat{\begin{bmatrix} -1 & 0 \\ 0 & -\iota \end{bmatrix}} \cdot \begin{bmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{bmatrix}_{-1 = NP} = \begin{bmatrix} \mp 1 & 0 \\ 0 & \pm i \end{bmatrix}_{-i} = \text{PP} \tag{16d}$$

Nothing in (16) is semantically stipulated: it follows from the algebraic structure of our axioms and simple assumptions about (i) exhaustively deploying all matrices in the group, (ii) in a connected fashion, and (ii) presuming the Anchoring Postulate—thus covering the algebraic space the diagonal matrices in the $\boldsymbol{G_{CP}}$ permit in Chomsky fashion.

In sum, while in physics the Hilbert space from the Pauli group represents the quantum state of a physical system (a *qubit* in quantum information terms), here the $\boldsymbol{\mathcal{H}_2}$ from the $\boldsymbol{G_{CP}}$ represents the state of "syntactic stuff" within the $\boldsymbol{L}$ "landscape" as it gets selected into $\boldsymbol{A_L}$—a finite multiset of heads from $\boldsymbol{L}$ and subject to Jarret-graph constraints. While $\boldsymbol{A_L}$ summons "what's to be said" in any given derivation, it has no phrasal topology as such, which still needs to be invoked. Matrices here are operators tracking linguistic manipulations. Just as the Pauli group structure reflects symmetries and transformations of quantum mechanics, so too the $\boldsymbol{G_{CP}}$ structure is a tool providing a formal, algebraic, way to model the combinatorial nature of linguistic features and local dependencies.

https://doi.org/10.3390/math14214037

3. Tensorization Materials and Methods for Long-Range Syntax

Workspace ***WS***, which is formalized below as a Fock lift introducing tensorization and superposition, is where the minimalist Internal MERGE (IM) is taken to occur. A Fock space is built from Hilbert space, allowing us to represent *variable numbers* of items as basis vectors (see Appendix A.3). It is a *lift* of Hilbert space to all possible multi-item configurations, an "arity" that matters as syntactic dependencies are built, as compared to the non-associative state of isolated items in ***L***. Smolensky had focused in 1990 on representing, in such Fock terms, filler-role bindings within a neural network context [41]. We start with background on how that view compares to the also classical outlook that current minimalism presumes—and a non-classical interpretation that also seems possible.

3.1. Background for Workspace Specifications

For Smolensky, vectors were activation patterns. With the goal of bridging the gap between symbolic and connectionist systems, his tensor products sought to bind word identities to syntactic positions in a vector space, resulting in high-dimensional vectors representing compositional structures. Symbolic structure was thus taken to map onto distributed representations. His superpositions were an aggregation of bindings via vector addition, nothing like a quantum superposition. Simpler binding mechanisms had been proposed within connectionism, although they faced limitations. For instance, merely concatenating vectors for roles and fillers would not allow for variable binding. In contrast, the tensor product representation provides a solution that allows for fully distributed, yet compositional, representations in a connectionist framework—offering a way to analyze how such representations are manipulated and learned.

With little relation to that project, while formalizing MERGE within a workspace, Marcolli, Berwick, and Chomsky (MCB) uses a graded connected Hopf algebra to track how syntactic objects associate in ***WS*** (via the product operation) *and* decomposed prior to interpretation (the coproduct [16]). Even though such an algebra features in perturbative quantum field theory, these authors only involve it in the process of *renormalization*, as used to split a divergent mathematical expression into a regularized and a divergent part. MCB draws an analogy to this process for "mining meaning" (from syntactic structures built by MERGE into a non-associative commutative magma). Others too have interpreted MERGE as an information renormalization process within tensor networks through an information-theoretic coarse-graining [42]. In so doing, the mathematical physics mechanics these authors borrow brings constraints along that pertain to corresponding Lie algebras, which are deeply connected to a Lie Group like our **SU(2)** in Section 2.1.

For **SU(2)** as a curved world of rotations, a Lie algebra $\mathfrak{su}$**(2)** allows us to transport the tangent space of **SU(2)** along any given path. This is analogous to a GPS, which constantly centers a flat map of a round earth directly underneath, as our car advances, rebalancing the coordinates in every new position. The primary directions of $\mathfrak{su}$**(2)** are called *generators*, which on the Hilbert space are realized as Hermitian operators, and thus, the Hilbert space serves as the medium for both the infinitesimal (Lie algebra generators) and finite (Lie group transformations) symmetries to be realized as said operators, bound together by the *exponential map* (recall from Section 2.1).

MCB defines a formal pre-Lie algebra structure whenever phrase structures are built after inserting subtrees. This is close to our paradigmatic considerations in Section 2, even if they do not go into the specificities of our selective approach. It is at this point that these authors show how a graded connected Hopf algebra contains an associated Lie algebra, made up of the primitive elements of its dual Hopf algebra. Although we will not dwell on this now, over a field of characteristic zero, cocommutative Hopf algebra and Lie algebra categories are equivalent via the universal enveloping algebra construction [43].

As a state (infinite-dimensional Hilbert) space, a Fock space as such is just a container that neither presupposes nor requires a Lie or Hopf algebraic structure: it is the arena where operators act (like those from a Lie or Hopf algebra). Smolensky used the power of the Fock space structure (summing tensor products) for representation only. In contrast, rather than asking how a sentence could be stored as a vector, MCB is examining the algebraic machine that constructs/manipulates sentences. Their MERGE involves a recursive structure that can be described by a Hopf algebra in renormalization theory—the procedure for dealing with infinite computations. But Fock spaces can also be equipped with a Hopf algebra structure to model how multiparticle states split or merge, non-classically. In short, the same mathematical structure can be interpreted in different ways, including non-classical systems like quantum-inspired models, which use the formal machinery of quantum theory as an inquisitive model for language [44].

A prime example of those is the Distributional Compositional Categorical (DisCoCat) model that Coecke and collaborators developed, invoking the mathematical formalism of category theory and quantum mechanics to model compositional semantics [45]. That framework employs a grammatical calculus—like Lambek's pregroup grammar [46]—to define the syntactic structure of a sentence. Word meanings in the model are represented as vectors or tensors in a Hilbert space, with grammatical rules (as symbolized in the category theory formalism) corresponding to the structure of quantum circuits, composing via tensor products. Meaning composition is then fundamentally linked to where a verb entangles the denotations of its subject and object to produce a new meaning for the sentence. In this approach, the grammatical structure behaves essentially like a system of quantum wires, with information flowing through entangled states.

Coecke's perspective bypasses Lie algebras by building its notions from category theory, mapping linguistic features to the quantum interpretation of Hilbert spaces [45], with non-commuting operators and entanglement as central for compositional semantics. Evidently, the mathematical structure is what it is, and the fact that the formalism is powerful enough to be interpreted in classical and non-classical ways is intriguing.

3.2. Separable Chains as Instantiating Internal MERGE (IM)

The present proposal is meant as largely compatible with the minimalist MCB. Smolensky's [41,47], though a conceptual precursor, presuppose both too much (about phrases) and too little (about syntactic features) to be of direct use. The Coecke moves are inspirational [45], although by focusing on Lambeck calculus and making relatively ad-hoc decisions about syntactic structure [48], it ends up being impractical for syntactic processes. It is certainly not meant as neurophysiologically or psychologically real, but we take the major Chomskyan result to be the insistence on a "biolinguistics" enterprise [49].

We may recall how, formally, a Fock space is a graded expansion of a Hilbert space:

$$\mathcal{F}(\mathcal{H}) = \mathbb{C} \oplus \mathcal{H} \oplus (\mathcal{H} \otimes \mathcal{H}) \oplus (\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}) \oplus \ldots \quad (17)$$

Here, each grade corresponds to configurations of 0, 1, 2, … lexical items from $\boldsymbol{L}$, with tensor products somehow encoding (as in DisCoCat, but with a very different implementation) how these items combine. In particle physics, Fock spaces describe multi-particle systems [50]; in cognition, the tensor product may encode combinatorial possibilities, as needed for syntax to move from paradigmatic lexical items in $\boldsymbol{L}$ to syntagmatic structured expressions, via combination in $\boldsymbol{WS}$. We can, then, describe a *Fock lift* as the move from the initial Hilbert space of stable lexical states (2 × 2 diagonal operators) in $\boldsymbol{L}$ to the Fock space in $\boldsymbol{WS}$ of structured configurations (4 × 4 tensors and their compressible superpositions as shown below), via the intermediate, selecting $\boldsymbol{A}_{\text{I}}$ step.

It is only in the syntagmatic Fock space that phrasal entanglement (chains in "displacement", "control", "ellipsis", etc.) and exclusion for grammaticality appear, as the IM

tensorization creates nontrivial correlations between phrasal contexts. The question then becomes: which of the entangled structures separate upon observation/interpretation (Observation Postulate, Born's rule, see chapter 2 of [51])—*and* may compress back to the paradigmatic Hilbert lexicon $\boldsymbol{L}$ after computation? This "cyclic" PSP component is central to our approach, in that we take an animal not just to use stored concepts for novel thoughts, but also to enrich the concepts through use, via newly acquired information.

The tensor product of any two matrices A and B, $A \otimes B$, combines two separate systems. If we start with two linguistic categories in $\boldsymbol{A}_{\mathrm{L}}$, each represented in its own 2-D $\boldsymbol{\mathcal{H}_2}$ built from the $\boldsymbol{G}_{CP}$, the tensor product combines them into a larger, 4-D Hilbert space $\boldsymbol{\mathcal{H}_4}$ that describes a new composite system [50]. Just as a similar lift is presumed in DisCoCat, this structure is used in our matrix syntax to model how linguistic elements combine to form larger structures. But in the present model, this is in terms of the IM procedure that links distinct contexts Λ, K, where Λ is contained within a phrase-theoretic root K in the ***WS*** area, in terms of some (internal) merger of a linking element α:

Axiom 7 (Tensorization): Internal MERGE IM in ***WS*** is a tensor product in $\boldsymbol{\mathcal{H}_2}$. (18)

Strictly, $\boldsymbol{L}$ itself paradigmatically lives in $\boldsymbol{\mathcal{H}_2}$, a Hilbert space, but the Fock lift will temporarily invoke an $\boldsymbol{\mathcal{H}_4}$ in its syntagmatic state, although of a very restricted kind. That limitation arises in terms of what constitutes a grammatical chain, as shown next.

To fix the mathematical form of the Fock lift: starting from the one-particle Hilbert space $\boldsymbol{H}_2 \cong \mathbb{C}^2$, the second-quantized space of interest is $\boldsymbol{H}_4 = \boldsymbol{H}_2 \otimes \boldsymbol{H}_2 \cong \mathbb{C}^4$, the two-particle sector of the full Fock space. On $\boldsymbol{H}_4$, the standard exchange operator $P_{12}(u \otimes v) = v \otimes u$ gives rise to two projection operators, $P_{\mathrm{sym}} = \frac{1}{2}(1 + P_{12})$ and $P_{\mathrm{asym}} = \frac{1}{2}(1 - P_{12})$, whose ranges are the symmetric (dimension 3) and anti-symmetric (dimension 1) subspaces of $\boldsymbol{H}_4$. This is formally the same construction that in quantum field theory yields *bosonic* and *fermionic* sectors, obeying (anti)commutation relations. We use "paradigmatic/syntagmatic" as the linguistic analogues, but the underlying algebraic content is that of (anti)commutation on second-quantized states. The $\boldsymbol{U}$-matrix construction of Sections 3.1 and 3.4 operates within a specific subspace of $\boldsymbol{H}_4$ (the 2D antisymmetric subspace) and the compression-back-to-$\boldsymbol{L}$ operation is a partial-trace-type contraction from a $\boldsymbol{H}_4$ state back to a $\boldsymbol{H}_2$ reduced density-matrix-like object (see Appendices A.3 and A.4 for technical statements).

We want a system that transitions from a lower paradigmatic state to a syntagmatic state, as necessary to the PSP cycle. These limitations are meant as syntactic—not semantic, pragmatic, or in other information terms. The system as described lumps all nouns together in terms of a Chomsky matrix $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -i \end{smallmatrix}\right]$ that we call C_1, thus with -C_1 as its polar verbal representation $\left[\begin{smallmatrix} -1 & 0 \\ 0 & i \end{smallmatrix}\right]$ (the same holding for the adjectival C_2 $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$ and its -C_2 polar counterpart $\left[\begin{smallmatrix} -1 & 0 \\ 0 & -i \end{smallmatrix}\right]$) Obviously, something else, then, must distinguish the thousands of diverse nouns, verbs, adjectives, etc., that exist for any individual in any given language. But this is algebraically trivial: one must presume this syntactic scaffolding has participated in tensor products involving other kinds of information, in terms of any conceptual algebra. In the present framework, individual lexical items are thus distinguished by their tensor product with high-dimensional conceptual vectors in domains external to the syntactic algebra, which the $\boldsymbol{G}_{CP}$ scaffolding acts on but does not determine.

Consider how tensor (Kronecker) products involve a matrix like the diagonal ones in $\boldsymbol{G}_{CP}$ and another matrix of a different kind and dimensionality, the resulting object being a higher-order tensor [50]. The tensor product of two matrices, A, B is basically a larger block matrix where each element of A is multiplied by the entire matrix B, the dimension of the resulting matrix being the product of the dimensions of the original matrices. Thus, in (19) and accompanying visual in Figure 2 we are tensorizing the nominal C_1 with some A as a pixel representation for the Pyramids of Giza, which in the output gets replicated as shown: with the upside-down version meant as a metaphor of whatever multiplier $-i$

may have done to the original matrix (which is irrelevant to the syntactic computation). In cognitive terms, nothing prevents comparable (sensorial or arbitrarily abstract) associations in any direction pertinent for categorial information encoding, so long as the 2 × 2 syntactic scaffolding is preserved for the system to recognize the element as a noun.

$$\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \otimes A = \begin{bmatrix} 1 \cdot A & 0 \cdot A \\ 0 \cdot A & -i \cdot A \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & -iA \end{bmatrix} \quad (19)$$

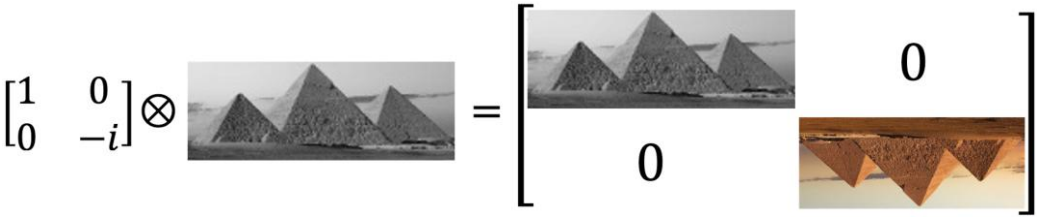


Figure 2. Visual representation for Equation (19).

Dimension growth after the IM product implied in Axiom (18) is significant, and internal to the syntax. Chains *within* chains clearly arise (thus, not just *Rome's destruction*, but also *Yesteryear's Rome's destruction*, *Nero's mother's Rome destruction*, etc.), so the explosive potential is unbounded. Hence, the practical desire to compress the resulting matrices, which has a more important consequence: it allows information syntactically acquired via IM for a category in its chain state to return to the paradigmatic state. Because in the tensor product the dimension of the resulting vector space grows as the product of the dimensions of the original vector spaces [50], a tensor *contraction* can be used to reduce dimension under certain conditions [47].

To formulate these ideas, we need a couple of final axioms: the crux of the "quantum" format of our proposal. The "chain matrices" are the 4×4 square matrices that result from the Kronecker product of the 2×2 lexical matrices as stated in Axiom 7. It should be these that represent the combined state of multiple linguistic dependents within what, hopefully, will be recognized by the system as a phrase. The term "chain" here denotes a sequence of such combinations prior to its observation at the interfaces.

Axiom 8: Chains are normalized sums of vectors from $\mathcal{H}_4$ that separate. (20a)

Axiom 9: When a chain *Ch* is sent to an interface, *Ch*'s vector $|\psi\rangle$ gets projected in one of *Ch*'s components $|\varphi\rangle$, according to probability $\langle\varphi|\psi\rangle^2$. (20b)

Axiom 10: Chain matrices have a non-trivial null (irrelevant) eigenspace. (20c)

Recall from Section 1.3 how an *eigenspace* is the set of eigenvectors corresponding to an *eigenvalue*. For any matrix *M*, these vectors are transformed by that matrix: only being scaled by the eigenvalue, not changed in direction; a *null eigenspace* corresponds to the eigenvalue zero (technically, all vectors that are mapped to the zero vector when transformed by *M*). "Non-trivial" in Axiom 10 means the null eigenspace is not just the zero vector itself, or it contains other non-zero vectors. We discuss that point next.

3.3. Phrasalization Conditions Associated to Chains

A non-trivial null eigenspace for a chain matrix contains all the information *irrelevant* to the syntactic structure. When the chain matrix operates on a vector, any component of that vector that lies in the null eigenspace is *compressed to zero*. In a situation as in Figure 2, there is (there, visual) information that we know immaterial to the syntax: in that such information never triggers any syntactic process. But, more importantly, information that does arise syntactically in the process of chain formation may fall in a scenario of such

deep symmetry, internal to these processes themselves, that formally opposing conditions make the result non-existent. In linear algebra, this constitutes a general mechanism for *dimensionality reduction*. In the present syntactic context, it is used for filtering out redundant information, with the goal of sending the syntagmatic object back to ***L***, if or when it successfully achieves such a useful dimensionality reduction that the dynamically obtained result may fit back in the paradigmatic format.

The model thus posits that only a subset of all possible IM operations, and their resulting tensorized states (generating the chains), are viable, with the null eigenspace as the grammaticality filter. If a chain results in a state vector that falls entirely within the null space of the chain matrix, the derivation leading to that configuration *crashes* (may not converge into the interpretive boundaries).

The axiom thus addresses the filtering process taken to happen at the sensory-motor ***PF*** and conceptual/intentional ***LF*** interfaces. When an entanglement has been built by IM, this *observation process* (Axiom 9, or the Born rule as is often called) can only apply to the remaining, non-null, relevant components of the state vector. The null eigenspace ensures that only meaningful linguistic information is available for interpretation and further processing. In particle physics, an operator's null space corresponds to the states the operator has no effect on. In a linguistic situation, the chain matrix represents the grammatical dependencies, with its null space representing configurations that are grammatically out of consideration. That is the mechanism by which the chain-formation process invoked via IM discards illicit dependencies. This, then, ensures that only valid syntactic information is carried algorithmically forward to where linearization, at ***PF***, and scope assessment, at ***LF***, take place [26]—so the information can eventually return to ***L*** (if updated).

Axiom 10 makes Axioms 8 and 9 work (given the general presuppositions in Axiom 7 and the rest of more basic axioms holding of ***L*** and $\boldsymbol{A}_{\mathrm{L}}$). The filtering provided by the null eigenspace ensures that only the relevant, well-formed, *and separable* components of the chain, in the sense of Axiom 8 (of Separation) converge to the interfaces, allowing for a coherent interpretation. In turn, the Born rule projection postulated in Axion 9 (of Observation) happens on the *remaining* state vector, after the irrelevant components have been filtered out by the null eigenspace. This guarantees that the final observed state corresponds to a collapse of the chain in one of the configurations it links, at that point terminating its (distributed) chain state, as a configurational syntactic object. It is this that links the syntactic structure to the observed linguistic output.

So: it is ultimately Axiom 10 that formally prunes the state space, ensuring that the compositional process, though otherwise generating many entangled possibilities, is constrained to produce only grammatically relevant and interface-interpretable outcomes. Somewhat surprisingly, when this happens, further results are achieved: a) a phrasal configuration ensues (as shown below), and b) the compressed chain under these conditions can fit back into the ***L*** confines. We discuss this aspect next.

3.4. U-Matrices and Separation

The key to, both, the phrasal configurations (the essence of the syntagmatic state) and the compression conditions (fundamental to the paradigmatic state) is in what may be called *Unit* or ***U***-matrices. We name these rank-1 sum-matrices 'Unit matrices', not because they are *unitary* (in the familiar sense of their inverse being equal to its conjugate transpose, thus preserving lengths and angles, with columns forming an orthonormal basis representing a transformation). These matrices central to chain formation are, in fact, *not* unitary (unlike the $\boldsymbol{G}_{\mathrm{CP}}$ elements), but they constitute the elementary *context units* from which chain states are formed. To avoid confusion with unitary elements, henceforth ***U***-matrix means "a rank-1 sum of two distinct $\boldsymbol{G}_{\mathrm{CP}}$ elements (each of which is unitary)."

Examples of ***U***-matrices are sums such as $\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=\begin{bmatrix}1 & -i\\ i & 1\end{bmatrix}$ (i.e., Pauli's $\sigma_Y + I$) or $\begin{bmatrix}i & 0\\ 0 & 1\end{bmatrix}+\begin{bmatrix}i & 0\\ 0 & -1\end{bmatrix}=\begin{bmatrix}2i & 0\\ 0 & 0\end{bmatrix}$ (i.e., Chomsky's $iC_1 + iC_2$)—and see Figure 3 for possible syntactic interpretations of the individual summands, as well as other ***G*CP** elements. In each case, the resulting matrix is not itself an element of ***G*CP**, but an aggregate formed from two of its elements. The 96 U-matrices (listed in the final figures in Appendix A.5) can therefore be viewed as the distinct rank-one matrices obtained by adding pairs of matrices from ***G*CP** (the resulting nonzero matrix having only one linearly independent row or column). For a 2×2 matrix, this is equivalent to saying that its determinant is zero. Consequently, each ***U***-matrix has one zero eigenvalue, while its other eigenvalue is equal to its trace.

-1 *Pauli Group*				***i* Chomsky Set**			
				Adjective	*Adposition*		
$Z=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$	$-Z=\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}$	$X=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$	$-X=\begin{bmatrix}0 & -1\\ -1 & 0\end{bmatrix}$	$C2=\begin{bmatrix}1 & 0\\ 0 & i\end{bmatrix}$	$-C2=\begin{bmatrix}-1 & 0\\ 0 & -i\end{bmatrix}$	$S1=\begin{bmatrix}0 & 1\\ -i & 0\end{bmatrix}$	$-S1=\begin{bmatrix}0 & -1\\ i & 0\end{bmatrix}$
NP		NumP		VP		*v*P	
				Pred	*C*		
$iI=\begin{bmatrix}i & 0\\ 0 & i\end{bmatrix}$	$-iI=\begin{bmatrix}-i & 0\\ 0 & -i\end{bmatrix}$	$Y=\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}$	$-Y=\begin{bmatrix}0 & i\\ -i & 0\end{bmatrix}$	$iC1=\begin{bmatrix}i & 0\\ 0 & 1\end{bmatrix}$	$-iC1=\begin{bmatrix}-i & 0\\ 0 & -1\end{bmatrix}$	$iS2=\begin{bmatrix}0 & i\\ -1 & 0\end{bmatrix}$	$-iS2=\begin{bmatrix}0 & -i\\ 1 & 0\end{bmatrix}$
NULL [-1]		PredP		NULL [-*i*]		DP	
1				***-i***			
		a	*n*	*Noun*	*Verb*		
$I=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$-I=\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}$	$iY=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$	$-iY=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$	$C1=\begin{bmatrix}1 & 0\\ 0 & -i\end{bmatrix}$	$-C1=\begin{bmatrix}-1 & 0\\ 0 & i\end{bmatrix}$	$S2=\begin{bmatrix}0 & 1\\ i & 0\end{bmatrix}$	$-S2=\begin{bmatrix}0 & -1\\ -i & 0\end{bmatrix}$
AP		DegP		PP		*p*P	
		p	*v*	*Det*	*T*		
$iZ=\begin{bmatrix}i & 0\\ 0 & -i\end{bmatrix}$	$-iZ=\begin{bmatrix}-i & 0\\ 0 & i\end{bmatrix}$	$iX=\begin{bmatrix}0 & i\\ i & 0\end{bmatrix}$	$-iX=\begin{bmatrix}0 & -i\\ -i & 0\end{bmatrix}$	$iC2=\begin{bmatrix}i & 0\\ 0 & -1\end{bmatrix}$	$-iC2=\begin{bmatrix}-i & 0\\ 0 & 1\end{bmatrix}$	$iS1=\begin{bmatrix}0 & i\\ 1 & 0\end{bmatrix}$	$-iS1=\begin{bmatrix}0 & -i\\ -1 & 0\end{bmatrix}$
NULL [1]		TP		NULL [*i*]		CP	

Figure 3. "Periodic table" with *G*CP matrices. Colors track det. labeling in Figure 1.

A chain $\{\{\alpha, K\}, \{\alpha, \Lambda\}\}$ linking distinct phrasal contexts Λ, K (for Λ contained in a phrase-theoretic root K within ***WS***), via the internal merger of a linking element α, boils down to the following (assuming normalized vectors):

$$|Chain\rangle = \frac{1}{\sqrt{2}}\left(|\hat{\alpha}\rangle \otimes |\widehat{K}\rangle + |\hat{\alpha}\rangle \otimes |\widehat{\Lambda}\rangle\right) = \frac{1}{\sqrt{2}}|\hat{\alpha}\rangle \otimes (|\widehat{K}\rangle + |\hat{\alpha}\rangle) \quad (21)$$

In a syntactic notation for this vector, for "two occurrences of the same token in a derivation" to amount to a *vector sum* in the space of matrices obtained after IM (with a tensor product structure), summed matrices must correspond to *orthogonal* vectors:

$$[\text{Chain}] = [\alpha \otimes K] + [\alpha \otimes \Lambda] = \alpha \otimes [K + \Lambda]. \quad (22)$$

Mutual orthogonality happens only *when the contexts K + Λ, sum up to a* ***U****-matrix* [52].

In a valid chain, a specifier is distributed over different positions, the linking element $|\hat{\alpha}\rangle$ in (21)/(22) acting *simultaneously* as specifier of $|\widehat{K}\rangle$ and $|\widehat{\Lambda}\rangle$. A chain is a sum of states sharing said specifier and, thus, based on the equally weighted superposition of (orthogonal) normalized vectors in $\mathcal{H}_4$. This allows us to incorporate a central idea from the Quantum Cognition frameworks: when the chain is sent to the ***PF*** and ***LF*** interfaces, only one of the configurations for the chain occurrences is observed.

The Measurement Postulate is, thus, invoked at this stage: once a superposed quantum system is observed in the appropriate basis, it "collapses" in one of the superposed options because of the measurement, which dictates the outcome [51]. The chain collapse applies at *both* interfaces: at ***PF***, the chain is spelled out as a single pronounced occurrence (one of the chain "copies" being externalized); at ***LF***, the chain is interpreted as a single scope-bearing occurrence (one of the "copies" anchoring interpretation). Indeed, the outcome need not be for the same occurrence at each interface! We return in Section 4 to how

this presumes hermiticity—so the outcome is a well-defined observable in either sensory-motor or conceptual/intentional projection.

That is how our approach models the puzzling behavior of syntactic chains, whose occurrences are distributed through the configurations they span over, but only manifest themselves observationally (at either ***PF*** or ***LF***) in exactly one of said configurations. It is worth showing this with some actual data involving "multiple copies" (occurrences) of displaced phrases to appreciate the scope of the result. Consider, for instance:

Someone unknown to Polymarket is likely to win. (23)

In English, we want this structure to be pronounced as in (24a), while in languages like the Northwest Caucasian Adyghe [53], articulation is as in the equivalent of (24b). But we never find either (24c) (copy pronounced nowhere) or (24d) (each copy pronounced):

[TP [DP *Someone*…]DP [VP *is likely* [TP [DP *someone*…]DP *to win*]TP]VP]TP
Someone unknown to Polymarket is likely to win. (24a)

[TP [DP *Someone*…]DP [VP *is likely* [TP [DP *someone*…]DP *to win*]TP]VP]TP
Is likely someone unknown to Polymarket to win. (24b)

* [TP [DP *Someone*…]DP [VP *is likely* [TP [DP *someone*…]DP *to win*]TP]VP]TP
(*Is likely to win.*) (24c)

* [TP [DP *Someone*…]DP [VP *is likely* [TP [DP *someone*…]DP *to win*]TP]VP]TP
(*Someone unknown to Polymarket is likely someone unknown to Polymarket to win.*) (24d)

Our approach to chains predicts, as shown in detail in the next section, that PF realization is either UP or DOWN (not both): because of the Observation Postulate.

Consider also (25), which can be interpreted in various ways—but not as in (26), noting that a hashmark in linguistics indicates semantic anomaly (vis-à-vis some structure):

No one known to Polymarket is likely to win. (25a)

[TP [DP *No one*…]DP [VP *is likely* [TP [DP *no one*…]DP *to win*]TP]VP]TP (25b)

No one known to Polymarket is in a position to win. (26a)

[TP [DP *No one*…]DP [VP *is likely* [TP [DP *no one*…]DP *to win*]TP]VP]TP (26b)

It is likely no one (known to Polymarket) will win—they won't run the bet. (27a)

[TP [DP *No one*…]DP [VP *is likely* [TP [DP *no one*…]DP *to win*]TP]VP]TP (27b)

No one known to Polymarket is in a position **not** to win. (28a)

[TP [DP *No one*…]DP [VP *is likely* [TP [DP *no one*…]DP *to win*]TP]VP]TP (28b)

Both a high scope interpretation for *no one* as in (26) and a "lowering" interpretation as in (27), with the subject at LF under the scope of *is likely*, are viable for (25). Note that the "lowering" reading in (27) is compatible with there being no individual winning because the bet was not run; but this reading does not seem possible when *no one* has wide scope. The crucial example is (28), where one copy of *no* is associated to its restriction to yield the negative quantifier reading *no one*, and another copy is also interpreted (a possibility relevant to yield the reading for (27) for when no bet was run, where the *no* in *no one* is denying the event itself: as in *no one died in the earthquake*, meaning that "it is not the case that anyone died"). The issue is how to prevent that sneaky "extra copy", given its mere

set-theoretic presence [54]. For us, once again the Observation Postulate forces either the UP or the DOWN occurrence to materialize via observation, this time around at the ***LF*** interface. The system, thus, chooses only one copy at each interpretive interface—and not necessarily the same, a phenomenon that was originally called "reconstruction". We return in Section 5.3 to a comparison with how MCB approaches these very matters.

3.5. Chain Labels

There is a unique role for our ***U***-matrices. Lexicon ***L***, with categories in a paradigmatic state—brought-to-attention via a lexical array A_L and an active workspace **WS** where categories algebraically become syntagmatic—are all meant as neurophysiological and psycholinguistic conditions on the algebraic space, affecting the form of representations language users deem grammatical. If entanglement is a linguistic resource, the language faculty works with the cheapest set of states: those ultimately *lacking* entanglement. While all grammatical states that correspond to a chain have the format in (29) (which generalizes (21) by allowing unequal weights *a*, *b*, subject to normalization as denoted by $\widehat{hat}$), as mathematically separable states, not all separable states are grammatically valid.

$$|Chain\rangle = \frac{1}{\sqrt{2}}\left(a|\hat{\alpha}\rangle \otimes |\hat{K}\rangle + b|\hat{\alpha}\rangle \otimes |\hat{\Lambda}\rangle\right) = \frac{1}{\sqrt{2}}|\hat{\alpha}\rangle \otimes \left(a|\hat{K}\rangle + b|\hat{\Lambda}\rangle\right) \quad (29)$$

The issue, then, is what else is a limiting condition for grammatically separable states.

For a grammatical chain compressible back to ***L*** in a PSP cycle, block K + Λ must have exactly one nonzero eigenvalue; equivalently, pairs $A, B \in \boldsymbol{G}_{CP}$ exist with $U = A + B$ such that U is rank-1. Those ***U***-matrices, 96 of which arise in the present construct (Appendix A.5), are thus essential but not exhaustive in determining grammaticality: several chains of the form $Chain = \alpha \otimes U$, where U is a ***U***-matrix as relevant, are still unacceptable to speakers. Interestingly, using *determinant* properties for tensor products that we review in Appendices A1 and A2, as well as a *normalization factor* N, one is taken to characterize a chain label in a way that is worth pondering already. Call the compressed 1 × 1 matrices U_C. The label of the chain formed by using ***U***-matrices as context, *after* compression (i.e., with U_C as context), follows the usual rules of a tensor product determinant and a rescaling by the norm, as given by the following equation, with the rescaling factor in (31).

$$Label\ (Chain) = \frac{1}{N}\det\ (\alpha \otimes U_C) = \frac{1}{N}\det(\alpha)\ (\det(U_C))^2 \quad (30)$$

$$Rescaling\ factor\colon N \equiv |\det(\alpha)\ (\det(U_C))^2| = |\det(U_C)|^2 \quad (31)$$

Then the *phase in the complex plane* of product $(\det(\alpha))\cdot(\det(U_c))^2$ constitutes a chain label, a composite of the specifier label ($\det(\alpha)$) and the (combined) "contexts" label ($\det(U_c))^2$—the norm of this quantity being the rescaling factor. So the chain label is the specifier label (matrix determinant) *times the square* of the U_C label, with a rescaling by its norm. Rescaling, as such, is necessary to keep the *consistency of linguistic labels*, a renormalization step. This refinement forces us to state Axiom 6 in the following, more accurate, fashion:

Axiom 6 (Labeling): A label is the complex phase of the determinant. (32)

To make that fully computable, we may define *label* as follows:

$$\text{For any } M \in \boldsymbol{G}_{CP},\ \text{Label}(M)_{\text{DEF}} := \det(M)/|\det(M)| \in \{+1, -1, +i, -i\}. \quad (33)$$

The four values correspond bijectively to the syntactic category slots (colors in the Jarret graph of Figure 1; see also extensions as in Figure 4). For a chain formed by *U*-matrix compression with linker α, the chain label combines specifier and context contributions:

$$\text{Label(chain)} = (\det(\alpha)\cdot(\det(U_c)^2)/N, \text{ with } N = |\det(U_c)|^2. \quad (34)$$

https://doi.org/10.3390/math14214037

That result again lies in {±1, ±*i*} (see Appendix A.4 on the normalization factor). A mapping label is, thus, a group homomorphism from $\boldsymbol{G}_{CP}$ onto the cyclic group of fourth roots of unity $\mathbb{Z}/4\mathbb{Z} \cong \{\pm 1, \pm i\}$, computable in constant time from the matrix representation.

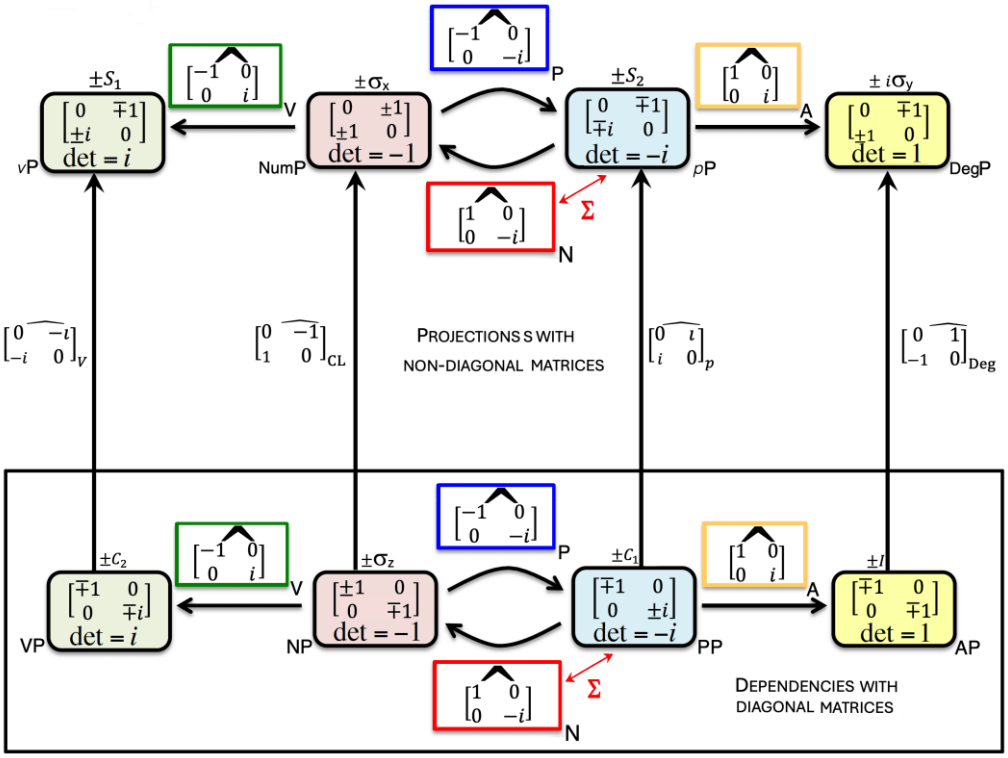


Figure 4. Extended graphic of $\boldsymbol{G}_{CP}$ combinations. Dependencies with diagonal matrices correspond to the Jarret graph. Projections to consistently labeled categories arise from non-diagonal matrices.

But a significant conceptual question remains: Why do we even need labelings in a system like ours? Below we conjecture that this is because we must relate terminal items in an effectively Markovian array $\boldsymbol{A}_{\mathrm{I}}$ to ensembles as arising in the IM format we are now exploring for ***WS***, where phrases are supposed to emerge. The issue is how that can be. Note that the objects in $\boldsymbol{A}_{\mathrm{I}}$, tokenized as they may be in the Jarret-graph(s), are still unitary matrices from $\boldsymbol{G}_{CP}$, as paradigmatic in that format as they are in ***L***. However, by merely associating various such paradigm-like elements in the grammatically viable ways we have been describing as syntactic chains, a curious sum becomes possible, as follows.

3.6. Phrasal Topology and Spectral Radius

Recall the elements in a chain: contexts $K + \Lambda$ [a ***U***-matrix] that IM links via some α, call it the "linker", and finally the projected result, whatever it may be—call that the chain "projection" Π. Suppose we sum these. For a formal object as in (14) that yields:

$$\underset{}{\text{AGGREGATE}} = \underset{\text{U-MATRIX}}{[K + \Lambda]} + \underset{\text{LINKER}}{\alpha} + \underset{\text{PROJECTION}}{\text{P}} . \tag{35}$$

To make the mathematical species of the summands explicit: all four terms in (35) are elements of $\boldsymbol{G}_{CP}$—2 × 2 unitary matrices with entries in $\mathbb{C}$—and the operation "+" is ordinary matrix addition (in the *vector space* generated by $\boldsymbol{G}_{CP}$). The bracketed $[K + \Lambda]$ is required to be a ***U***-matrix (i.e., $K, \Lambda \in \boldsymbol{G}_{CP}$ and $K + \Lambda$ has rank 1); the linker α is any element of $\boldsymbol{G}_{CP}$; and Π is likewise an element of $\boldsymbol{G}_{CP}$ (the "phrasal element" associated to the chain projection as the derivation continues). Although we call Π a (chain) "projection", for the

syntactic reason that it is the algebraic proxy for the phrasal head that the chain projects; clearly *Π is not a projection operator in the algebraic sense* (we do not require $\Pi^2 = \Pi$).

It is also interesting to note that, of 96 possible ***U***-matrices (***L***-compressible chain contexts), one-third consistently results in a matrix 2 × 2 whose spectral radius $\rho = 2\varphi$. This spectral radius ρ is the magnitude of the largest eigenvalue in a square matrix: a non-negative number that describes the growth rate associated with the matrix when used to scale vectors [52]. Unitary matrices have a $\rho = 1$, but when $\rho < 1$, a sequence of matrix powers converges towards the zero matrix, while if $\rho > 1$, the same sequence grows to infinity, meaning matrix multiplication by such a matrix makes vectors larger. Note how the $\rho > 1$ associated to the aggregate in (35) is of this last sort.

These matters are all most relevant to *phrasal topology*. Consider Lindenmayer (L) systems [55], originally designed to apply classical Chomsky grammar to biological growth. While in a Chomsky grammar there must be a *Traffic Convention* (so that only one rule applies at a time), in an L-system all rewritable symbols rewrite simultaneously, at each derivational step, with an entire symbol string transforming in one sweep [56]. For instance, consider a Fibonacci derivational system with specifications as in (36). The number of symbols at each derivational step grows like the Fibonacci sequence.

Alphabet: {+, −} (36a)

Rules: − → +, + → − + (36b)

Starting Axiom: (36c)

Iterations:
Step 0: − …… 1 token symbol
Step 1: + ……. 1 token symbol
Step 2: − + …….. 2 token symbols
Step 3: + − + ….. 3 token symbols
Step 4: − + + − + … 5 token symbols
… … … … … … 8 token symbols (36d)

That count is captured in a square (substitution) matrix, whose powers encode the effect of repeated rewriting. Such an operator is a $k \times k$ matrix M, where entry M_{ij} is the number of tokens of symbol a_i in w_j—for row index i the input symbol and column index j, how many of a_i tokens expand it. For $v^{(n)} \in \mathbb{N}^k$ the count vector after n derivations is:

$$v^{(n)} = \begin{bmatrix} \#a_1 \text{ in step } n \\ \#a_2 \text{ in step } n \\ \vdots \\ \#a_k \text{ in step } n \end{bmatrix} \quad (37)$$

The rewriting step updates counts by matrix multiplication $v^{(n+1)} = M\, v^{(n)}$. Thus, where $v^{(0)}$ is determined by the starting axiom, then $v^{(n)} = M^n\, v^{(0)}$. The string length at step n is the sum of entries of $v^{(n)}$, and the asymptotic growth rate is governed by the spectral radius $\rho(M)$ [52]. In the Fibonacci example, the incidence matrix is $\left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ (the Fibonacci matrix), sequential multiplication by which corresponds to repeated rewriting. The total "word length" (in computer science terms: meaning "repeated symbols") after n steps is proportional to F_n, the Fibonacci numbers, which the grand sum of $\left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]^n$ iterations tracks.

Next, in (38) we introduce a 3 × 3 matrix because our alphabet there contains two different phrasal types +, − plus a constant *h* that terminates without expanding—hence, the row of zeroes in the bottom line in (39), defining a singular matrix. Formally, it is immaterial whether we state the phrasal dependency in top-down (rewrite) terms as in (36) or, instead, in bottom-up (*merge*) conditions as in (38); what matters is that the presumed object, in Lindenmayer fashion, covers the *maximal expansion of the phrasal space* in such terms. Note how the object in (30d) systematically distinguishes head (*h*), complement

(i.e., *h* siblings), and specifier (elsewhere). Powers of that matrix correspond to the systemic iterations in (38), the third line in the Wythoff Array of Fibonacci recurrences [57].

Alphabet: {+,−, 0} (38a)

Rules: M {0, +} as −, M {+, −} as + (38b)

Top Axiom: + (38c)

Iterations: Step 0: + ……1
Step 1: + − ……2
Step 2: + − *h* + ……....4 (38d)
Step 3: + − *h* + + − ……6
Step 4: + − *h* + + − + − *h* + ……10
… … … … … … … … ……....16

$$X' = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},\ X'^2 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ X'^3 = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},\ X'^4 = \begin{bmatrix} 5 & 3 & 2 \\ 3 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix},\ X'^5 = \begin{bmatrix} 8 & 5 & 3 \\ 5 & 3 & 2 \\ 0 & 0 & 0 \end{bmatrix},\ldots \quad (39)$$

Sum: 4 Sum: 6 Sum: 10 Sum: 16 Sum: 26

Equivalences are known between the Fibonacci sequence and this "X′-sequence"—as we may call it because it correlates with the syntactic X′-scheme [7], as discussed next:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55…
+0, 1, 1, 2, 3, 5, 8, 13…
2, 4, 6, 10, 16, 26, 42, 68… (40a)

$$2 \times 1 = 2,\quad 2 \times 2 = 4,\ 2 \times 3 = 6,\ 2 \times 5 = 10,\ 2 \times 8 = 16,\ 2 \times 13 = 26,\ldots \quad (40b)$$

The Fibonacci sequence summed to itself with offset of three equals expansion (39) for the X′-recurrence (38a), as does doubling Fibonacci sequence terms (38b) (a correspondence between the recurrences, which stems from the X′ matrix in (39) also having $\rho = \varphi$).

The fact that the aggregate in (35) (of $\boldsymbol{U}$-matrices, their linker, and the chain projection) may yield a matrix whose $\rho = 2\varphi$ seems interesting, as this is the spectral radius that one expects of an aggregate seed matrix $\left[\begin{smallmatrix} 2 & 2 \\ 2 & 0 \end{smallmatrix}\right]^n = 2\left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]^n = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]^n + \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]^n$ [58]. A connection between the 2 × 2 aggregate with spectral radius 2φ and the 3 × 3 X′ matrix with spectral radius $\rho = \varphi$ is, thus, established by noting that 2φ is the dominant eigenvalue of the seed matrix $2 \cdot \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, which equals the sum of two copies of the Fibonacci matrix: corresponding exactly to the doubled Fibonacci sequence that is the X′-theoretic recurrence.

The aggregate of (four) unitary matrices from $\boldsymbol{G}_{CP}$ that properly entangles into a chain that may compress back to $\boldsymbol{L}$ may, in key instances, show not just a Fibonacci topology, but *a power expansion equivalent to the syntactically relevant X′-theoretic growth* in (39):

$$X'^1 = 2\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^0 = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ X'^2 = 2\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^1 = 2\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},\ X'^3 = 2\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^2 = 2\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},\ldots \quad (41)$$

Sum: 4 Sum: 6 Sum: 10

In other words: aggregating unitary matrices, we end up with what amounts not only to a syntagmatic configuration, but the one that has been observed for natural language: the X′-theoretic recurrence. The spectral fingerprint just described is verified by direct calculation in Section 4.2. What remains to be understood is how said aggregates yield such a result, as the elements selected in $\boldsymbol{A}_L$ map their way into syntactic existence via IM chains in **WS**, while other entanglements in the system do not. The characterization of which aggregates satisfy this spectral condition—and why only these—is shown in Appendix A.5.

4. Results: Phrasalization and a Proof of Concept

We have been calling *paradigmatic* and *syntagmatic* the language faculty macro-phases in PSP cycles: taking stock from an $\boldsymbol{L}$ syntactically existing in $\boldsymbol{\mathcal{H}_2}$ (with connections to other cognitive systems) and, through a selection $\boldsymbol{A}_{\text{L}}$ stage, generating a Fock lift into a $\boldsymbol{WS}$ living in $\boldsymbol{\mathcal{H}_4}$, where, after compression back to $\boldsymbol{\mathcal{H}_2}$, a phrasal disposition emerges (as well as the possibility of deconstructing representations so that new information returns to $\boldsymbol{L}$). A quantized expression encoded in $\boldsymbol{G_{CP}}$ allows different junctures in terms of the formal properties of each: either the paradigmatic stage of categories in $\boldsymbol{L}$ coexisting in the thousands for an individual's lifetime, or the syntagmatic stage of categories from $\boldsymbol{A}_{\text{L}}$ partnering within less than a second, to yield IM chains in **WS**, eventually to "quiet down" back to $\boldsymbol{L}$. Like dancers occupy space/time, chains generate a choreography where category occurrences at some, pre-interpreted phase (pre-observed into a configurational collapse) establish themselves as information carriers for a few instants. In that sense, not just IM as the entanglement discussed in Section 3, but also the aggregate in (35) is decisive: at a collapsing stage, the cards fall into an observable arrangement that, in the case of humans anyway, can be shared among conspecifics, leading to rich cultures.

4.1. Summary of Elementary Results

One issue is to be able to track what human syntax does (insert categories in selective ways, arrange them into phrases and chains of dependencies, have the latter manifest occurrence properties, handle global conditions, etc.) and very different one is to predict why any of those conditions should present the shape they exhibit. We are among those who take linguistic structure seriously; but it is hardly the same to postulate it ("VP is projected from V when it selects NP", etc.) than to deduce it from conditions that, if MP is generally correct, ultimately should reduce to mind/brain properties (in animals). From our Fundamental Assumption (Axiom 1) onwards we are taking that project to heart.

It is in that very first axiom that our system builds its algebraic relations based on what eventually turns into the $\boldsymbol{G_{CP}}$ via Axiom 2 (the Chomsky matrices) and reflective symmetry presumptions to let the system go beyond diagonal matrices. Our entire edifice is postulated within the confines of the $\boldsymbol{G_{CP}}$ and its tensorization (which leads to a much larger group of 4×4 matrices), so it starts in algebraic entities that arrange themselves algebraically to yield the interactions and objects of syntax. Although we have not gone into this in the present context, our wager is that it will be easier to rationalize the neurophysiological binding of objects like those in the $\boldsymbol{G_{CP}}$ to actual animal brains than it is to work with traditional syntactic representations (nouns, verbs, NPs, VPs, etc.).

Our approach assumes the minimalist sensitivity to the faculty of language separating paradigmatic lexical types in $\boldsymbol{L}$ from syntagmatic syntactic tokens thereof. For the latter, we presume a derivational $\boldsymbol{WS}$, which we ultimately expect to reside in a specific brain area (possibly the hippocampus and related networks). Our $\boldsymbol{L}$ is postulated to be a collection of multi-dimensional matrices, in that sense probably not unlike those in Large Language Models (LLMs). However, our high-D matrices have a crucial difference from LLM's: their overall shape is dictated by one of the Chomskys. Derivationally, only that "syntactic shape" matters in organizing dependencies like the Jarret graph (with the Chomskys and diagonal Paulis) or its extensions (featuring non-diagonal matrices in $\boldsymbol{G_{CP}}$ as in Figure 4). The empirical consequence of this move is a "first pass" selection structure. Although this has not been our focus, the approach allows us to construct a "periodic table" arranging the 32 categories in $\boldsymbol{G_{CP}}$ as (Figure 3).

The fate of such a project depends on the actual size of the human linguistic inventory—and whether it reduces back to some fundamental dimensions. Our system makes algebraic predictions in that direction. To be clear, nothing deep depends on specifically the $\boldsymbol{G_{CP}}$ being the system's basis, so long as some such algebraic basis is presumed. It

would be correspondingly natural for our approach if our essential 2×2 matrices contained, instead, quaternions $(\pm 1, \pm i, \pm j, \pm k)$ as entries, which would double-down on the Cayley/Dickson aspect of the algebra: effectively sending the edifice from a 4D non-commutative algebra as presented here to an 8D non-associative algebra; beyond that, since >8D algebras lose normed division, any computational system would be hard to state [59]. In the greater space, naturally the possibilities augment for a larger inventory of items. The point is, simply, that our system has a formal way to proceed. This is entirely compatible with the empirical discovery of the categories of language, but it postulates, as in Figure 3, a palette of possibilities to choose from in an informed and a priori matter.

All of that is classical, as discussed in Section 2. But then in Section 3, we upped the ante in a non-classical direction via a tensorization of the $\boldsymbol{G}_{CP}$ within a Fock lift. All minimalist systems presume a fundamental role of structure, with classical proposals—including MCB—"projecting" it from elementary syntactic categories while keeping their type stable. But our system predicts that structure arises when certain symmetries are met. Just as selection stabilizes into the Jarret graph, so too our syntactic chains presume a higher-order symmetry. And just as the Jarret graph prevents combinations that do not adhere to its strictures, the chain format determines aggregates that can be shown to correspond to (and, thus, sanction) a given phrasal topology. Our process is not a constructive one that instructs the system to "make a phrase for the task". Because our specific tensorized network (within its Fock lift) contains all possible entanglements among all possible phrases that elements selected into an $\boldsymbol{A}_L$ may yield, our non-classical approach is *selective* instead: only a fraction of entangled chains are observable as separable phrases.

Consequently, we are purposefully tackling two separate issues. One is to predict the behavior of chain occurrences. Chains arising from displacement have geared (transformational) generative grammar since inception [60], although the phenomenology and consequences arguably became clearer within MP [7,15]. The phenomenon of chain realization at either ***PF*** or ***LF***—and the fact that those need not coincide on the same occurrence—makes human language remarkable (as this extends from displacement to ellipses, cross-sentential gaps, long-range gate effects, and more). Denying any of this is not an option for us [61], and we are effectively postulating that nothing short of the possibilities associated to entangled networks can deal with its specific tenets. In this piece, we have only shown "tips-of-icebergs" in situations as in Sections 3.4 and 3.5, but monographs have been written about each of the phenomena that constitute long-range syntax.

However, equally important to our approach is that the very same chains at the core of said long-range correlations have a phrasalizing effect. While not all generative grammarians assume transformations as just alluded to, all minimalists do, analyzing displacement as IM conditions obeying minimality, locality, or general economy constraints (of the "third factor" sort). MCB is remarkable in that it explores an algebraic treatment of said conditions. But while the goal of their syntactic chains is to track representations that we agree exist in language, such a role is articulating for us: how we sanction a phrasal topology. This is a difference arising because of the non-classical nature of our system.

4.2. On Phrasal Observation

There is no equivalent to the aggregate in (35) in physics or chemistry (a better analogy in terms of interaction with the classical world). This is fine, as we are studying minds, not chemistry, even if the former arises within the confines of the organic version of the latter. In this sense, just as $\boldsymbol{G}_{CP}$ is postulated as unique to this system of thought, so too an aggregate as in (35) may be uniquely cognitive or linguistic—that being an empirical matter. The structural duality we need is clear: reservoir types in $\boldsymbol{L}$ storage vs. processing occurrences thereof in ***WS***, with observation at interfaces that lead to actual *token uses of the lexical types*. Formally, we must go from unitary, diagonal types stored in permanent

memory after a Fock lift to categories dynamically entangled superposed chains that are transient, exclusion-sensitive, and collapse back into interpretable phrases.

For physical particle dynamics, Hermitian operators are the observables: the class of matrices whose *eigenvalues are real* and *eigenspaces yield measurable states* [51]; without hermiticity, one could have unitary evolution, but no stable measurement. In ***WS***, entanglements may involve non-Hermitian (even non-diagonal) matrices—superpositions that cannot be "written back" in ***L***. But a $\boldsymbol{G}_{CP}$ subset contains Hermitian matrices only: the Paulis (including I). Making those privileged in the system leads us to postulate:

Phrasal Observation: For ***U***-matrix + link + chain-projection aggregates to be syntactically observable, the result of the sum must be Hermitian. (42)

This, which pertains to situations as in (35), restricts the class of collapse-triggering ***U***-matrices, guaranteeing that, upon interface measurement, eigenvalues thereof are real and corresponding eigenspaces, orthogonal. The following are the Hermitian matrices in $\boldsymbol{G}_{CP}$:

$$\text{For } \widehat{N}\text{: Pauli's } \sigma_Z \begin{bmatrix} \pm 1 & 0 \\ 0 & \mp 1 \end{bmatrix}_{-1/\text{NP}} = \text{NP [ANCHORS TYPING]} \tag{43a}$$

$$\text{For } \hat{A}\text{: The } I \text{ matrix } \begin{bmatrix} \mp 1 & 0 \\ 0 & \mp 1 \end{bmatrix}_{1/\text{AP}} = \text{AP [ANCHORS MODIFICATION]} \tag{43b}$$

$$\text{For } \widehat{Num}\text{: Pauli's } \sigma_X \begin{bmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{bmatrix}_{-i/\text{NumP}} = \text{NumP [ANCHORS REFERENTIALITY]} \tag{43c}$$

$$\text{For } \widehat{Pred}\text{: Pauli's } \sigma_Y \begin{bmatrix} 0 & \mp i \\ \pm i & 0 \end{bmatrix}_{i/\text{PredP}} = \text{PredP [ANCHORS PREDICATION]} \tag{43d}$$

After the hypothesized Fock lift in **WS**, postulate (42) has direct consequences for the entanglements in $\boldsymbol{\mathcal{H}_4}$ that converge as syntactic chains. Subscripts in (44) indicate a putative syntactic interpretation—for empirically established syntactic categories as in Figure 3—of some of the elements from $\boldsymbol{G}_{CP}$, given what is currently understood:

$$\underset{\boldsymbol{U}\text{-matrix}}{[Y_{\text{PredP}} + I_{\text{AP}}]} + \underset{\text{Linker } \alpha}{Z_{\text{NP}}} + \underset{\text{Projection } \Pi}{Y_{\text{PredP}}} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = 2\begin{bmatrix} 1 & -i \\ i & 0 \end{bmatrix} \quad \rho = 2\varphi \tag{44a}$$

$$\underset{\boldsymbol{U}\text{-matrix}}{[iC_1 + iC_2]} + \underset{\text{Linker } \alpha}{iX_{\text{TP}}} + \underset{\text{Projection } \Pi}{iX_{\text{TP}}} = \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} i & 0 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} + \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} = 2\begin{bmatrix} i & i \\ i & 0 \end{bmatrix} \quad \rho = 2\varphi \tag{44b}$$

In addition to Pauli's σ_X (X) as in (35c), σ_Y (Y) as in (35d), σ_Z (Z) as in (35a), or the I matrix (as in (35b)), here we include other categories within $\boldsymbol{G}_{CP}$ like iX, as coding what syntacticians call TP (Tense phrase)—for reasons that we cannot go into now beyond the "periodic table" in Figure 3, which extend dependencies as in the Jarret graph.

Consider the aggregate in (44a): $U = Y + I$ (a rank-1 ***U***-matrix), $\alpha = Z$, $\Pi = Y$. The 3-term partial sum $M_3 = U + \alpha$ equals the 2 × 2 matrix [[2, −i], [i, 0]], with characteristic polynomial $\lambda^2 - 2\lambda - 1 = 0$ and roots $1 \pm \sqrt{2}$; hence $\rho(M_3) = 1 + \sqrt{2}$, the silver mean. This mean appears as an intermediate spectral invariant; its role in the chain formation is analyzed in Appendix A.5. Addition of $\Pi = Y$ gives the four-term aggregate M = [[2, −2i], [2i, 0]], with characteristic polynomial $\lambda^2 - 2\lambda - 4 = 0$ and roots $1 \pm \sqrt{5}$, so $\rho(M) = 1 + \sqrt{5} = 2\varphi$. This value is the dominant eigenvalue of the scaled seed matrix $2 \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix}$, whose nth powers track the doubled Fibonacci sequence 2, 4, 6, 10, 16, 26, …—which the X′-recurrence maximal expansion counts as in (39). Enumeration over admissible (U, α, Π) quadruples confirms that 32 of the 96 ***U***-matrices (one third) yield aggregates with $\rho(M) = 2\varphi$.

Adding the hermiticity constraint of postulate (42) narrows the class of observable aggregates even more. Aggregate (44a) is Hermitian—its off-diagonal entries (−2i, 2i) are a conjugate pair—whereas the aggregate (44b) yields $\begin{bmatrix} 2i & 2i \\ 2i & 0 \end{bmatrix}$, which shares the spectral radius $\rho = 2\varphi$ but is not Hermitian. The linguistic reading of this contrast is that the standard

NP-AP-PredP predicational chain of (44a) is well-formed, while a putative "TP-of-TP" configuration as in (44b) is not. In point of fact, situations as in (44a) (where an NP linker takes an AP context into a predicational PredP projection [62]) are standard in the world's languages, while a hypothetical in which some imaginary projections of the Chomskys are linked by a TP to yield yet another TP of some sort… are not attested.

In sum, our constraints—(i) Compressibility: $K + \Lambda = U$ rank-1 (96 admissible contexts); (ii) Spectral signature: $\rho(M) = 2\varphi$ (32 of those); and (iii) Hermiticity: $M = M\dagger$—jointly define the class of grammatical, X′-projecting, interface-observable chain configurations (see Figure 3 below for a flow chart). This is, however, an area where we seek unification and empirical validation, so we leave further discussion until Appendix A.5, where Figure A1 shows the 32 (anti-)Hermitian $\boldsymbol{U}$-matrices of the grammatically relevant Set 1.

By separating the space of dynamical evolutions from that of measurable outcomes, hermiticity allows us to sort out through the categories in $\boldsymbol{G}_{CP}$ into what amounts to a "periodic table" of syntactic types as in Figure 3—with consequences for both interpretive interfaces (***PF*** and ***LF***). In particular, the operators in (43) correspond to semantic primitives (grammatically combining to encode kinds, properties, entities, and judgments, respectively) but also to default spell-out domains (although factors that we cannot review here can cause interesting misalignments from the defaults in each instance).

By triangulating hypothesized dependencies from $\boldsymbol{G}_{CP}$ elements, as they entangle into conditions arising via IM in the Fock-lift presumed for ***WS*** with compression, φ-based spectral conditions for the aggregate phrasal topology, and Hermitian assumptions, a research program may hypothesize *plausible cognitive anchors* for each of the sixteen unitary categories in the Abelian portion of our core group, hosted as paradigms in $\boldsymbol{L}$. In addition, by pairing equivalent projections—sharing the same determinant complex phase (per the revision of Axiom 6 in (32))—we can analytically hypothesize *consistent phrasal projections* (sixteen "twin" pairs in the entire $\boldsymbol{G}_{CP}$) for said lexical items.

A plan of that sort cannot be completed in a presentation like this. But the package is a guided approach to the enterprise Chomsky started in 1974 [20], which the Jarret graph as in Figure 1 gives a hint of. Even without going into the more elaborate entanglements explored in the previous section, that graph could be enriched as in Figure 4.

Regardless of how we extend categorial hypotheses and presumed dependencies beyond those conjectured eight (the Chomskys and non-diagonal "functional" or "grammatical" variants thereof), and putative projections arising in such terms, even the sketch in Figure 4 ought to suffice to see how systemic states are not observationally equal: only the Hermitians (Paulis) are observable in terms of the chain-collapse logic—per Born's rule, as relevant to mapping the derivation to its interfaces).

4.3. A Proof of Concept with an Actual Sentence

Consider the proposition *a cat is curious in nature*. Intuitively, such a sentence divides into a subject (*a cat*) and a predicate (*is curious in nature*), each of which gears a separate derivational thread, with a Σ START (self-merge) state and various other combinations depending on selected elements in each instance. In an article like this, we cannot go into a detailed justification of each of the analytical decisions in (45), as underlying the mapping from an array $\boldsymbol{A}_L$ of lexical items from the lexicon $\boldsymbol{L}$ to articulated configurations thereafter, via conditions as in Figure 3; but we can make a few informed comments in passing:

$$\{cat, a, \ldots\} \rightarrow [_{\text{NumP}}\ a\ [_{\text{NP}}\ cat\ ]_{\text{NP}}\ ]_{\text{NumP}} \quad (45a)$$

$$\{\text{-}s, be, nature, in, curious, \ldots\} \rightarrow [_{\text{TP}}\ \text{-}s\ [_{\text{VP}}\ be\ [_{\text{AP}}\ curious\ [_{\text{PP}}\ \text{in}\ [_{\text{NP}}\ nature]_{\text{NP}}\ ]_{\text{PP}}\ ]_{\text{AP}}]_{\text{VP}}\ ]_{\text{TP}}] \quad (45b)$$

Note how in (45a), we invoke not just a noun phrase NP, but a more articulated numeral phrase NumP, which is taken to syntactically project from the Num head *a*.

https://doi.org/10.3390/math14214037

To understand the analysis for the adjective *curious* in (45b), recall that the Start point in the Jarret graph in Figure 1 is presumed to be nominal (per the Anchoring Postulate). If this is that general, it must be presumed not just for verbal/adpositional projections but elsewhere too. Adjectives—which are often bare—are supposed in our terms to be more elaborate: *green (in color)*, *old (in age*), *resilient (in character)*, *curious* (*in nature*), etc.

The other important point to raise about (45b) is that the verbal morphology is separated as its own tense marker *−s*—which could have been a distinct auxiliary too (e.g., as in *a cat will (always) be curious*)—although it must be said that the output representation *be*+*-s* in English is obviously not **be-s* but, instead, the suppletive form *are*.

More relevantly to our syntactic concerns, we must invoke a chain via *internal* MERGE (IM) when associating (45a) and (45b). This triggers our entire machinery (the tensor products within the Fock lift, superpositions leading to an entangled network, and separation upon observation under orthogonal conditions). It is at this juncture that we need to consider situations as in (44) Section 4.2: determining a $\boldsymbol{U}$-matrix (relevant contexts K and Λ), a linker, and an ensuing projected (appropriately labeled) chain, as in (46) and Figure 5:

A cat is curious (*in nature*) (46)

a. $[_{\text{NumP}}\ a\ [_{\text{NP}}\ cat\ ]_{\text{NP}}\ ]_{\text{NumP}}$

b. $[_{\text{PredP}}\ \textbf{PRED}\ [_{\text{TP}}\ \textbf{-s}\ [_{\text{VP}}\ be\ [_{\text{AP}}\ curious\ [_{\text{PP}}\ \textbf{in}\ [_{\text{NP}}\ nature]_{\text{NP}}\ ]_{\text{PP}}\ ]_{\text{AP}}]_{\text{VP}}\ ]_{\text{TP}}]\ ]_{\text{PredP}}$

$[_{\text{PredP}}\ \textbf{PRED}\ [\quad \ldots\ \underline{[_{\text{AP}}\ curious\ ]_{\text{AP}}}\quad \ldots\quad]_{\text{PredP}}$

c. CONTEXT K: $\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$ + CONTEXT Λ: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

d. $[\ a\ [_{\text{NP}}\ cat\ ]_{\text{NP}}\ ]\ ]$

LINKER: $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

Figure 5. Separate derivational portions for (46) forming a Predication Chain.

The linker *a cat* enters a tensor product with the phrasal contexts: $\Lambda = I$ (the AP *curious*), contained within $K = Y$ (a PredP). This is a difference between the unassembled chunks in (45) and the assembly stage in (46): a "predicator" Pred as in (35d) is introduced [63], which prepares the (extended) verbal projection to meet the subject NumP. Recall, also, from Section 2.2 that MERGE is anti-symmetrical in terms of syntactic projection (Axiom 4); for an internal merger, this implies direct asymmetry (as there cannot be a *self* IM): the linker merges the contexts $[_K \ldots [_\Lambda \ldots {}_\Lambda] \ldots {}_K]$, extending the inclusive one (K). This asymmetry features centrally in how components are summed for phrasalization:

Phrasalization Assumptions (47)

The lower context of a chain is substantive. (47a)

The upper context of a chain is formal. (47b)

The chain linker is represented at its substantive projection. (47c)

The chain projection label is a function of its components' labels. (47d)

As a result of the phrasalizing assumptions, which generalize those in Section 4.2, we find:

$$[\ a\ [_{\text{NP}}\ cat\ ]_{\text{NP}}\ ]\ ]\ [_{\text{PredP}}\ \textbf{PRED}\ [\ \ldots\ \underline{[_{\text{AP}}\ curious\ ]_{\text{AP}}}\ \ldots\ ]_{\text{PredP}} \tag{48a}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \tag{48b}$$

https://doi.org/10.3390/math14214037

The sum $K + \Lambda$ (= $Y + I$) is rank-1 (a Hermitian $\boldsymbol{U}$-matrix in Set 1, Figure A1). The chain is the tensor $Z \otimes (Y + I)$—for (39c) demands that the linker representation is its substantive projection (so NP = Z, not NumP or any functional projection that may arise thereafter for quantificational purposes). This compresses to a 2 × 2 form (allowing the writing of the long-range correlation back to $\boldsymbol{L}$) by the mechanism further discussed in Appendix A.4. All three constraints (i)–(iii) in Section 4.2 hold, so the ensuing chain is grammatical. One must also bear in mind that, after observation, the resulting phrasal configuration continues to act as a PredP $\left[\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right]$ (constituting the chain projection), so the consequent sum is that added to the result in (48): $\left[\begin{smallmatrix} 2 & 1 \\ 1 & 0 \end{smallmatrix}\right] + \left[\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2 & 2 \\ 2 & 0 \end{smallmatrix}\right] = 2\left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. This is the aggregate of example (35a) worked out in Section 4.2: characteristic polynomial $\lambda^2 - 2\lambda - 4 = 0$, spectral radius $\rho = 2\varphi$, and Hermitian—our target matrix for the X′ equivalence.

Implicit to IM is the fact that the linker *a cat* starts the derivation as a specifier within the extended projection of the VP *is curious*, while (in English) it ends as specifier of the PredP, a separate chain occurrence. But either occurrence at ***PF*** is possible in many languages. Different interpretations ensue, e.g., in Spanish (and recall the "reconstruction" data in (25)/(28), Section 3.4). A prototypical construal for *un* gato, "a cat", is favored in (49a)—although an indefinite (contrastive) reading works too. The indeterminate reading is, however, obligatory with the word order in (49b), which disallows any genericity.

Un gato es curioso (de naturaleza).
a cat is curious of nature (49a)

Es curioso (*de naturaleza*) *un gato* (*pero no el otro*).
is curious of nature a cat but not the other (49b)

Such interpretive nuances are present in English too, though not reflected in word order. The ***LF*** differences correlate with which chain occurrence is semantically observed [7,54].

To be sure, human syntax is much more complex: aside from being curious by nature, cats can be said to nap, be on the mat, disappear when searched for, and so on. Many further nuances arise in such constructions, which force us to extend our notions from the relatively straightforward Jarret graph in Figure 1 to more elaborate dependencies of the sort in Figure 4 and beyond: to the full power of the $\boldsymbol{G}_{CP}$ (and bear in mind the "periodic table" in Figure 3). If the program holds, just as a combination of $\boldsymbol{U}$-matrices and relevant linkers, plus projection, works for predication, more elaborate dependencies should be possible. That is to be determined, remembering that language presents no shortage of phenomena that appear to "prepare" the dependencies for these sorts of compositions.

For instance, nominal expressions come equipped with an adpositional "case" morphology. This is obvious in a simple contrast between Tagalog and Japanese:

Umidlip ***ang****-pusa.* [Tagalog]
sleeps nominative-cat
"The cat sleeps." (50a)

*Neko-****ga*** *nemuru* [Japanese]
cat-nominative sleeps
"The cat sleeps." (50b)

While relevant for ***PF***, word order (whether the verb comes before/after the subject, whether the case-marker is pre- or post- positional) is not significant at ***LF***, as the compositional semantics proceed. English is hardly the best language to show case realization, although it does show up in pronominal forms (***it*** *sleeps* vs. ***its*** *sleep*) and in old English one would have distinguished ***ān catt*** *(hnappaþ)*, "a cat (naps)" (with *ān catt*, "a cat" in subject format), from *se hund æhte* ***æne catt***, "the dog chased a cat," where ***æne catt*** is in the object form. In any event, the theory will have to analyze how these nuances affect the

overall algebraic dependencies—similar questions being posed for agreement forms, "empty" categories, ellipses, and all other syntactic long-range correlations.

Schematic as this presentation must be, it shows a linguistic instantiation of the algebraic scaffolding. The three-constraint characterization (Section 4.2) is verified by exhaustive enumeration; the proof of concept (Section 4.3) applies this machinery to a specific sentence, with steps (48a-b) derived from Axioms 3, 4, and 7, and the phrasalization assumptions (47) constituting additional stipulations whose full derivation is deferred to future work. Aside from testing these ideas in relevant contexts, ways should be explored of unifying axioms or auxiliary assumptions as in (47)—which ought to follow from relevant combinations of algebraic conditions and constraints specific to linguistics (whatever their ultimate nature). In what follows, we discuss, in this regard, the interlock between items selected from ***L*** into $\boldsymbol{A}_{\text{L}}$ which bears on the role of syntactic labeling.

5. Discussion: The ($\boldsymbol{A}_{\text{L}}$, WS) Correspondence

MP presumes the human brain does not process language like a flat string (as LLMs currently do) but that it operates instead as a higher dimensional, dynamic, structured algebraic entity, which the IM process instantiates best with its multiple occurrence realizations. How radical that dynamicity can become is what is at stake in comparing various minimalist models, and ours takes that view to a non-classical extreme. This has various empirical consequences and a conceptual one: the role of labeling in the system.

5.1. On Interlocking A_{L} and WS

Consider an interlock between representations (in $\boldsymbol{A}_{\text{L}}$ and in ***WS***) that stems from correspondences as in (40), Section 3.6. The intuition is that, as an X′-theoretic object topologically expands, it can be seen as connecting two more elementary objects: one keeping track of heads from ***L*** in the form of an $\boldsymbol{A}_{\text{L}}$, another tracing projections in ***WS***, as in Figure 6. This is a consequence of the X′-theoretic expansion being within the Wythoff array, as in Section 3.6 and discussed in Appendix A.5: right after the Fibonacci and Lucas series. The yellow lines in Figure 6 sketch an X′-tree like as presumed in (37)–(39), Section 3.6. This, however, is a very idealized situation, with no reason existing to suspect that terminal items, as selected into $\boldsymbol{A}_{\text{L}}$, should arrange themselves into anything that elaborate.

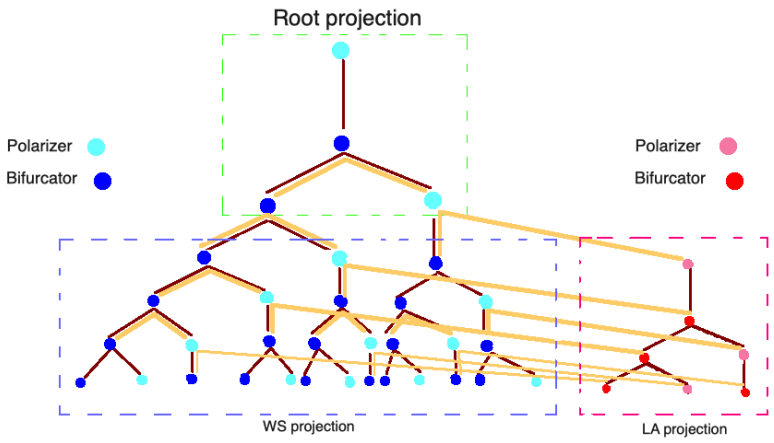


Figure 6. Correspondences between Fibonacci L-trees as hypothesized for **WS** and $\mathbf{A}_{\text{L}}$. Colors track different types of projections depending on whether they bifurcate (bifurcators) or not (polarizers).

Formally, the interlock could be a *monotone injective map* $\iota : \boldsymbol{A}_{\text{L}} \rightarrow \mathit{Term}(\boldsymbol{WS})$, where the domain $\boldsymbol{A}_{\text{L}} = (h_1, \ldots, h_n)$ is a Markovian sequence of n heads from ***L*** (with adjacencies constrained by the Jarret graph) and the codomain $\mathit{Term}(\boldsymbol{WS})$, the ordered set of terminal positions in an X′-tree derived from admissible chain aggregates $M = U + \alpha + \Pi$ (rank-1

compressibility, $\rho(M) = 2\varphi$, hermiticity). *Injectivity* requires that no two distinct heads of $\boldsymbol{A}_{\text{L}}$ occupy the same terminal position; *monotonicity*, that the Markovian order in $\boldsymbol{A}_{\text{L}}$ be preserved by the yield-order of *Term*(***WS***); that is, if h_i precedes h_j in $\boldsymbol{A}_{\text{L}}$, then $\iota(h_i)$ precedes $\iota(h_j)$ in the terminal string. Surjectivity is not required: an X′-template may leave terminal positions unfilled (corresponding to elliptical or default-valued "empty" slots). Interlock ι is, thus, an order-preserving partial injection between two well-defined discrete structures, ranked by a cost function as discussed below (unmatched positions plus span). For instance, the interlock ι : (*cat*, a, -s, *be*, *curious*, *nature*, *in*) → *Term*(***WS***) must assign each lexical head to its corresponding X′-terminal in the resulting phrasal object, with the cost of such a correspondence being minimal (no unmatched positions, unit span).

But presuming a Fibonacci topology for phrasal projection, a simpler disposition can be assumed for the terminals in $\boldsymbol{A}_{\text{L}}$, preserving the X′-theoretic assumptions in Figure 6. In Figure 7, we see a mapping to the Fibonacci object from an unassuming terminal list. Topologically, phrasal objects there are of the same kind as the Fibonacci object in (38)-(39), Section 3.6, although we are not representing expansion beyond given nodes. Ponder, then, what complexities $\boldsymbol{A}_{\text{L}}$ may bring to its IM projection, presuming an algorithmic START Σ and selectional specificities as the Jarret graph tracks (or extensions as in Figure 4). Since the $\boldsymbol{A}_{\text{L}}$ idea was proposed, it became relevant to consider whether such formal objects exhibit *cyclic* conditions—as presumed by "phasal" multiclausality in the MP sense [64]. That is to be empirically elucidated, but it ought to affect the proper mapping between $\boldsymbol{A}_{\text{L}}$ and the phrasal realization for ***PF*** and ***LF*** purposes, in terms of IM at ***WS***.

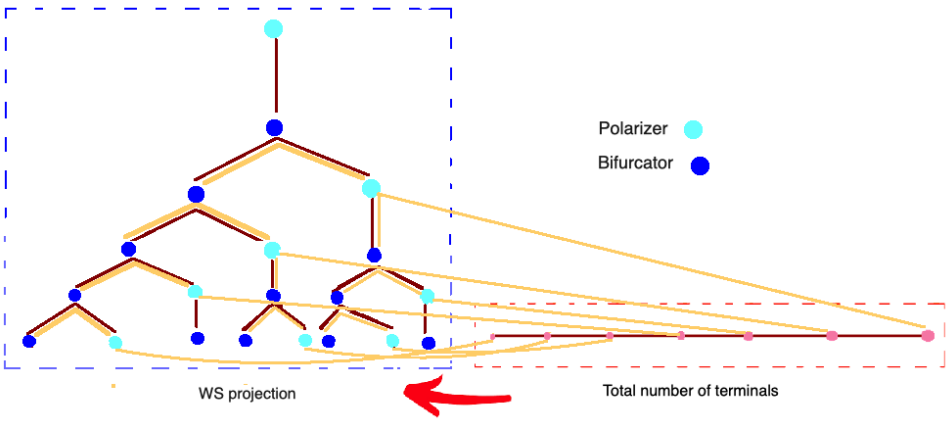


Figure 7. Correspondence between Fibonacci L-tree hypothesized for **WS** and sequential $\mathbf{A}_{\text{L}}$. Distinctions between polarizers and bifurcators are kept for the projections, but the lexical heads are presented in a mere list that gets injected into the projected structure, as signaled by the arrow.

As noted, at the very least, $\boldsymbol{A}_{\text{L}}$ must be a finite multiset of lexical heads summoned from ***L*** and constrained by the Jarret graph (and extensions), in terms of an algorithmic launch and termination, cyclicity considerations, and perhaps other such fundamentals.

Labels ought to be central in keeping information consistent from the elements selected from ***L*** (via $\boldsymbol{A}_{\text{L}}$) onto the ones that successfully entangle in ***WS*** (with the ability to observably phrasalize and effectively compress back to ***L***-size elements). One wants not just to randomly activate items in ***L***. During awareness, when claiming that *the cat is on the mat*, we do not mean the inverse (that the mat is on the cat), let alone the dog being on the rug, etc. While any of that may, indeed, randomly happen in a surrealist dream, none of it seems to affect normal thinking or speech. Our system, thus, in principle allows for a disposition where the various bits of information entangle in bizarre ways, which we would not want to forbid from a cognitive reality; but, more importantly, the system ends up being computational only if there are correspondence procedures between ***L*** and ***WS***.

This cannot be achieved without the addressing role of labels in that charting; moreover, it can be seen to bear on the connection between Lie algebras and the putatively enveloping algebra(s), which we return to in Sections 5.3 and 5.4 as we compare proposals.

An L-system as in Section 3.6 defines a topological dynamical organization (a *subshift* of finite type [65]). Its topological entropy can be measured in relation to the seed matrix that generates it as it self-scales (via matrix powers), measuring how much "complexity" (distinct configurations) the topology of the shift space allows. Beyond systemic growth, a spectral radius ρ thus provides a topological invariant of the symbolic system. This is relevant as we consider quadruple aggregates leading to a phrasally observable configuration, the sum K↑ + α + L↓ + Π (i.e., $U + \alpha + \Pi$). How does this addition—which in successful cases yields a signature $\rho = 2\varphi$—address elements from $\boldsymbol{A}_{\mathrm{L}}$ in a fashion that may collapse into a phrasal configuration interpretable to the ***PF*** and ***LF*** interfaces?

If the discussion above is on track, our expectation is that the match will be generally fruitful only for small to average sentences, although tentativeness may arise for more elaborate sentences or under conditions taxing awareness. If this is the right approach, one could use a cost function (unmatched + span) to pick preferred mappings and show the distribution of selected mapping shapes (clustered vs. spread), which becomes harder to compute for sentence planning involving elaborate thoughts or without full awareness (in a mapping from an $\boldsymbol{A}_{\mathrm{L}}$ to a grammatical output).

That cost function is, then, a map cost : $\mathit{Interlocks}(\boldsymbol{A}_{\mathrm{L}}, \boldsymbol{WS}) \to \mathbb{R}^{\geq 0}$ defined on the set of admissible interlock maps $\iota : \boldsymbol{A}_{\mathrm{L}} \to \mathit{Term}(\boldsymbol{WS})$, with cost(ι) = #{unmatched terminal positions in $\mathit{Term}(\boldsymbol{WS}) \setminus \mathrm{image}(\iota)$} + span(ι), where $\mathrm{span}(\iota) = \max\{|t_{k+1} - t_k| : 1 \leq k < n\}$ computed over the terminal indices t_k = position of $\iota(h_k)$. The optimization may, thus, be an argmin (*argument of the minimum*, the input set of values that gives the lowest output for a given function [66]) over the finite set of order-preserving injections. Unmatched count penalizes wasted phrasal capacity, while span penalizes non-local dependencies that stretch memory buffers; the argmin picks the tightest packing. Both terms are computable in $O(n)$ per candidate, with the number of candidates bounded by binomial coefficient $C(|\mathit{Term}(\boldsymbol{WS})|, n)$ [67]. So, for actual sentence lengths, the optimization is polynomial in n for fixed $|\mathit{Term}(\boldsymbol{WS})|$ (as the number of order-preserving injections from an n-element set to a k-element ordered set is $C(k, n)$, and evaluation of cost per candidate is $O(n)$).

5.2. Final Speculations

The discussion above, in a nutshell, presupposes an addressing mechanism from $\boldsymbol{A}_{\mathrm{L}}$ to ***WS***. Even if invoking Markovian settings as presumed in the Jarret graph (for starting algorithmic conditions, sub-categorization viabilities, cyclic interfaces), we still need to track whichever selections happen in $\boldsymbol{A}_{\mathrm{L}}$ to whichever projections may optimally correspond to them in ***WS***. The extension of the Jarret graph in Figure 4 assumes that much:

$$\text{Categories in red like } \begin{bmatrix} \overbrace{1} & 0 \\ 0 & -\iota \end{bmatrix} \text{ map to a matrix space with determinant } -1 \tag{51a}$$

$$\text{Categories in green like } \begin{bmatrix} \overbrace{-1} & 0 \\ 0 & \iota \end{bmatrix} \text{ map to a matrix space with determinant } i. \tag{51b}$$

$$\text{Categories in yellow like } \begin{bmatrix} \overbrace{1} & 0 \\ 0 & \iota \end{bmatrix} \text{ map to a matrix space with determinant } 1. \tag{51c}$$

$$\text{Categories in blue like } \begin{bmatrix} \overbrace{-1} & 0 \\ 0 & -\iota \end{bmatrix} \text{ map to a matrix space with determinant } -i. \tag{51d}$$

Grammatically, those will be NumP, *v*P, DegP, and *p*P, respectively [62], but that notation is even less far-reaching than the colors—which keep track of "extended projection" beyond the confines of a given phrase. A label in present terms is the complex phase of a matrix determinant (per Axiom 6). When a ***U***-matrix is rank-1 with nonzero eigenvalue u

(a scalar), the compressed U gives the chain label via $\det(\alpha)\, u^2/|u|^2$ (so, for admissible write-back, we expect the label to be in $\{\pm 1, \pm i\}$) [52]. That is, basically, the information the system must preserve when mapping from $\boldsymbol{A}_{\mathrm{L}}$ to ***WS***, or it will not correspond to attested observations. Note how, for the operator matrices in (52) (the Chomskys), that is actually not their determinant. These are *operators* on $\boldsymbol{\mathcal{H}_2}$, whose identity for ***PF*** and ***LF*** is no more crucial than that of an inventor is for a given breakthrough (like the distinction of AC vs. DC currents, during the 19th century). Matrices in $\boldsymbol{\mathcal{H}_2}$ can be understood as operator $\hat{M}$ or vector M within the overall vector space, and only in the latter guise is the matrix determinant taken to carry substantively to the interpretive components.

Future developments obviously must go beyond the extended Jarret graph into domains where determinant/label preservation matters. In aggregates as in (35), Section 3.6, what is meant by "chain projection" Π is constrained by harmony conditions along the lines in Section 4.2, ensuring successful chains associate to the combined action of some ***U***-matrix (relevant contexts) and some linker α that undergoes IM within ***WS***. Although this poses no ontological issues for the system that (43) in Section 4.2 does not raise already, it is key to understanding what gears the transition between paradigmatic and syntagmatic states, as central to the PSP cycle. The nature of these processes is the most intriguing aspect of our program. Combinatorial decisions for the implied model abstract away from motivations internal to the system's algebraic logic: whether in terms of working-memory, attention span, awareness, etc. [68], *or* interface conditions. The latter may well be relevant to the manifestation of systemic implementations, as in the "semantic flavor" of the Anchoring Postulate in (14), Section 2.2. But underlying it all, an algebra must carve out evolutionary and developmental paths—much as geophysical forces shape continental configurations.

The grammar, reflecting a language acquisition path from where these considerations first emerged (the Projection Problem [20]), may have stumbled onto some sort of minimal bottom-up fit relating the $\boldsymbol{A}_{\mathrm{L}}$ terminals to functional expansions consistent with IM specificities in the WS, subject to conditions as in (52) below. The idea is to yield a tool to explain why sentences, built from lexical items, "project" with familiar X′-structures.

Extended Projection Conditions: Map $\boldsymbol{A}_{\mathrm{L}}$ terminals consistent with (i), (ii), and (iii):

(i) Only to polarizers (non-branching terminals in the Fibonacci tree).
(ii) Preserving $\boldsymbol{A}_{\mathrm{L}}$ (Markovian) adjacency-driven orderings.
(iii) To the smallest phrasal topology consistent with (i) and (ii). (52)

$\boldsymbol{A}_{\mathrm{L}}$ has always been presumed as more articulated than an unordered multiset, keeping track of conditions as in the Jarret graph (and extensions). But how information from ***L*** enters $\boldsymbol{A}_{\mathrm{L}}$'s weakly ordered confines has consequences arguably related to Dehaene's description of animals tracking numerosities up to ~4 (subitizing), shifting to approximate estimation beyond that [69]. The reliability of subitizing at small totals may depend on the brain binding each increment to a running total, thereby creating redundancy between "what just happened" and "where we stand." This underlying mechanism may allow $\boldsymbol{A}_{\mathrm{L}}$ to gauge itself sequentially as it maps to ***WS***, restricting any presumed match.

The first summatory function of a sequence entails, by definition, that whatever increments ones chooses, their cumulative sum "remembers" them [70]. For instance, summing the terms of the (random) sequence 1, 2, 3, 1, 6, 2, 5… as they accrue, the n^{th} number in the cumulative sequence is less than the n+2^{th} number in the regular sequence by the second number thereof as in Figure 8, with consequences for an approximate gauge of the sequence. Note how, as the series advances (no matter what it may be), the *less reliable a cumulative gauge one obtains*—beyond small numbers that stay small when added.

https://doi.org/10.3390/math14214037

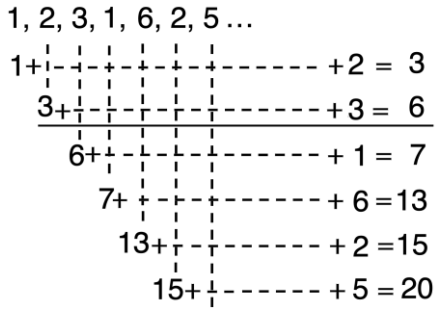


Figure 8. Geometric decomposition of a first summatory function, illustrating the bijective preservation of discrete increments through interval concatenation.

Evaluating the order numbers are presented against their accumulation:

a. 1st step: **1** token accessed can be sensed as **1** token total.
b. 2nd step: **2** tokens accessed can be sensed as **3** tokens total.
c. 3rd step: **3** tokens accessed thereafter can be sensed as **6** tokens total.
d. 4th step: **1** token accessed thereafter can be sensed as **7** tokens total.
e. 5th step: **6** tokens accessed thereafter can be sensed as **13** tokens total.
f. 6th step: **2** tokens accessed thereafter can be sensed as **15** tokens total. (53)
g. 7th step: **5** tokens accessed thereafter can be sensed as **20** tokens total.

If $\boldsymbol{A}_{\mathrm{L}}$ receives a lump from $\boldsymbol{L}$, the system faces the issue of determining if it is, say, seven. If animals lack a number detector—presenting, instead, some mechanism for tracking successive events—we could think of a numerosity system as an integrator, with each new token generating a pulse to be aggregated to a running tally. Unless summoning items from $\boldsymbol{L}$ into $\boldsymbol{A}_{\mathrm{L}}$ is seen as a careful incremental selection, it would not naturally yield subitizable states. Binding increments to totals yields redundancy: at each step, the system may detect increment size and running total. For a small n, the dual code is mutually reinforcing: both the increments and cumulative state carry the information that can be matched for accuracy, grounding the match. For a larger n, there is no reliable correlation. This approach presumes language grammars lack counters [71], an idea going back to Miller and Chomsky's contrast between infinite grammar and finite processing [72].

Cognitive mechanisms like subitizing are clearly needed outside grammar, and thus it is relevant to find them in other animals. This is important when pondering how the ***WS*** Fibonacci frame determines, to yield the X′-structure, a given number of polarizers to activate, as corresponding to a particular $\boldsymbol{A}_{\mathrm{L}}$. If only lump totals could be estimated, the mapping would be fragile; but if the $\boldsymbol{A}_{\mathrm{L}}$ includes an *increment-plus-accumulation gauge* as sketched in (53) for small n, every derivational step can be read ambiguously: (i) "how many tokens are retrieved" (raw sequence), and (ii) "how many tokens have been accumulated" (partial sums). For small n (be these tokens or derivational steps in accumulating them), a step-by-step growth could be matched against a Fibonacci growth in ***WS*** consistent with the $\rho = 2\varphi$ we postulated for aggregative chain ensembles and their presumed projection, thus ensuring what amounts to an injective interlocking of $\mathbf{A}_{\mathrm{L}}$ and **WS**.

This approach may also give a rationale for the upper limit researchers have found of three verbal arguments in human languages, with anything beyond being vanishingly rare [73]. Little debate exists beyond arguable instances of verbs taking four arguments [74]; situations involving pentavalent verbs or more are unattested—although no logical reason exists for not verbalizing that. So aside from connecting the restriction to Miller's "magical number 7" [75], the present tactic is also consistent with why a multi-valence verb of the form X′-form involves very small argument numbers in $\boldsymbol{A}_{\mathrm{L}}$ configurations.

5.3. A Comparison with the Marcolli, Chomsky, and Berwick Model

Although detailed comparisons among models is beyond our current scope, consider where ours differs from related ones. This was already shown in Section 3.1 regarding Smolensky's tensor product variable bindings. There, we also discussed DisCoCat, which is closer in spirit to our approach: in the non-classical presuppositions and their use to understand syntactic representations themselves. Like Coecke's work and ours, MCB also examines the putative algebraic machine that constructs/manipulates sentences, but although that piece comes from MP presuppositions we also adhere to, it does not invoke primitives that go beyond the vocabulary inherited from a classical tradition materializing into Chomsky's 1974 categorial features [20], which MP presumes in its essentials. Our Fundamental Assumption (Axiom 1) sets us in a related, but also rather different, path.

It should be easy to see that, while our entire edifice is postulated within the confines of the $\boldsymbol{G_{CP}}$ and its tensorization, the Hopf algebra that MCB builds manipulates standard syntactic objects based on symbolic elements (like NP, TP, etc.), together with an ontologically primitive $\boldsymbol{WS}$ where these are assembled into sets. Our more basic system, instead, starts in purely algebraic entities, which forces us to postulate the anti-symmetric self-merger of nouns at the bottom of every derivation, a need not arising for MCB. For us, this is a consequence of our Axiom (4), a crucial relation between the subset of $\boldsymbol{L}$ that arranges as an array $\boldsymbol{A_L}$ as it projects into $\boldsymbol{WS}$, which we postulate as anti-symmetrical.

An important desideratum of the MCB system is non-associativity. This is an excellent result if achieved, for one cannot have strict compositionality (as presumed for the human language faculty) without that: the structurally ambiguous *watch me with binoculars!* simply cannot be such that one of the readings entails the other (as is the case for associative relations: $(2 \times 3) \times 4 = 2 \times (3 \times 4) = 24$). As stated here, our system is, indeed, associative, even if non-commutative (it basically has the general properties of a 4D algebra), so it needs to be enriched. Extending the present framework to an 8D non-associative algebra, via quaternionic matrix entries, seems like the natural next step. Now, the way MCB achieves associativity is not directly algebraic either, for the Hopf algebra per se *is* associative. It is worth considering that point.

In MCB, MERGE is postulated to generate a free, commutative, non-associative *magma*. This is an algebraic structure consisting of a set and a single binary operation within it that must be closed. The collection of active computational resources (multisets of syntactic objects present in a "scratchpad" of the language faculty during a derivation: $\boldsymbol{WS}$) is what gets endowed with a Hopf algebra structure with a coproduct (Δ) to model the dismantling, extraction, or accessing of sub-trees (accessible terms) from $\boldsymbol{WS}$, and a product to re-assemble them. To make the Hopf-algebraic structure interact with non-associative syntactic generation, MCB defines the action of MERGE through a linear endomorphism on a vector space, thus bypassing constraints of internal multiplication associativity. It is in that fashion that a restriction of zero-cost operations to (external and internal) MERGE arises structurally from the Hopf framework's grading and compatibility conditions, letting the non-associative tree-building operations be tracked by the associative bookkeeping of the $\boldsymbol{WS}$ Hopf algebra (see [76] for discussion).

That summary is enough to see that the non-associativity, as such, is not native to the algebra (as it would be for an 8D algebra of octonions, for instance). It is imposed by postulate, which is of course fine—but we too can do that, as the constraint per se is not in doubt. The point harkens back to whether (in our terms) a larger group than $\boldsymbol{G_{CP}}$ is needed, the next natural candidate utilizing quaternions (which would ipso facto yield non-associativity in 2×2 matrix terms). All these issues should be analyzed in the future.

One other important aspect of MCB is its ability to deconstruct. The proposal is deployed to make the relationship to the interfaces work in their constrained ways: the Hopf algebra antipode acts as the filter that can algebraically treat complex phrase structures so

they can be externalized into a flat, sequential, stream. But this also raises the possibility of a twist that is central to our PSP cycle: the ability for information to *return* to ***L*** after syntactic use, crucially with (slightly) altered presuppositions. There is no direct way of achieving this aspect of the system without going beyond computing truth conditions. Those are central to understanding and judgement, but language does not stop at judging the world around: trusted findings are integrated into permanent knowledge. However this is achieved, we clearly need "mining" meaning tools, of the sort MCB investigates.

5.4. Classicality and Evolutionary Considerations.

That dovetails into the most important conceptual difference between the two minimalist proposals: classicality (which hinges on non-commutativity). That is why we went into label considerations in Sections 5.1 and 5.2. For, if the relation between that sub-portion of ***L*** that we have called the lexical array $\boldsymbol{A}_L$ and ***WS*** is itself not necessarily algebraic, then all algebraic bets are off: what happens after the Fock lift (syntagmatically in ***WS***) is only connected in non-algebraic ways to what happens *before* (in $\boldsymbol{A}_L$, paradigmatically as is the case for everything in ***L*** more generally). While this may seem like anathema for using algebra to add rigor, it is nonetheless worth pondering. Even if (like us) one takes algebra as central to understand natural processes, that does not entail that nature used algebra "all the way down", in every possible interaction between systems (keeping in mind celestial body collisions, tectonic clashes, evolutionary races, and the like).

Chomsky has taken narrow syntax as a biological invariant: a "perfect system" that operates under rigid, unchanging, structural laws of nature, like a snowflake. Indeed, MCB is effectively selling the idea that the language faculty, as presented in the terms they postulate, obeys the same mathematical symmetries (like a Lie-Hopf correspondence) that physicists use to describe the space-time fabric—so MP is, in that arguable sense, mathematically inevitable. The first practical problem we have with that is finding it surprising that such mathematical inevitability should have been evolutionarily realized so recently and only in our lineage. Second, with some skeptics [77,78], we find other animals smarter than this view presumes. If, as we said in Section 1.2, other creatures also present a language of thought (LoT), ipso-facto that commits us to other animals presenting at least some form of *merge*, if not the full-blown MERGE that MCB have proposed for humans. At that point, the evolutionary issue becomes what a transition from *merge* to MERGE amounts to, and whether it is, itself, ruled by algebraic considerations.

Chomsky has always drawn a line between internal, autonomous, syntax, and any form of externalization (communication/pragmatics). We have no problem generalizing that view so much that it holds also of other systems of thought—possibly as far back as the origins of the bilaterian clade. If the syntax of *merge* is such a good idea for evolution to have given animals for thinking, it should be no surprise to have its essentials emerge eons ago, perhaps in pre-Cambrian (snowball) Earth. But then very recently, within the last half million years, the essentials of MERGE apparently evolved in the primates that entertain these very musings. At issue is, in that sense, whether such an evolutionary event could have been along the open-ended lines of what our speculations in Section 5.2 suggested or, rather, deep algebraic considerations got engaged in relating thoughts from ***L*** into a focused ***WS***, where they could become externalized among conspecifics.

The question boils down to whether a mapping along the lines in Figure 7 is deeply algebraic (like the rest of the considerations we are the first to presume for LoT) or, instead, it was an evolutionary hack (which our subitizing-inspired proposal suggests).

For all their talk about normalization and coarse-graining, MCB assert something stronger: the "endocentricity" of phrases (the condition that a phrase preserves the categorial type of its head/nucleus)—a persistent category identity that is far from obvious in other natural fields for which (re-)normalization matters. This is why they algebraically

connect the Lie algebra in $\boldsymbol{L}$ to the $\boldsymbol{WS}$ Hopf algebra for recursive building, copying, and splitting of structures. The Hopf algebra stands in a *universal enveloping algebra* in relation to the underlying Lie algebra. In such terms, labels are preserved because $\boldsymbol{WS}$ is an algebraic expansion of $\boldsymbol{L}$, so the endocentricity is baked into the math. While this works, what seems perplexing here is the somewhat "anthropic" aspect it presumes.

For us, too, labels are crucial. We have an Axiom for labeling and an entire section on chain labels Section 3.5, which includes an algebraic definition of labels based on matrix determinants and normalized conditions. Monographs could be devoted to why a biological system should even exhibit labels—or what that really means, ontologically. Language is weird in that regard, as other biophysical systems involving "third factors" hardly seem to involve the sort of computational addressability or sheer pointing that labels appear to invoke: nothing short of a cognitive data-routing system. But note that our labels, given our system based on algebraic features (not just algebraic categories), are just matrix determinants. In that sense, our mechanism for relating labels (as determinants) in $\boldsymbol{A}_L$ and labels (also as determinants) in $\boldsymbol{WS}$ need not be anything as powerful as an algebraic translation: it could be, as suggested, a humbler outgrowth of subitizing that got lucky in evolution and, perhaps, could arise again in another species in the future.

The consequence of not invoking anything deeply algebraic—appealing, instead, to an evolutionary hack—raises no question as to why this putative evolutionary event would not have arisen with the rest of cognition or be widespread across animal species. Moreover, we admit not to want the relationship between a paradigmatic Lie algebra and whatever carries the syntagmatic algebra (be it a Hopf algebra or, as is the case in our proposal, a Fock lift within our assumptions about chains) to mess with algebraic dependencies bearing on commutativity. Our account clearly needs the non-commutativity of dependencies that rests at the core of non-classicality; by its very nature, the weak subitizing-style move suggested in Section 5.2 has no formal import in this regard.

Be that as it may, words change after being learned and constantly adopt nuance. Shoving this out of the algebraic workspace and relegating it to "performance" systems has a venerable tradition (Saussure too emphasized that diachronic linguistics presumes a synchronic understanding). Of course, taking dynamicity to the extent of letting the system "go quantum" (in relevant regards, no matter how restricted) is not the only viable theoretical move. We ourselves would have hardly gone to such extremes if we had better ways of treating the syntactic chains we have analyzed in this piece. That is one distinct area in which the MCB proposal and ours are making different, if related, moves.

MCB treats External MERGE as a product of generators, while it utilizes the Hopf coproduct for movement as IM, thus addressing how a single phrase can be syntactically evaluated in multiple places at once. In the MP version, MCB formulates assumptions about how Minimal Search (MS) limits such a process to "the top tier" of the workspace set. MS is challenged in IM, which is why vector spaces are invoked. In that formal space, a "top tier restriction" is modeled as an algebraic projection operator. If an element is buried deep in a phrase, the linear operators representing MERGE cannot interact with its coordinates without destroying the conservation properties of the vector space. The Hopf algebra utilizes Quotient Workspaces, so when IM occurs, the algebra treats occurrences as a singular mathematical entity mapped across a relational diagram.

Within the phenomenology of syntactic chains, our model and MCB make different predictions. Our logic allows for the realization of chain occurrences in UP *or* DOWN states (not both). Examples as in (23)/(28), Section 3.4 or (46) in Section 4.3, directly suggest that any treatment like MCB's, limited to some operator grafting the extracted piece to the root of phrase-markers during the IM process, is too strong. Both $\boldsymbol{PF}$ and $\boldsymbol{LF}$ must choose *one* site to execute the "displaced" phrase (usually phonetically UP in English, but sometimes DOWN in Spanish, and always DOWN in relevant instances in backwards-raising

languages like Adyghe). Knowing of the existence of such challenges—and rather than invoking anything like a quantum collapse—MCB hand-waves in the direction of unspecified filtering mechanisms of Externalization or Conceptual-Intentional systems. Readers can decide which of these moves has better testable consequences.

6. Conclusions

The present work founds a finite algebraic architecture in which lexical states, tensorized chain states, compression, spectral growth, and interface observability are related by explicit, enumerable constraints. The mathematical claims are finite and exhaustively checkable. The linguistic interpretation remains a program of empirical hypotheses. The principal open problem is not whether the algebra can be constructed, but whether its selective structure—the grammatical chains out of a much larger space of formal possibilities—corresponds to independently observable properties of language/cognition.

6.1. What Has Been Established

Our paper makes four contributions, which differ in epistemic status.

Constructed: We have defined the Chomsky-Pauli group we call $\boldsymbol{G_{CP}}$, a discrete group of 32 2 × 2 complex matrices, and constructed from it an Abelian Hilbert space $\boldsymbol{\mathcal{H}_2}$ serving as the algebraic model of the lexicon $\boldsymbol{L}$. The axioms governing selection (Axioms 1–6), tensorization (Axiom 7), chain formation (Axioms 8–10), and labeling (Axiom 6, revised) are stated explicitly and collectively define the formal system (Sections 2 and 3).

Proved by exhaustive enumeration: Of rank-1 sums $U = K + \Lambda$ for $K, \Lambda \in \boldsymbol{G_{CP}}$, 96 satisfy the compressibility condition—the resulting chain $\alpha \otimes$ U admits a write-back to a lexicon-compatible diagonal form. Of these 96, 32 yield four-term aggregates $M = U + \alpha + \Pi$ with spectral radius $\rho(M) = 2\varphi$, the spectral fingerprint of the X′-family. These counts are finite and exhaustively checkable (see Figure 9, Sections 3.5 and 3.6, Appendices A.4 and A.5).

Derived analytically: For admissible triples (U, α, Π), partial sum $M_3 = U + \alpha$ has spectral radius $\rho(M_3) = 1 + \sqrt{2}$ (silver mean); adding phrasal Π shifts this to $\rho(M) = 2\varphi$ for the 32 cases identified. Chain label is definable as the complex phase of a matrix determinant, yielding a group homomorphism from $\boldsymbol{G_{CP}}$ onto $\mathbb{Z}/4\mathbb{Z} \cong \{\pm 1, \pm i\}$ (Section 3.5, Appendix A.4).

Demonstrated by proof of concept: The sentence *a cat is curious in nature* has been analyzed within the framework, exhibiting a grammatical chain satisfying compressibility, spectral signature, and Hermiticity constraints and yielding the X′-equivalent aggregate $2\left[\begin{smallmatrix}1 & 1\\ 1 & 0\end{smallmatrix}\right]$ with $\rho = 2\varphi$ (Section 4.3). This demonstration shows that the formalism is applicable to actual linguistic data. Evidently, empirical validation of the framework requires systematic testing across a range of constructions, languages, and phenomena.

6.2. What the Architecture Suggests

The system outlines the PSP cycle: $\boldsymbol{L} \rightarrow \boldsymbol{A_L} \rightarrow \boldsymbol{WS} \rightarrow$ (chain compression) $\rightarrow \boldsymbol{L}$. Each arrow in this cycle has a different status, so distinguishing them is important.

Going from $\boldsymbol{L}$ to $\boldsymbol{A_L}$ is modeled as a Markovian selection constrained by the Jarret graph (Section 2.2, Figure 1), with subitizing-style incremental gauging as a cognitive implementation (Section 5.2). A hypothesis about selection order, this is not a derived result.

The move from $\boldsymbol{A_L}$ to $\boldsymbol{WS}$ is the Fock lift in the system (Section 3.2, Axiom 7): the tensorization of 2 × 2 $\boldsymbol{G_{CP}}$ elements into a 4 × 4 Fock-space sector generating superpositions of chain states. This is a mathematical construction, explicitly defined.

https://doi.org/10.3390/math14214037

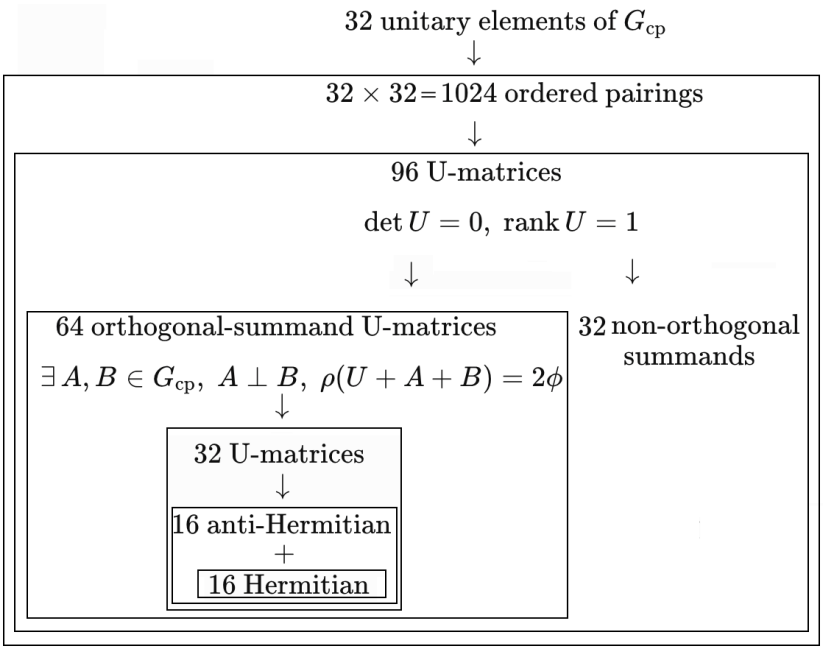


Figure 9. Flow chart illustrating the screening process of ***U***-matrices in Section 4.2.

Within ***WS***, three constraints jointly identify grammatically relevant chains:

A. *Compressibility* (rank-1 condition on $U = K + \Lambda$): which chains can return to a lexical format, a route from the Fock-lifted ***WS*** back to ***L***.
B. *Spectral condition* ($\rho(M) = 2\varphi$): which compressible chains carry the X′-type phrasal topology, linking the algebraic structure to the Fibonacci growth pattern observed in natural language phrase structure.
C. *Hermiticity* ($M = M^{\dagger}$): which chains are interface-observable, restricting grammatical chains to those whose aggregates have real eigenvalues and orthogonal eigenspaces (stable measurability at ***PF*** and ***LF***). (54)

These are independent constraints. The compression route from ***WS*** to ***L***—***a*** write-back—is established by the 96-matrix compressibility result. Whether this corresponds to any actual cognitive mechanism, including sleep-dependent memory consolidation, is a question the present formalism can only raise.

6.3. Open Problems and Extensions

The formal results in the model (Figure 10) raise some well-defined open problems.

Mathematically: An open problem is to derive the 32/96 sieve from first principles—to prove, rather than enumerate, why one-third of the compressible ***U***-matrices yield aggregates with $\rho = 2\varphi$. A proof should draw on the group-theoretic structure of the Abelian diagonal restriction within G_{CP} and spectral theory for the relevant 2×2 families. A related mathematical question is whether the Hermiticity constraint (the Phrasal Observation Postulate (42)) can be derived from the compressibility and spectral conditions. If it is independent, its justification might come from interface conditions at ***PF*** and ***LF***.

A*lgebraic extensions:* The present system is associative and non-commutative, with the general properties of a 4D algebra. Strict compositionality—uncontroversially required for human language—demands *non*-associativity, which our framework as stated does not natively provide. A natural algebraic extension is to replace the 2×2 G_{CP} matrices with 2×2 matrices whose entries are quaternions, effectively moving the system to an 8D non-associative algebra. Exploring this extension, and whether it changes the compressibility and spectral results, is a well-defined mathematical task.

Computationally: Automated enumeration of compressible chains, spectral verification, and cost-function optimization for the $A_L \rightarrow WS$ interlock ι are all computationally tractable and would allow the framework to be tested at scale. Of particular interest is how interlock reliability degrades as A_L size grows (or when awareness becomes challenged in any fashion)—and even whether probabilistic relaxations of the cost function can model known performance effects in sentence processing.

Neurocognitively: The formalism predicts that certain algebraic conditions—tensorization, separation, compression—correspond to distinct stages of linguistic computation. Whether these have identifiable neurophysiological correlates is an empirical question the model makes precise enough to test. Natural targets include whether the oscillatory dynamics of hippocampal or prefrontal networks during sentence comprehension reflect the Fock-lift structure; whether the (non-)Hermitian cut maps onto distinctions between observable and unobservable syntactic categories in brain recording data; and whether the compression route back to L is reflected in sleep-dependent lexical updating.

Conceptually: As such, present axioms do not distinguish between properties of cognition and those specific to language. Axiom 1 (the fundamental algebraic assumption) and Axiom 7 (tensorization) are plausible candidates for general cognitive properties present across species with sufficiently complex nervous systems. Axioms 4 (anti-symmetry of projection) and the Hermiticity postulate are better candidates for specifically linguistic properties. Articulating this distinction—what is and is not specific to the language faculty—and testing it against comparative cognitive data across model organisms are among the most important conceptual extensions of our framework.

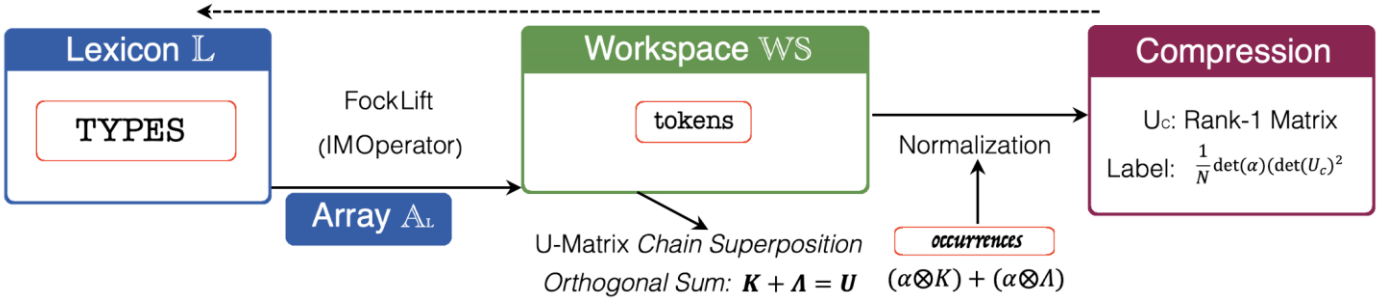


Figure 10. A visualization of the PSP cycle within the Matrix Syntax model.

Author Contributions: Conceptualization, R.O. and J.U.; Methodology, J.U.; Software, R.O.; Validation, R.O. and J.U.; Formal analysis, R.O. and J.U.; Investigation, J.U.; Writing—original draft, J.U.; Writing—review & editing, R.O.; Supervision, J.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: This work is dedicated to the memory of Roger Martin. We appreciate useful comments from the reviewers of this journal, as well as Drew Baden, Mostafa Barzegar, Tom Bever, Jerome Busemeyer, Victor Chamot, Jackson Confer, Laurent Dekydtspotter, Naomi Feldman, Ángel Gallego, Victor Galitski, Andrew Hanson, Harry van der Hulst, Bill Idsardi, Srini Iyengar, Mike Jarret, Phil Larkin, Dave Medeiros, Mithün Paul, Larry Moss, Elliot Murphy, Kharen Musaelian, Paul Pietroski, Rodrigo Treviño, Yutaro Ueno, Jesús Ugalde, and (always) the late Roger Martin, who was part of our team until his untimely departure. None of them should be responsible for any of our errors. Both authors have used freely available LLMs for help with English, references,

diagrams, and the validity of non-technical inferences, as well as The MathWorks, Inc. "MATLAB" [https://www.mathworks.com] for specific formal testing.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

We finally address some technical derivations our results rest on. They are based on linear algebra concepts and a more detailed analysis of our work with Roger Martin [79].

Appendix A.1. From Chomsky Matrices to Hilbert Spaces

Chomsky matrices, together with the Identity matrix I, form an abelian group (considering as operation matrix multiplication). This is given by

$$G_8 = \{\pm I, \pm Z, \pm C_1, \pm C_2\},$$

with

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, C_1 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}, C_2 = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}.$$

The group is abelian (commutative) because all matrices are diagonal and, therefore, they mutually commute. It is easy to check that this set satisfies the usual mathematical properties of a group: existence of a neutral element I and an inverse for every element. The group is also generated by the repeated multiplications of C_1, or of C_2.

The abelian group can be expanded to a non-abelian group by inclusion of the 2 × 2 Paulis. We call the resulting group the "Chomsky Pauli Group" ($\boldsymbol{G_{CP}}$):

$$G_{cp} = \{\pm I, \pm X, \pm Y, \pm Z, \pm iI, \pm iX, \pm iY, \pm iZ, \pm C_1, \pm C_2, \pm S_1, \pm S_2, \pm iC_1, \pm iC_2, \pm iS_1, \pm iS_2\},$$

where I is the 2×2 identity matrix, X, Y, Z are the 2×2 Paulis, C_1, C_2 are two of the original Chomsky matrices, and (to fix notation) S_1 and S_2 are "anti-Chomsky" matrices:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

$$C_1 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}, C_2 = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, S_1 = \begin{bmatrix} 0 & 1 \\ -i & 0 \end{bmatrix}, S_2 = \begin{bmatrix} 0 & 1 \\ i & 0 \end{bmatrix}.$$

We call this the Chomsky-Pauli group because it contains the original Chomsky matrices and the Paulis X, Y, Z. The new group includes the Pauli group as a subgroup, as shown below. The determinants of all the 32 matrices are either ±1 or ±i, matching the fact that there are four lexical categories indexed by their labels, computed as determinants.

Let us now build a Hilbert space from the matrices in the group. We consider the matrices as abstract *vectors;* i.e., a vector space where vectors are represented by matrices. The scalar product between matrices A, B, thought as vectors, is given, for instance, by:

$$\langle A | B \rangle = \mathrm{tr}(A^{\dagger}B).$$

For convenience, we use the Dirac (bra-ket) notation from quantum mechanics. Matrix A is represented as $|A\rangle$, its adjoint $A^{\dagger} = (A^T)^*$ is represented by $\langle A|$ (T is the transpose, which exchanges rows by columns, and $*$ is the complex conjugate, which substitutes $i \leftrightarrow -i$), and the scalar product of A and B is represented as $\langle A|B\rangle$. The scalar product produces a single complex number out of the two matrices. The product provides a way to define *angles* between vectors in the space. For orthogonal vectors, this product is zero.

We are to build a vector space out of matrices, which can themselves be understood as elements of an abstract vector space. Whenever a scalar product is zero, we say that the corresponding matrices—as abstract vectors—are *mutually orthogonal*, in the sense that they are "perpendicular" to one another. For instance, one can check how I and Z are mutually orthogonal, i.e., $\langle I|Z\rangle = 0$. The *norm* N of a vector $|A\rangle$ (matrix A) is given by:

$$N = \left|\sqrt{\langle A|A\rangle}\right| = \left|\sqrt{\mathrm{tr}(A^{\dagger}A)}\right|$$

For a *normalized* vector, $N = 1$; for instance (where the "hat" means "normalized"):

$$|\hat{A}\rangle = \frac{1}{\sqrt{\langle A|A\rangle}}\,|A\rangle.$$

For whichever matrix pairs are mutually orthogonal when taking their tensor products $A \otimes B$, as relevant to our modeling of syntactic chains, one can observe how:

$$<A \otimes B \,|\, C \otimes D> = \mathrm{tr}(A^{\dagger}C)\ \mathrm{tr}(B^{\dagger}D).$$

We then proceed with the construction of the Hilbert space associated to $\boldsymbol{G_{cp}}$, call it $\boldsymbol{\mathcal{H}_{CP}}$. Pairs of orthogonal matrices A, B in $\boldsymbol{G_{cp}}$ obey at least one of the following constraints:

1. A is diagonal, B antidiagonal, or vice-versa (i.e., A antidiagonal, B diagonal).
2. A is from the Pauli group, B is also from that group, except $A = \pm B$, $A = \pm iB$.
3. A is not from the Pauli group, neither is B from that group, except $A = \pm B$, $A = \pm iB$.

All the scalar products between elements of the group can be derived from Table A1. Rows correspond to the matrix for which there is a † in the scalar product formula (the "bra"), and columns for those with no † (the "ket"). Products between other elements in G_{cp} are also given by the numbers in the table, up to possible ±1 and ±i multiplying factors.

Table A1. Table of scalar products of some elements in the G_{cp}.

	I	X	Y	Z	C_1	C_2	S_1	S_2
I	2	*0*	0	0	$1-i$	$1+i$	0	0
X	0	2	0	0	0	0	$1-i$	$1+i$
Y	0	0	2	0	0	0	$-1+i$	$1+i$
Z	0	0	0	2	$1+i$	$1-i$	0	0
C_1	$1+i$	0	0	$1-i$	2	0	0	0
C_2	$1-i$	0	0	$1+i$	0	2	0	0
S_1	0	$1+i$	$-1-i$	0	0	0	2	0
S_2	0	$1-i$	$1-i$	0	0	0	0	2

By inspecting Table A1, we see how the set of scalar products of elements in $\boldsymbol{G_8}$ is a subset of those in $\boldsymbol{G_{cp}}$—a consequence of $\boldsymbol{G_8}$ being a subgroup. Moreover, we can see, from our definition of scalar product, that these tables provide the dimensionality as well as possible choices of basis of a vector space constructed by the elements in the group.

Next, we find the dimensionality and bases for $\boldsymbol{\mathcal{H}_{CP}}$. Table A1 shows $\boldsymbol{\mathcal{H}_{CP}}$'s dimension: $\dim(\mathcal{H}_{CP}) = 4$. This is because in $\boldsymbol{G_{cp}}$ there are at most 4 linearly independent matrices, e.g., the three Paulis together with the identity {I, *X*, *Y*, *Z*}, but also the four Chomskys {C_1, C_2, S_1, S_2}. Notice how those can be written in terms of the Paulis and the identity:

$$C_1 = \left(\frac{1-i}{2}\right) I + \frac{1+i}{2}\, Z, S_1 = \left(\frac{1-i}{2}\right)(X-Y),$$

$$C_2 = \left(\frac{1+i}{2}\right) I + \frac{1-i}{2}\, Z, S_2 = \left(\frac{1+i}{2}\right)(X+Y).$$

These relations can be inverted to obtain the Paulis from the Chomskys:

$$\mathrm{I} = \left(\frac{1+i}{2}\right) C_1 + \frac{1-i}{2}\, C_2, X = \left(\frac{1+i}{2}\right) S_1 + \left(\frac{1-i}{2}\right) S_2,$$

$$\mathrm{I} = \left(\frac{1-i}{2}\right) C_1 + \frac{1+i}{2}\, C_2, Y = -\left(\frac{1+i}{2}\right) S_1 + \left(\frac{1-i}{2}\right) S_2,$$

As to orthonormal bases, just as it was the case for H_8, there are two natural choices for $\boldsymbol{\mathcal{H}_{CP}}$. One is given by the "normalized" Paulis plus the identity:

$$\text{Pauli basis: } \left\{\frac{I}{\sqrt{2}}, \frac{X}{\sqrt{2}}, \frac{Y}{\sqrt{2}}, \frac{Z}{\sqrt{2}}\right\} \equiv \{|\hat{I}\rangle, |\hat{X}\rangle, |\hat{Y}\rangle, |\hat{Z}\rangle\},$$

where we use the Dirac notation. So, any vector in $|\psi\rangle \in \mathcal{H}_{CP}$ can be written as:

$$|\psi\rangle = \alpha |I\rangle + \beta |X\rangle + \gamma |Y\rangle + \delta |Z\rangle,$$

where α, β, γ, and δ are complex numbers—the "coordinates" of vector $|\psi\rangle$ in this Pauli basis. The second natural option is given by "normalized" Chomsky matrices:

$$\text{Chomsky basis: } \left\{\frac{C_1}{\sqrt{2}}, \frac{C_2}{\sqrt{2}}, \frac{S_1}{\sqrt{2}}, \frac{S_2}{\sqrt{2}}, \right\} \equiv \{|\widehat{C_1}\rangle, |\widehat{C_2}\rangle, |\widehat{S_1}\rangle, |\widehat{S_2}\rangle\}$$

This is an equally valid basis, and any vector in $\boldsymbol{\mathcal{H}_{CP}}$ can be written as:

$$|\psi\rangle = \alpha' |\widehat{C_1}\rangle + \beta' |\widehat{C_2}\rangle + \gamma' |\widehat{S_1}\rangle + \delta' |\widehat{S_2}\rangle,$$

where α', β', γ' and δ' are the "coordinates" of vector $|\psi\rangle$ in this Chomsky basis.

The relation between these two bases is the following unitary matrix:

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = \begin{pmatrix} \frac{1-i}{2} & \frac{1+i}{2} & 0 & 0 \\ 0 & 0 & \frac{1-i}{2} & \frac{1+i}{2} \\ 0 & 0 & \frac{-1+i}{2} & \frac{1+i}{2} \\ \frac{1+i}{2} & \frac{1-i}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \\ \delta' \end{pmatrix}.$$

The above matrix-vector multiplication gives the rule to change from "Chomsky coordinates" to "Pauli coordinates" for any vector in $\boldsymbol{\mathcal{H}_{CP}}$. Both "coordinate systems" are valid as far as the description of the space is concerned, and other bases of $\boldsymbol{\mathcal{H}_{CP}}$ are possible.

Note how Hilbert space $\boldsymbol{\mathcal{H}_8}$ is a 2D subspace of $\boldsymbol{\mathcal{H}_{CP}}$, just as G_8 is a subgroup of G_{CP}.

Appendix A.2. Unit (U) Matrices and Chain Formation

Within the scope of our model, states without chain conditions are of the kind

$$|\text{Regular}\rangle = |\alpha\rangle \otimes |K\rangle.$$

While (allowing for possible coefficients a, b), chain states are of the following kind:

$$|\text{Chain}\rangle = {}^{1}/_{\sqrt{2}} \left(a |\hat{\alpha}\rangle \otimes |\hat{\mathrm{K}}\rangle + b |\hat{\alpha}\rangle \otimes |\hat{\Lambda}\rangle\right) = {}^{1}/_{\sqrt{2}} |\hat{\alpha}\rangle \otimes \left(a|\hat{\mathrm{K}}\rangle + b|\hat{\Lambda}\rangle\right).$$

In the context of our construction, mathematically these are separable states: those that have no entanglement in the quantum mechanical jargon.

Next, consider which chain constructions are valid in our Hilbert space $\boldsymbol{\mathcal{H}_{CP}}$. We take chains to obey the following constraints (which may follow from deeper conditions):

1. Chains must be compressible back to 2 × 2 matrices.
2. A resulting 2 × 2 matrix must have label (complex phase of the determinant) ±1, ±i.
3. The chain is formed by summing perpendicular vectors in $\boldsymbol{\mathcal{H}_{CP}}$ (i.e., in the superposition of orthogonal contexts).

By examining all possible cases, it is easy to see that there are 96 combinations $U = A + B$, with $A, B \in G_{CP}$, $A \neq B$, such that U has one non-zero eigenvalue only. We call these 96 combinations *unit (**U**-)matrices.* In terms of such matrices, a chain reads as follows:

$$\text{Chain} = \alpha \otimes A + \alpha \otimes B = \alpha \otimes (A + B) = \alpha \otimes U.$$

The above matrix is compressible back to a 2 × 2 matrix, since α is 2 × 2 and U is compressible to a 1 × 1 diagonal matrix (implying it is proportional to a projector). Let us call these compressed 1 × 1 matrices U_c (for "compressed U"). The label of the chain formed by the **U**-matrices as context, after compression (i.e., with U_c), follows the usual rules of the determinant of a tensor product and, importantly, a *rescaling* by the norm.

To see why this constraint forces a unique non-zero eigenvalue, recall that $\alpha \in G_{CP}$ always has two non-zero eigenvalues, since $\det(\alpha) = \pm 1, \pm i$ for any element of the group.

https://doi.org/10.3390/math14214037

The eigenvalues of $\alpha \otimes (K + \Lambda)$ are the four pairwise products of the eigenvalues of α and $K + \Lambda$. For the 4 × 4 chain matrix to be compressible to a non-trivial 2 × 2 form, via projection onto its non-null eigenspace, we require the resulting matrix to have two non-zero eigenvalues. This can happen only if $K + \Lambda$ itself has exactly one non-zero eigenvalue—equivalently, only if $K + \Lambda$ is rank 1. By exhaustive enumeration over ordered pairs $A, B \in G_{CP}$ with $A \neq B$, exactly 96 sums $U = A + B$ yield a rank-1 matrix. These are the $\boldsymbol{U}$-matrices.

The 96 $\boldsymbol{U}$-matrices partition into three subsets of 32 elements each, distinguished by the orthogonality of the contexts being summed and by the spectral character of the resulting U (as listed in Figures A1–A3 at the end of Appendix A.5):

1. Set 1 (32 matrices): sums of mutually orthogonal contexts in $\boldsymbol{\mathcal{H}_{CP}}$. Therein, 16 matrices are Hermitian (context label $(\det(U_c))^2/N = +1$); 16, anti-Hermitian (context label −1). Each matrix in Set 1 has a unique non-zero eigenvalue proportional to ±1, ±i.
2. Set 2 (32 matrices): sums of mutually orthogonal contexts whose unique non-zero eigenvalue is proportional to (±1 ± i). The corresponding context label is ±i.
3. Set 3 (32 matrices): sums of *non-orthogonal* contexts. All matrices therein are unitary diagonal, with unique non-zero eigenvalue proportional to (±1 ± i), context label ±i.

The 64 $\boldsymbol{U}$-matrices of Sets 1 and 2 are formed by sums of orthogonal contexts, while the 32 of Set 3 involve non-orthogonal contexts in violation of constraint 3 above. Of the 64 admissible $\boldsymbol{U}$-matrices in Sets 1 and 2, only the 32 of Set 1 are Hermitian or anti-Hermitian. The 16 combinations available within $\boldsymbol{G_8}$ alone are a particular case of the 64 admissible chain contexts that combine orthogonality with rank-1 compressibility. Within Set 1, the Hermitian $\boldsymbol{U}$-matrices play a privileged role: Hermitian operators have real eigenvalues and orthogonal eigenspaces, corresponding in physics to observables whose values can be measured. In our linguistic setting, hermiticity emerges as a stronger condition than rank-1 compressibility alone, singling out the $\boldsymbol{U}$-matrices whose chain-aggregate projects can be observed at the ***PF*** and ***LF*** interfaces. We return to this in Appendix A.5.

A chain label after compression follows the rules of a tensor product determinant, with a rescaling by the norm. For $\alpha \in G_{CP}$ with $|\det(\alpha)| = 1$, U_c the 1 × 1 compression of U:

$$\text{Label(Chain)} = (1/N)\ \det(\alpha \otimes U_c) = (1/N)\ \det(\alpha)\ (\det(U_c))^2,$$
$$\text{for } N = |\det(\alpha)\ (\det(U_c))^2| = |\det(U_c)|^2.$$

The phase in the complex plane of $\det(\alpha)\ (\det(U_c))^2$ is the chain label, while the norm of the quantity is the rescaling factor N. Physically, this amounts to the rescaling of variables after a coarse graining of linguistic information, completing a renormalization step [42]. The chain label is thus the specifier label times the square of the U_c label, with a rescaling by its norm—motivating the formulation of Axiom 6 (Labeling) in (32) of the main text.

Without compression, the $\boldsymbol{U}$-matrices never form a group since the 2 × 2 identity is not rank 1. After compression, however, restricted subsets of $\boldsymbol{U}$-matrices do form groups. In particular, sets of four $\boldsymbol{U}$-matrices sharing the same eigenvector (e.g., {$I + X$, $-I - X$, $iI + iX$, $-iI - iX$}) generate, via tensor product with α-factors, the restricted-chain group $G_{r\text{-chain}} = G_{CP} \otimes \mathbb{Z}_4$, where the compressed U_c matrices, four in total, act as a 1 × 1 representation of $\mathbb{Z}_4$. Such restricted r-chains constitute candidate workable chains.

The 2D Hilbert subspace spanned by the diagonal ($\boldsymbol{G_8}$) $\mathbf{G_{CP}}$ portion is isomorphic to $\mathbb{C}^2$, the qubit space. The orthonormal Pauli basis {$I/\sqrt{2}$, $Z/\sqrt{2}$} (Chomsky basis {$C_1/\sqrt{2}$, $C_2/\sqrt{2}$}) provides a unitary identification with the standard $\mathbb{C}^2$ basis. The 4D space $\boldsymbol{\mathcal{H}_{CP}}$ decomposes as $\mathbb{C}^2 \oplus \mathbb{C}^2$: one qubit corresponds to the $\boldsymbol{G_8}$-subspace (diagonal/lexical sector); the other, to the antidiagonal/functional sector. This realizes our claim that the $\boldsymbol{L}$ Hilbert space $\boldsymbol{\mathcal{H}_2}$ is isomorphic to $\mathbb{C}^2$, while the Chomsky-Pauli ambient space underlying workspace ***WS*** is $\mathbb{C}^2 \oplus \mathbb{C}^2$, the Fock-lift adding graded sectors from $\boldsymbol{\mathcal{H}_4}$ upward, as discussed next.

https://doi.org/10.3390/math14214037

Appendix A.3. Fock Space from Hilbert Space

Let $\boldsymbol{\mathcal{H}}$ be a complex separable Hilbert (*one-particle)* space. The construction of the associated Fock space is a graded Hilbert space built from tensor powers of $\boldsymbol{\mathcal{H}}$, as follows.

For each integer $n \geq 1$, we may denote by $\mathcal{H}^{\otimes n}$ the n-fold Hilbert tensor product of $\boldsymbol{\mathcal{H}}$. We adopt the convention $\mathcal{H}^{\otimes 0} := \mathbb{C}$, whose canonical unit vector $1 \in \mathbb{C}$ is identified with the vacuum vector. Each $\mathcal{H}^{\otimes n}$ carries the natural Hilbert structure induced by the tensor product, characterized on tensors (extended by linearity and completion) by:

$$\langle f_1 \otimes \cdots \otimes f_n, g_1 \otimes \cdots \otimes g_n \rangle = \prod_{k=1}^{n} \langle f_k, g_k \rangle_{\mathcal{H}},$$

The (unsymmetrized) Fock space over $\boldsymbol{\mathcal{H}}$ is defined as the Hilbert direct sum:

$$F(\mathcal{H}) := \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n}.$$

This defines a $\mathbb{N}_0$-graded Hilbert space:

$$F(\mathcal{H}) = \bigoplus_{n=0}^{\infty} F^{(n)},$$

where $F^{(n)} := \mathcal{H}^{\otimes n}$. An element $\Psi \in F(\mathcal{H})$ is a sequence $\Psi = (\psi_0, \psi_1, \psi_2, \ldots)$, with $\psi_n \in \mathcal{H}^{\otimes n}$, such that $||\Psi||^2 = \sum_{n=0}^{\infty} ||\psi_n||^2 < \infty$.

The grading is implemented by the number operator N defined by $N|_{F(n)} = n\,\mathrm{id}$, which is a densely defined self-adjoint operator on $F(\mathcal{H})$. In applications to identical particles, one restricts each *n*-particle sector to the appropriate symmetry class under permutations. Let $\boldsymbol{S_n}$ denote the symmetric group on *n* letters. In turn, for each $\sigma \in S_n$, let $U_\sigma : \mathcal{H}^{\otimes n} \to \mathcal{H}^{\otimes n}$ be the unitary operator defined by permutation of tensor factors. Then in the paradigmatic case, the symmetric subspace of $\mathcal{H}^{\otimes n}$ is defined as

$$\mathrm{Sym}^n(\mathcal{H}) = \{ \psi \in \mathcal{H}^{\otimes n} \} \mid U_\sigma \psi = \psi \text{ for all } \sigma \in S_n \}.$$

Equivalently, $\mathrm{Sym}^n(\mathcal{H}) = P_{\mathrm{sym}(n)}\, \mathcal{H}^{\otimes n}$, with:

$$P_{\mathrm{sym}(n)} = (1/n!) \sum_{\{\sigma \in Sn\}} U_\sigma$$

The orthogonal projection onto the symmetric subspace. The Fock space over $\boldsymbol{\mathcal{H}}$ is:

$$F_s(\mathcal{H}) = \bigoplus_{n=0}^{\infty} \mathrm{Sym}^n(\mathcal{H}).$$

In the syntagmatic case, the anti-symmetric subspace of $\mathcal{H}^{\otimes n}$ is:

$$\wedge^n \mathcal{H} = \{ \psi \in \mathcal{H}^{\otimes n} \mid U_\sigma \psi = \mathrm{sgn}(\sigma)\, \psi \text{ for all } \sigma \in S_n \}.$$

Equivalently, $\wedge^n \mathcal{H} = P_{\mathrm{asym}(n)}\, \mathcal{H}^{\otimes n}$, where:

$$P_{\mathrm{asym}(n)} = (1/n!) \sum_{\{\sigma \in Sn\}} \mathrm{sgn}(\sigma)\, U_\sigma.$$

The syntagmatic Fock space is, then, defined as:

$$F_a(\mathcal{H}) = \bigoplus_{n=0}^{\infty} \wedge^n (\mathcal{H}).$$

The algebraic tensor algebra over $\mathcal{H}$, $T(\mathcal{H}) = \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n}$, is a graded algebra under tensor multiplication. Paradigmatic and syntagmatic constructions each correspond to the symmetric and the exterior algebra over $\boldsymbol{\mathcal{H}}$. Our Fock spaces are the Hilbert completions of these graded algebras with respect to the inner product induced from $\boldsymbol{\mathcal{H}}$. The Fock construction may be viewed as the passage from a one-particle Hilbert space to the Hilbert completion of its free graded tensor algebra, within symmetry constraints.

Appendix A.4. Compressibility, Write-Back to L, and the U-Matrix Span

The central asymmetry of the system is between the paradigmatic sector, where lexical items live as 2 × 2 unitary diagonal operators in $\boldsymbol{\mathcal{H}_8}$, and the syntagmatic sector emerging in workspace ***WS***, where Internal MERGE (IM) tensorizes elements into 4 × 4 superpositions in $\mathcal{H}_{CP} \subset F(\mathcal{H})$. Important to the system is that not every 4 × 4 chain

configuration can be projected back to the $\boldsymbol{\mathcal{H}_8}$ paradigmatic format: only those whose contexts admit $\boldsymbol{U}$-matrix compression (per A.2) yield states that *write back* to $\boldsymbol{L}$ after IM.

To gauge the algebraic rarity of the write-back property, consider the space of all ordered pairs of 4 × 4 Kronecker tensor sums of the form $\alpha \otimes K + \alpha \otimes \Lambda$, where $\alpha, K, \Lambda \in G_{CP}$. There, the 4 × 4 ambient space contains 65,536 ordered tensor-sum candidates when one ranges exhaustively over $\boldsymbol{G}_{CP}$ entries in each block (256^2 admissible quadrant configurations). Of these, exhaustive computer-assisted inspection reveals that only 96 contain a 2 × 2 sub-block that lies in the linear span of the $\boldsymbol{U}$-matrices identified in A.2.

That rarity has an algebraic explanation: a 2 × 2 block of $\alpha \otimes (K + \Lambda)$ belongs to the U-span if and only if $K + \Lambda$ is itself rank-1 (i.e., U-shaped); and rank-1 sums of distinct $\boldsymbol{G}_{CP}$ elements are the 96 $\boldsymbol{U}$-matrices of A.2. After projection onto the non-null eigenspace and spectral-radius normalization (the rescaling $N = |\det(U_c)|^2$ of A.2), the surviving 2 × 2 sub-blocks are diagonal forms with entries in $\{\pm 1, \pm i\}$—exactly the four lexical labels of $\boldsymbol{L}$.

We may state that result as the following:

Compression Postulate: For any $|\text{Chain}\rangle = (1/\sqrt{2})\,|\alpha\rangle \otimes (|K\rangle + |\Lambda\rangle)$, $K + \Lambda$ a $\boldsymbol{U}$-matrix, the compressed image lies in the diagonal $\boldsymbol{L}$-compatible subspace of $\boldsymbol{\mathcal{H}_8}$. Conversely, no chain whose context-sum $K + \Lambda$ lies outside the U-span admits write-back to $\boldsymbol{L}$.

We interpret this as the algebraic substrate of the $\boldsymbol{L \rightarrow (A_L \rightarrow) WS \rightarrow L}$ cycle: there is an route from a *syntagmatic chain* in the Fock-lift back to a *paradigmatic lexicon* item in $\boldsymbol{\mathcal{H}_8}$, governed by the rank-1 condition on context sums. Most syntactic configurations IM could generate are not compressible and, therefore, do not project back to long-term storage. Only those with the $\boldsymbol{U}$-matrix structure survive the cycle, supporting the projection to phrasal topology (Appendix A.5) and the return of enriched information to $\boldsymbol{L}$.

Are the surviving $\boldsymbol{U}$-matrices privileged because they admit compression back to $\boldsymbol{L}$, or because their four-term aggregates can project to phrasal topologies (Appendix A.5)? The conditions are formally independent—compressibility is a rank-1 condition on $K + \Lambda$, whereas the spectral signature $\rho(M) = 2\varphi$ is a condition on the four-term sum $U + \alpha + \Pi$. Yet, empirically, they co-vary: the same 96 $\boldsymbol{U}$-matrices that compress are those from which the Fibonacci spectral fingerprint can emerge, and the further restriction to the 32 Hermitian elements of Set 1 yields the configurations that are observable at the interpretive interfaces. We suspect this co-variation is not coincidental. Recurrence of hermiticity in the Set 1 write-back configurations and in the observability conditions at $\boldsymbol{PF}$/$\boldsymbol{LF}$ (per (42) of Section 4.2) raises the possibility that a common selectional principle may be provided by said property, simultaneously enabling write-back and supporting phrasal projection.

From that perspective, neither compressibility nor phrasal topology is logically prior: both are facets of the same (anti-)Hermitian sub-structure within $\boldsymbol{G}_{CP}$. It remains an open question whether this represents a deep algebraic coincidence or a derivable theorem about the interplay between rank-1 sums and the Fibonacci-spectral seed matrix.

A further filter stems from hermiticity (Section 4.2, Appendix A.2): the 32 (anti-)Hermitian $\boldsymbol{U}$-matrices of Set 1 (Figure A1) correspond to the contexts that not only compress, but whose four-term aggregates are observed at the interfaces. Within the 96 compressible contexts, elements of Set 1 are distinguished by their (anti-)Hermiticity and compatibility with the orthogonality condition imposed on chain formation. These appear most directly relevant to grammatical configurations, including the predicational chains in Section 4.

Appendix A.5. Spectral Radius of Phrasal Aggregates and the X′-Topology

A Fibonacci-style signature emerges in a subset of phrasal aggregates formed from compressible chains. As described in Section 4.2, given a compressible chain with contexts K, Λ (with $K + \Lambda = U$, *a* $\boldsymbol{U}$-matrix), linker α, and projection Π, the *aggregate* is:

$$M = (K + \Lambda) + \alpha + \Pi,$$

Defined as a sum within the 2 × 2 matrix space of $\boldsymbol{G}_{CP}$. This is a direct matrix sum—the aggregate capturing the additive arithmetic content of the participating elements as 2 × 2 matrices, relevant information for the chain to compress back to the $\boldsymbol{\mathcal{H}_8}$ format.

Consider the partial sum:

$$M_3 = (K + \Lambda) + \alpha = U + \alpha.$$

By inspection over admissible (U, α) pairs from Set 1 of the $\boldsymbol{U}$-matrices (Figure A1), a non-trivial subset of these triples yields a 2 × 2 matrix with spectral radius $\rho(M_3) = 1 + \sqrt{2}$.

To see that, take the example (35a) of the text: $U = Y + I$, $\alpha = Z$. Then:

$$U = Y + I = \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}, \quad \alpha = Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad M_3 = U + \alpha = \begin{bmatrix} 2 & -i \\ i & 0 \end{bmatrix}.$$

M_3's characteristic polynomial is $\lambda^2 - 2\lambda - 1 = 0$, roots $\lambda = 1 \pm \sqrt{2}$. The spectral radius is, then, $\rho(M_3) = 1 + \sqrt{2}$, the *silver mean*—the dominant eigenvalue of matrix $\left[\begin{smallmatrix} 2 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. When the projection Π is added to complete aggregate M above, one finds (for a structured subset of inputs drawn from Set 1 in Figure A1) that the spectral radius jumps to $\rho(M) = 2\varphi$, where $\varphi = (1 + \sqrt{5})/2$ is the *golden mean*. Continuing with example (35a):

$$M = U + \alpha + \Pi = \begin{bmatrix} 2 & -i \\ i & 0 \end{bmatrix} + Y = \begin{bmatrix} 2 & -2i \\ 2i & 0 \end{bmatrix}.$$

M's characteristic polynomial is $\lambda^2 - 2\lambda - 4 = 0$, roots $\lambda = 1 \pm \sqrt{5}$. Hence $\rho(M) = 1 + \sqrt{5} = 2\varphi$.

By enumeration over admissible (U, α, Π) quadruples for $\boldsymbol{U}$-matrices within Set 1 (Figure A1), one finds that 32 out of the 96 $\boldsymbol{U}$-matrices give rise to four-term aggregates with $\rho(M) = 2\varphi$. This ½ the $\boldsymbol{U}$-matrices is the source of the empirical claim that 2φ is the spectral fingerprint of the Fibonacci family within the model. This proportion might reflect a deeper symmetry of the Hermitian sub-sieve (cf. the open questions in Section 6.3).

Value 2φ is the dominant eigenvalue of the 2 × 2 seed matrix:

$$M_{\text{seed}} = 2 \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix},$$

Whose powers track the third row of the Wythoff array of Fibonacci recurrences [80]; by (38) Section 3.6, this 2, 4, 6, 10, 16, 26 … is the doubled Fibonacci sequence, which—by (39)—coincides with maximal expansion counts for the X′-recurrence (Lindenmayer—→ *h*, + and + → +, −). The 3 × 3 incidence matrix of that rewriting has spectral radius $\rho = \varphi$.

$$X' = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The key algebraic identity appears to be M_{seed} above. By the standard scalar-multiple property of primitive non-negative matrices [63], scaling preserves eigenvectors, scaling the dominant eigenvalue by that factor. The dominant eigenvalue of this M_{seed} is, hence, 2φ, matching the spectral radius of the four-term aggregate M when $\boldsymbol{U}$-matrix inputs are drawn from Set 1 (Figure A1). In other words: aggregating four $\boldsymbol{G}_{CP}$ unitary matrices—where the contexts sum to a $\boldsymbol{U}$-matrix, the linker is an admissible specifier, and the projection is (anti-)Hermitian—we obtain a 2 × 2 matrix whose spectral fingerprint is identical to that of the X′-theoretic seed driving Fibonacci-style phrasal expansion. This is meant as an algebraic justification for the claim in Section 4.3 that compressible chain quadruples are precursors of the X′-topology. A non-trivial issue is that we are dealing with 2 × 2 matrices without constants (in the Lindenmayer sense) in one instance, whereas in the other they are 3 × 3 matrices *with* constants. Without constants, a formal language cannot be specified as the expansions do not terminate, but we leave this for future research.

While $\rho(M) = 2\varphi$ is necessary for the aggregate to project to the X′-topology, it is not sufficient for *observability*. Per (42) of Section 4.2, observable phrasal aggregates must additionally be Hermitian. To illustrate, consider the contrast (35a) vs. (35b) of the main text:

$$M_a = [YPredP + IAP] + ZNP + YPredP = \begin{bmatrix} 2 & -2i \\ 2i & 0 \end{bmatrix} \quad \text{(Hermitian, } \rho = 2\varphi\text{)}$$

$$M_b = [iC_1 + iC_2] + iXTP + iXTP = \begin{bmatrix} 2i & 2i \\ 2i & 0 \end{bmatrix} \quad \text{(non-Hermitian, } \rho = 2\varphi\text{)}$$

Both aggregates have the Fibonacci spectral signature $\rho = 2\varphi$, but only M_a is Hermitian: $M_a^\dagger = M_a$, since the off-diagonal entries are conjugate pairs $(-2i, 2i)$. Matrix M_b, by contrast, has off-diagonal entries (i, i) that are not conjugates of one another ($M_b \neq M_b^\dagger$).

Therefore, the hermiticity filter selects, among the 32 $\boldsymbol{U}$-matrices yielding $\rho = 2\varphi$, those whose four-term M aggregate is itself Hermitian. In the linguistic translation of (35a) and (35b), this corresponds to the empirically attested distinction between standard NP-AP-PredP predicational chains and, e.g., unattested "TP-of-TP" configurations. As noted in Section 3.1—while this type of result must be generalized to all possible combinations arising in the system—this is an empirical commitment of the present model.

The complete set of constraints on phrasal aggregates thus boil down to:

(i) Compressibility: $K + \Lambda = U$ is rank 1 (96 admissible contexts; A.2, A.4).

(ii) Spectral signature: $\rho(M) = 2\varphi$ for $M = U + \alpha + \Pi$ (1/3 of the 96 contexts = 32).

(iii) Hermiticity: $M = M^\dagger$ (further reduction via open combinatorial work as in 3.5)

Conditions (i)–(iii) define X′-projecting, interface-observable chain configurations within our model. Constraint (i) by itself is the algebraic write-back condition supporting the ***L*** → ***WS*** → ***L*** cycle (A.4); (ii) is the topological growth condition supporting the X′-recursion; (iii) is the observability condition supporting the ***PF***/***LF*** collapse (Born rule, Axiom 9). Future work should establish whether these constraints, which conspire to produce configurations observed in natural-language phrase structure—can reduce further or be unified. They emerge from the linear-algebraic structure of $\boldsymbol{G}_{CP}$ alone, within the confines of the axioms in Sections 2 and 3. Future work should also analyze how many of those follow from general conditions of animal thought and which, instead, evolved only for humans, as we managed to "squeeze" said thoughts into spectral observables. Meanwhile, we state:

Proposition: Among the 96 admissible U-matrices, exactly 32 yield four-term aggregates $M = U + \alpha + \Pi$ with spectral radius $\rho(M) = 2\varphi$. Of these, the subset satisfying Hermiticity $M = M^\dagger$ corresponds to predicational chains attested in natural language.

https://doi.org/10.3390/math14214037

1	$\mathbb{I}+X$ $\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$	$\mathbb{I}+Y$ $\begin{pmatrix}1 & -i\\ i & 1\end{pmatrix}$	$\mathbb{I}+Z$ $\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix}$	$\mathbb{I}-X$ $\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$
1	$\mathbb{I}-Y$ $\begin{pmatrix}1 & i\\ -i & 1\end{pmatrix}$	$\mathbb{I}-Z$ $\begin{pmatrix}0 & 0\\ 0 & 2\end{pmatrix}$	C_1+C_2 $\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix}$	iC_1-iC_2 $\begin{pmatrix}0 & 0\\ 0 & 2\end{pmatrix}$
1	$-\mathbb{I}+X$ $\begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}$	$-\mathbb{I}+Y$ $\begin{pmatrix}-1 & -i\\ i & -1\end{pmatrix}$	$-\mathbb{I}+Z$ $\begin{pmatrix}0 & 0\\ 0 & -2\end{pmatrix}$	$-\mathbb{I}-X$ $\begin{pmatrix}-1 & -1\\ -1 & -1\end{pmatrix}$
1	$-\mathbb{I}-Y$ $\begin{pmatrix}-1 & i\\ -i & -1\end{pmatrix}$	$-\mathbb{I}-Z$ $\begin{pmatrix}-2 & 0\\ 0 & 0\end{pmatrix}$	$-C_1-C_2$ $\begin{pmatrix}-2 & 0\\ 0 & 0\end{pmatrix}$	$-iC_1+iC_2$ $\begin{pmatrix}0 & 0\\ 0 & -2\end{pmatrix}$
−1	$i\mathbb{I}+iX$ $\begin{pmatrix}i & i\\ i & i\end{pmatrix}$	$i\mathbb{I}+iY$ $\begin{pmatrix}i & 1\\ -1 & i\end{pmatrix}$	$i\mathbb{I}+iZ$ $\begin{pmatrix}2i & 0\\ 0 & 0\end{pmatrix}$	$i\mathbb{I}-iX$ $\begin{pmatrix}i & -i\\ -i & i\end{pmatrix}$
−1	$i\mathbb{I}-iY$ $\begin{pmatrix}i & -1\\ 1 & i\end{pmatrix}$	$i\mathbb{I}-iZ$ $\begin{pmatrix}0 & 0\\ 0 & 2i\end{pmatrix}$	$-C_1+C_2$ $\begin{pmatrix}0 & 0\\ 0 & 2i\end{pmatrix}$	iC_1+iC_2 $\begin{pmatrix}2i & 0\\ 0 & 0\end{pmatrix}$
−1	$-i\mathbb{I}+iX$ $\begin{pmatrix}-i & i\\ i & -i\end{pmatrix}$	$-i\mathbb{I}+iY$ $\begin{pmatrix}-i & 1\\ -1 & -i\end{pmatrix}$	$-i\mathbb{I}+iZ$ $\begin{pmatrix}0 & 0\\ 0 & -2i\end{pmatrix}$	$-i\mathbb{I}-iX$ $\begin{pmatrix}-i & -i\\ -i & -i\end{pmatrix}$
−1	$-i\mathbb{I}-iY$ $\begin{pmatrix}-i & -1\\ 1 & -i\end{pmatrix}$	$-i\mathbb{I}-iZ$ $\begin{pmatrix}-2i & 0\\ 0 & 0\end{pmatrix}$	C_1-C_2 $\begin{pmatrix}0 & 0\\ 0 & -2i\end{pmatrix}$	$-iC_1-iC_2$ $\begin{pmatrix}-2i & 0\\ 0 & 0\end{pmatrix}$

Figure A1. For $(\det(U_c)^2/N$, U-matrices with context label +1 are Hermitian; with −1, anti-Hermitian.

i	$S_2 + C_2$ $\begin{pmatrix} 1 & 1 \\ i & i \end{pmatrix}$	$-S_2 + C_2$ $\begin{pmatrix} 1 & -1 \\ -i & i \end{pmatrix}$	$S_2 - C_2$ $\begin{pmatrix} -1 & 1 \\ i & -i \end{pmatrix}$	$-S_2 - C_2$ $\begin{pmatrix} -1 & -1 \\ -i & -i \end{pmatrix}$
i	$iS_1 + C_2$ $\begin{pmatrix} 1 & i \\ 1 & i \end{pmatrix}$	$-iS_1 - C_2$ $\begin{pmatrix} -1 & -i \\ -1 & -i \end{pmatrix}$	$iS_1 - C_2$ $\begin{pmatrix} -1 & i \\ 1 & -i \end{pmatrix}$	$-iS_1 + C_2$ $\begin{pmatrix} 1 & -i \\ -1 & i \end{pmatrix}$
i	$iS_1 + iC_1$ $\begin{pmatrix} i & i \\ 1 & 1 \end{pmatrix}$	$-iS_1 - iC_1$ $\begin{pmatrix} -i & -i \\ -1 & -1 \end{pmatrix}$	$-iS_1 + iC_1$ $\begin{pmatrix} i & -i \\ -1 & 1 \end{pmatrix}$	$iS_1 - iC_1$ $\begin{pmatrix} -i & i \\ 1 & -1 \end{pmatrix}$
i	$S_2 + iC_1$ $\begin{pmatrix} i & 1 \\ i & 1 \end{pmatrix}$	$-S_2 - iC_2$ $\begin{pmatrix} -i & -1 \\ -i & 1 \end{pmatrix}$	$S_2 - iC_1$ $\begin{pmatrix} -i & 1 \\ i & -1 \end{pmatrix}$	$-S_2 - iC_1$ $\begin{pmatrix} -i & -1 \\ -i & -1 \end{pmatrix}$
$-i$	$S_1 + C_1$ $\begin{pmatrix} 1 & 1 \\ -i & -i \end{pmatrix}$	$-S_1 - C_1$ $\begin{pmatrix} -1 & -1 \\ i & i \end{pmatrix}$	$S_1 - C_1$ $\begin{pmatrix} -1 & 1 \\ -i & i \end{pmatrix}$	$-S_1 + C_1$ $\begin{pmatrix} 1 & -1 \\ i & -i \end{pmatrix}$
$-i$	$iS_2 + C_1$ $\begin{pmatrix} 1 & i \\ -1 & -i \end{pmatrix}$	$iS_2 - C_1$ $\begin{pmatrix} -1 & i \\ -1 & i \end{pmatrix}$	$-iS_2 + C_1$ $\begin{pmatrix} 1 & -i \\ 1 & -i \end{pmatrix}$	$-iS_2 - C_1$ $\begin{pmatrix} -1 & -i \\ 1 & i \end{pmatrix}$
$-i$	$iS_2 + iC_2$ $\begin{pmatrix} i & i \\ -1 & -1 \end{pmatrix}$	$-iS_2 - iC_2$ $\begin{pmatrix} -i & -i \\ 1 & 1 \end{pmatrix}$	$-iS_2 + iC_2$ $\begin{pmatrix} i & -i \\ 1 & -1 \end{pmatrix}$	$iS_2 - iC_2$ $\begin{pmatrix} -i & i \\ -1 & 1 \end{pmatrix}$
$-i$	$S_1 + iC_2$ $\begin{pmatrix} i & 1 \\ -i & -1 \end{pmatrix}$	$-S_1 - iC_2$ $\begin{pmatrix} -i & -1 \\ i & 1 \end{pmatrix}$	$S_1 - iC_2$ $\begin{pmatrix} -i & 1 \\ -i & 1 \end{pmatrix}$	$-S_1 + iC_2$ $\begin{pmatrix} i & -1 \\ i & -1 \end{pmatrix}$

Figure A2. Set 2 of 32 U-matrices with context label ±i (where $(\det(U_c)^2/N)$.

i	$\mathbb{I} - C_1$ $\begin{pmatrix} 0 & 0 \\ 0 & 1+i \end{pmatrix}$	$-\mathbb{I} + C_1$ $\begin{pmatrix} 0 & 0 \\ 0 & -1-i \end{pmatrix}$	$\mathbb{I} + iC_2$ $\begin{pmatrix} 1+i & 0 \\ 0 & 0 \end{pmatrix}$	$-\mathbb{I} + iC_2$ $\begin{pmatrix} -1-i & 0 \\ 0 & 0 \end{pmatrix}$
i	$i\mathbb{I} + C_1$ $\begin{pmatrix} 1+i & 0 \\ 0 & 0 \end{pmatrix}$	$-i\mathbb{I} - C_1$ $\begin{pmatrix} -1-i & 0 \\ 0 & 0 \end{pmatrix}$	$i\mathbb{I} - iC_2$ $\begin{pmatrix} 0 & 0 \\ 0 & 1+i \end{pmatrix}$	$-i\mathbb{I} + iC_2$ $\begin{pmatrix} 0 & 0 \\ 0 & -1-i \end{pmatrix}$
i	$Z - C_2$ $\begin{pmatrix} 0 & 0 \\ 0 & -1-i \end{pmatrix}$	$-Z + C_2$ $\begin{pmatrix} 0 & 0 \\ 0 & 1+i \end{pmatrix}$	$-Z - iC_1$ $\begin{pmatrix} -1-i & 0 \\ 0 & 0 \end{pmatrix}$	$Z + iC_1$ $\begin{pmatrix} 1+i & 0 \\ 0 & 0 \end{pmatrix}$
i	$iZ + C_2$ $\begin{pmatrix} 1+i & 0 \\ 0 & 0 \end{pmatrix}$	$-iZ - C_2$ $\begin{pmatrix} -1-i & 0 \\ 0 & 0 \end{pmatrix}$	$iZ - iC_1$ $\begin{pmatrix} 0 & 0 \\ 0 & -1-i \end{pmatrix}$	$-iZ + iC_1$ $\begin{pmatrix} 0 & 0 \\ 0 & 1+i \end{pmatrix}$
$-i$	$\mathbb{I} - C_2$ $\begin{pmatrix} 0 & 0 \\ 0 & 1-i \end{pmatrix}$	$-\mathbb{I} + C_2$ $\begin{pmatrix} 0 & 0 \\ 0 & -1+i \end{pmatrix}$	$\mathbb{I} - iC_1$ $\begin{pmatrix} 1-i & 0 \\ 0 & 0 \end{pmatrix}$	$-\mathbb{I} + iC_1$ $\begin{pmatrix} -1+i & 0 \\ 0 & 0 \end{pmatrix}$
$-i$	$i\mathbb{I} - C_2$ $\begin{pmatrix} -1+i & 0 \\ 0 & 0 \end{pmatrix}$	$-i\mathbb{I} + C_2$ $\begin{pmatrix} 1-i & 0 \\ 0 & 0 \end{pmatrix}$	$i\mathbb{I} - iC_1$ $\begin{pmatrix} 0 & 0 \\ 0 & -1+i \end{pmatrix}$	$-i\mathbb{I} + iC_1$ $\begin{pmatrix} 0 & 0 \\ 0 & 1-i \end{pmatrix}$
$-i$	$Z - C_1$ $\begin{pmatrix} 0 & 0 \\ 0 & -1+i \end{pmatrix}$	$-Z + C_1$ $\begin{pmatrix} 0 & 0 \\ 0 & 1-i \end{pmatrix}$	$-Z + iC_2$ $\begin{pmatrix} -1+i & 0 \\ 0 & 0 \end{pmatrix}$	$Z - iC_2$ $\begin{pmatrix} 1-i & 0 \\ 0 & 0 \end{pmatrix}$
$-i$	$iZ - C_1$ $\begin{pmatrix} -1+i & 0 \\ 0 & 0 \end{pmatrix}$	$-iZ + C_1$ $\begin{pmatrix} 1-i & 0 \\ 0 & 0 \end{pmatrix}$	$iZ - iC_2$ $\begin{pmatrix} 0 & 0 \\ 0 & 1-i \end{pmatrix}$	$-iZ + iC_2$ $\begin{pmatrix} 0 & 0 \\ 0 & -1+i \end{pmatrix}$

Figure A3. U-matrices formed by sums of non-orthogonal contexts. All are unitary and diagonal, with context label ±*i* (for $(\det(U_c)^2/N)$ but fail the orthogonality condition for chain formation.

References

1. Pāṇini. Aṣṭādhyāyī. In *Aṣṭādhyāyī of Pāṇini: Roman Transliteration and English Translation*; Katre, S.M., Trans.; University of Texas Press: Austin, TX, USA, 1987.
2. Bloomfield, L. *Language*; Holt, Rinehart and Winston: New York, NY, USA, 1933.

3. Covington, M. *Syntactic Theory in the High Middle Ages: Modistic Models of Sentence Structure*; Cambridge Studies in Linguistics; Cambridge University Press: Cambridge, UK, 1984.
4. Arnauld, A.; Lancelot, C. *Grammaire Générale et Raisonnée*; Paris, France, 1660. For English translation see *General and Rational Grammar: The Port-Royal Grammar*. Rieux, J., Rollin, B., Eds. and Trans.; Mouton: The Hague, The Netherlands, 1975.
5. Chomsky, N. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*; Harper & Row: New York, NY, USA, 1966.
6. Chomsky, N. *Aspects of the Theory of Syntax*; MIT Press: Cambridge, MA, USA, 1965.
7. Chomsky, N. *The Minimalist Program*; MIT Press: Cambridge, MA, USA, 1995.
8. Chomsky, N. *Knowledge of Language: Its Nature, Origin, and Use*; Praeger: New York, NY, USA, 1986.
9. Levelt, W. *Speaking: From Intention to Articulation*; MIT Press: Cambridge, MA, USA, 1989.
10. Gallistel, R. *The Organization of Learning*; MIT Press: Cambridge, MA, USA, 1990.
11. de Saussure, F. *Cours de Linguistique Générale*; Bally, C., Sechehaye, A., Eds.; Payot: Paris, France, 1916. For English translation see *Course in General Linguistics*. Baskin, W., Trans.; McGraw-Hill: New York, NY, USA, 1959.
12. Peirce, C.S. Prolegomena to an Apology for Pragmaticism. In *Collected Papers of Charles Sanders Peirce*; The Simplest Mathematics; Hartshorne, C., Weiss, P., Eds.; Harvard University Press: Cambridge, MA, USA, 1933; Volume 4; §4, pp. 423–424 (CP 4.537–4.538).
13. Fitz, H.; Hagoort, P.; Petersson, K. Neurobiological Causal Models of Language Processing. *Neurobiol. Lang.* **2024**, *5*, 225–247.
14. Collins, C.; Stabler, E. A Formalization of Minimalist Syntax. *Syntax* **2016**, *19*, 43–78.
15. Chomsky, N.; Seely, D.; Berwick, R.; Fong, S.; Huybregts, R.; Kitahara, H.; McInnerney, A.; Sugimoto, Y. Merge and the Strong Minimalist Thesis. In *Elements in Generative Syntax*; Cambridge UP: Cambridge, UK, 2023.
16. Marcolli, M.; Chomsky, N.; Berwick, R. *Mathematical Structure of Syntactic Merge: An Algebraic Model for Generative Linguistics*; MIT Press: Cambridge, MA, USA, 2025.
17. Fodor, J.; Pylyshyn, Z. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition* **1988**, *28*, 3–71.
18. Fodor, J. *The Language of Thought*; Harvard University Press: Cambridge, MA, USA, 1975.
19. Berwick, R.; Chomsky, N. *Why Only Us: Language and Evolution*; MIT Press: Cambridge, MA, USA, 2016.
20. Chomsky, N. *The Amherst Lectures*; Unpublished lecture notes from the LSA Summer Institute, University of Massachusetts, Amherst, MA; distributed by Documents Linguistiques, University of Paris VII: Paris, France, 1974.
21. Jakobson, R.; Fant, G.; Halle, M. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*; MIT Acoustics Laboratory, Technical Report No. 13: Cambridge, MA, USA, 1952.
22. McCarthy, J. Feature Geometry and Dependency: A Review. *Phonetica* **1988**, *45*, 84–108.
23. Clements, N.; Hume, E. The Internal Organization of Speech Sounds. In *Handbook of Phonological Theory*; Goldsmith, J., Ed.; Basil Blackwell: Oxford, UK, 1995; pp. 245–306.
24. Firth, R. *Papers in Linguistics 1934–1951*; Oxford UP: Oxford, UK, 1957.
25. Wittgenstein, L. *Philosophical Investigations*; Anscombe, G., Ed.; Blackwell: Oxford, UK, 1953.
26. Harris, Z. Distributional Structure. *Word* **1954**, *10*, 146–162.
27. Landauer, T.; Dumais, S. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychol. Rev.* **1997**, *104*, 211–240.
28. Bruza, P.; Busemeyer, J.; Gabora, L. Introduction to the special issue on quantum models of cognition and decision. *J. Math. Psychol.* **2009**, *53*, 345–361.
29. Pulvermüller, F. Words in the brain's language. *Behav. Brain Sci.* **2005**, *28*, 179–224.
30. Binder, J.; Desai, R.; Graves, W.; Conant, L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* **2009**, *19*, 2767–2796.
31. Burke, D.; MacKay, D. Memory, language, and ageing. *Philos. Trans. R. Soc. B Biol. Sci.* **1997**, *352*, 1845–1856.
32. Bybee, J. *Language, Usage and Cognition*; Cambridge UP: Cambridge, UK, 2010.
33. Kuhl, P. Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci.* **2004**, *5*, 831–843.
34. Gottesman, D. The Heisenberg Representation of Quantum Computers. In *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*; Corney, S.P., Delbourgo, R., Jarvis, P.D., Eds.; International Press: Cambridge, MA, USA, 1999; pp. 32–43.
35. Hall, B. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, 2nd ed.; Springer: New York, NY, USA, 2015.
36. Lee, J. *Introduction to Smooth Manifolds*, 2nd ed.; Springer: New York, NY, USA, 2013.
37. Nielsen, M.; Chuang, I. *Quantum Computation and Quantum Information*, 10th ed.; Cambridge University Press: Cambridge, UK, 2010.

38. Sakurai, J.; Napolitano, J. *Modern Quantum Mechanics*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2017.
39. Ajdukiewicz, K. Die syntaktische Konnexität. *Stud. Philos.* **1935**, *1*, 1–27. For English translation see *Polish Logic 1920–1939*. McCall, S., Ed.; Oxford University Press: Oxford, UK, 1967.
40. Bar-Hillel, Y. A Quasi-Arithmetical Notation for Syntactic Description. *Language* **1953**, *29*, 47–58.
41. Smolensky, P. Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems. *Artif. Intell.* **1990**, *46*, 159–216.
42. Orús, R.; Gallego, A. MERGE as Information Renormalization: Tensor Networks and Coarse-Graining in Syntax. *SN Comput. Sci.* **2022**, *3*, 1–27.
43. Cartier, P. A Primer of Hopf Algebras. In *Frontiers in Number Theory, Physics, and Geometry II*; Cartier, P., Moussa, P., Julia, B., Vanhove, P., Eds.; Springer: Berlin, Germany, 2007; pp. 537–615.
44. Busemeyer, J.; Bruza, P. *Quantum Models of Cognition and Decision*; Cambridge UP: Cambridge, UK, 2012.
45. Coecke, B.; Sadrzadeh, M.; Clark, S. Mathematical Foundations for a Compositional Distributional Model of Meaning. *Linguist. Anal.* **2010**, *36*, 345–384.
46. Lambek, J. Type Grammar Revisited. In *Logical Aspects of Computational Linguistics*; Lecture Notes in Computer Science; Lecomte, A., Lamarche, F., Perrier, G., Eds.; Springer: Berlin, Germany, 1999; Volume 1582, pp. 1–27.
47. Smolensky, P.; Legendre, G. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*; MIT Press: Cambridge, MA, USA, 2006.
48. Coecke, B.; Kissinger, A. *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*; Cambridge UP: Cambridge, UK, 2017.
49. Piattelli-Palmarini, M. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*; Harvard University Press: Cambridge, MA, USA, 1980.
50. Peskin, M.; Schroeder, D. *An Introduction to Quantum Field Theory*; Westview Press: Boulder, CO, USA, 1995.
51. Griffiths, D.J. *Introduction to Quantum Mechanics*, 2nd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2005; Chapters 1–2.
52. Horn, R.; Johnson, C. *Matrix Analysis*, 2nd ed.; Cambridge UP: Cambridge, UK, 2013.
53. Potsdam, E.; Polinsky, M. Backward Raising. *Syntax* **2012**, *15*, 75–108.
54. Lasnik, H.; Uriagereka, J. *Structure: Concepts, Consequences, Interactions*; MIT Press: Cambridge, MA, USA, 2022.
55. Lindenmayer, A. Mathematical models for cellular interactions in development. *J. Theor. Biol.* **1968**, *18*, 280–315.
56. Rozenberg, G.; Salomaa, A. *The Mathematical Theory of L Systems*; Academic Press: New York, NY, USA, 1980.
57. Allouche, J.-P.; Shallit, J. *Automatic Sequences: Theory, Applications, Generalizations*; Cambridge University Press: Cambridge, MA, USA, 2003.
58. Seneta, E. *Non-Negative Matrices and Markov Chains (Revised Printing)*; Springer: New York, NY, USA, 2006.
59. Baez, J. The Octonions. *Bull. Am. Math. Soc.* **2002**, *39*, 145–205.
60. Chomsky, N. *The Logical Structure of Linguistic Theory*; Plenum Press: New York, NY, USA, 1975; Originally circulated as a mimeographed manuscript, MIT, 1955.
61. Collins, C.; Groat, E. Distinguishing Copies and Repetitions. 2018. Available online: https://ling.auf.net/lingbuzz/003809 (accessed last on September 1, 2026).
62. Carnie, A. *Syntax: A Generative Introduction*, 3rd ed.; Blackwell: Malden, MA, USA, 2013.
63. Bowers, J. The Syntax of Predication. *Linguist. Inq.* **1993**, *24*, 591–656.
64. Chomsky, N. Derivation by Phase. In *Ken Hale: A Life in Language*; Current Studies in Linguistics; Kenstowicz, M., Ed.; MIT Press: Cambridge, MA, USA, 2001; Volume 36, pp. 1–52.
65. Lind, D.; Marcus, B. *An Introduction to Symbolic Dynamics and Coding*; Cambridge University Press: Cambridge, UK, 1995.
66. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
67. Jurafsky, D.; Martin, J. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2009; Chapters 12–15.
68. Baddeley, A. Working Memory: Theories, Models, and Controversies. *Annu. Rev. Psychol.* **2012**, *63*, 1–29.
69. Dehaene, S. *The Number Sense: How the Mind Creates Mathematics*; Oxford UP: Oxford, UK, 1997.
70. Hardy, G.; Wright, E. *An Introduction to the Theory of Numbers*, 6th ed.; Oxford UP: Oxford, UK, 2008.
71. Berwick, R.; Weinberg, A. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*; MIT Press: Cambridge, MA, USA, 1984.

72. Miller, G.; Chomsky, N. Finitary models of language users. In *Handbook of Mathematical Psychology*; Luce, R.D., Bush, R.R., Galanter, E., Eds.; Wiley: New York, NY, USA, 1963.
73. Malchukov, A.; Haspelmath, M.; Comrie, B. *Studies in Ditransitive Constructions: A Comparative Handbook*; De Gruyter Mouton: Berlin, Germany, 2010.
74. Fortescue, M. *West Greenlandic*; Routledge: London, UK, 1994.
75. Miller, G. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **1956**, *63*, 81–97.
76. Ueno, Y. Formal Properties of the Prerequisite Operation for Combining Syntactic Objects: Categorical Foundations of an Algebraic Theory of Workspace. Master's Thesis, Sophia University, Tokyo, Japan, 2026.
77. Pinker, S.; Jackendoff, R. The Faculty of Language: What's Special about it? *Cognition* **2005**, *95*, 201–236.
78. Gallistel, R.; King, A. *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*; Wiley–Blackwell: Malden, MA, USA, 2009.
79. Orús, R.; Martin, R.; Uriagereka, J. Mathematical foundations of matrix syntax. *arXiv* **2017**, arXiv:1710.00372.
80. Morrison, D. A Stolarsky Array of Wythoff Pairs. In *A Collection of Manuscripts Related to the Fibonacci Sequence*; Hoggatt, V., Bicknell-Johnson, M., Eds.; The Fibonacci Association: Santa Clara, CA, USA, 1980; pp. 134–136.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.