

Article

Categories with Complements

Juan Uriagereka

Department of Linguistics and SLLC, University of Maryland, College Park, MD 20742-7505, USA; juan@umd.edu

Abstract: Verbs and nouns gear θ -dependencies, Case, agreement, or construal relations. Building on Chomsky's 1974 decomposition of such categories into $\pm N$, $\pm V$ features, by translating said features into ± 1 , $\pm i$ scalars that allow for the construction of a vector space, this paper studies the possibility of organizing said features into 2×2 square matrices. In the system proposed to explore "head-complement" relations, operating on nouns yields a measurable/observable (Hermitian matrix), which in turn limits other potential combinations with abstract lexical categories. Functional/grammatical categories in the system deploy the same features, albeit organized differently in the matrix diagonal and off-diagonal. The algebraic result is a group with well-defined mathematical properties, which properly includes the Pauli group of standard use in quantum computation. In the system, the presumed difference between categories and interactions—here, in a context of the head-complement sort—reduces to whether the magnitude of the matrix eigenvalue is 1 or not, in the latter instance inducing asymmetric interactions.

Keywords: lexical categories; functional categories; (external) merge; matrix syntax; Pauli group

1. Featural Specifications as Categorical Dimensions

Let us examine the head-complement relation, which presupposes an understanding of categories that can be heads and what we presume to be their complements.¹ In current parlance, we take those to be instances of External Merge (EM), of the 1st-Merge type: this is inasmuch as the process is presumed to take an item from the lexicon (the head) and asymmetrically associate it to a projection, its complement. Evidently, that presupposes understanding what is meant by "head" and "projection", or the presumed "endocentricity" of phrases. Bear in mind that a Chomsky grammar is a 4-tuple $[\Sigma, T, N, P]$, specifying rules over arbitrary vocabularies of "non-terminals" N and "terminals" T , thus allowing a set of productions P starting on initial axiom Σ . Rules of the sort $X \rightarrow \psi$, where $X \in N$ and ψ is some arbitrary string of terminals in T^* ,² do not guarantee the endocentricity of phrases, since (by that formalism) one could rewrite a VP, say, as a noun followed by a prepositional phrase only (or any other string of terminals, but no verb). If these constructions do not exist in natural language, the formalism is inadequate [1].

The problem of endocentricity is intimately related to the problem of selection (e.g., that intransitive, transitive, or ditransitive verb takes different dependents). It is *because* a transitive verb takes a complement *that* it is immediately dominated by a VP of a (transitive) kind, and similarly with other such dependencies. That is the essence of the "lexicalist" idea behind bottom-up projections, sketched in Chomsky [2] and then pursued from Jackendoff [3] to Speas [4], the bases (together with Muysken [5] and Kayne [6]) for Chomsky's Bare Phrase Structure (BPS) [7]. To capture such syntagmatic interactions, linguists had to consider systematicities implied not just in the terminal vocabulary (nouns as opposed to verbs, marked vs. default exemplars, etc.), but also the generalizations within the non-terminals (subcategorization restrictions among the former, label endocentricity). The standard treatment of these ideas presumed "distinctive features" of some sort [8,9].

Chomsky [2] extended distinctive phonological features to the noun vs. verb distinction. Just as consonants are taken to be maximally distinct from vowels (which the orthogonality of features [+consonantal, –syllabic] vs. [–consonantal, +syllabic] expresses),



Citation: Uriagereka, J. Categories with Complements. *Philosophies* 2022, 7, 102. <https://doi.org/10.3390/philosophies7050102>

Academic Editor: Peter Kosta

Received: 26 May 2022

Accepted: 29 August 2022

Published: 15 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. 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 (<https://creativecommons.org/licenses/by/4.0/>).

so too verbs are seen, in demonstrable senses, as maximally distinct from nouns, as reflected in Chomsky's equally orthogonal categorial features $[-N, +V]$ vs. $[+N, -V]$. In fact, Chomsky seems to have been concerned with orthogonality between syntactic "dimensions" of some sort; just as the "vowel dimension" is as different to the "consonant dimension" as can be, so too can nominal expressions in syntax be seen as orthogonal to verbal expressions. This has (arguable) consequences that Varro already emphasized in *De Lingua Latina*, Vol 2: Book IX, XXIV-31, when dividing "speech into four parts, one in which the words have cases, a second in which they have indications of time, a third in which they have neither, a fourth in which they have both" (translated by Kent [10]). It is actually remarkable how similar this is to Chomsky's categorization for the X' -schema:

As far as the categorial component is concerned, it seems to me plausible to suggest that it is a kind of projection from basic lexical features through a certain system of schemata as roughly indicated in (1) and (2):

- (1) $[\pm N, \pm V]: [+N, -V] = N[\text{oun}]; [+N, +V] = A[\text{djective}]; [-N, +V] = V[\text{erb}],$
 $[-N, -V] = \text{everything else};$
- (2) $X^n \rightarrow \dots X^{n-1} \dots$, where $x^i = [\alpha = \pm N, \beta = \pm V]^i$ and $X^1 = X$

Let us assume that there are two basic lexical features N and V ($\pm N, \pm V$). Where the language has rules that refer to the categories nouns and adjectives ... they will be framed in terms of the feature $+N$. and where there are rules that apply to the category nouns and adjectives, they will be framed in terms of the feature $+V$. [2] (Lecture 3, p. 2)

What is the nature of those "dimensions"? One possibility is to liken them to the underlying "vision maps" that stem from cones and rods (see [11]). In this view, whatever the nature of features may ultimately be, it is mediated by brain physiology. This approach is promising, for instance, for voice onset time (VOT), associated by Poeppel [12] to specific "brain events". That seems just as reasonable for activating vocal folds, upon the appropriate motor commands, as the response of cones of different types is for light of diverse frequencies—if presuming efference vs. afference. For that approach, the project is empirical in a classical sense, and perhaps harder for linguistic features (if not present in other animals we allow ourselves to experiment with) than for vision features one detects even in insects. Much as I respect that line of reasoning, I have little to contribute to it in the present context. But there is a different, though compatible, approach to feature dimensions: that their nature is algebraic and, hence, arguably *prior* to any substantive feature system. In phonology, the \pm consonantal or \pm syllabic dimensions seem more fundamental than, for example, \pm continuant or \pm labial features, in terms (at least) of universal relevance in articulating morpho-phonemics and the orthogonality conditions presumed for the deeper type. The issue I want to explore here is whether similar fundamental dimensions exist in syntax.

Note that Chomsky could have chosen attributes A or P (for adjectives and prepositions) instead of N or V , since any two binary features obviously yield four distinctions. But he assumed, instead, "that there are two lexical features N and V ", which accords with traditional intuitions: in some sense, nominal and verbal dimensions are more essential than adjectival or prepositional ones.³ Chomsky [2] (lecture 3, p. 3) also asserts that the initial phrase-marker, prior to any transformations modifying it, is "projected from the lexical categories uniformly". He then clarifies his hypothesis that: "in a fundamental way the expansion of major categories like NP , VP , AP is independent of categorial choice of the head ... [as] instantiations of the same general schemata." This presupposes an understanding that syntax is somehow articulated around N and V dimensions. The suggestion I will be pursuing here is that the N vs. V distinction reflects algebraic conditions that also underlie our number system. This is for reasons Chomsky often emphasizes, going back to Alfred Wallace (see [13]): that mathematical knowledge could not have served any obvious evolutionary purpose—and, thus, is seen as an outgrowth of language. But before developing that point, I would like to reflect on the topological nature of phrases.

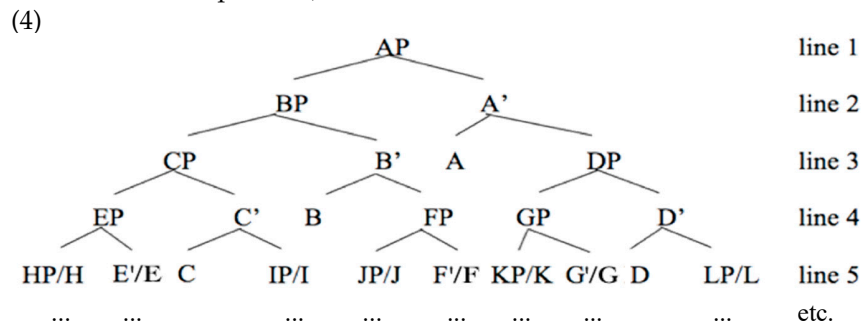
2. Lindenmayer Systems in Matrix Representation

A standard Chomsky-grammar involves rewriting mechanisms applying one-at-a-time (Chomsky’s [14] “Traffic Convention”). While rewrite rules were abandoned half a century ago (see [15,16]), actually projecting systems still preserve the fundamental properties of a Chomsky-grammar. This is demonstrably the case for string-based systems like those in [17], but it can also be argued for the BPS system in [7] or current extensions. A relation like (3a) is equivalent, at the relevant level of abstraction, to one as in (3b):

- (3) a. The merge of α and β results in $\{\alpha, \beta\}$
- b. $\{\alpha, \beta\} \rightarrow \alpha, \beta$

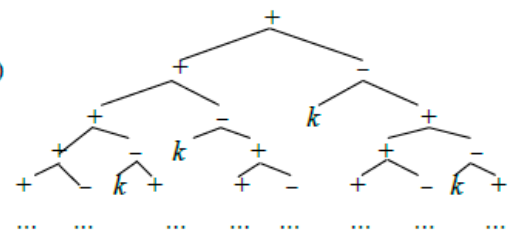
Needless to say, it makes good sense to *project* (whatever that ultimately means) α or β as in (3a), the relation being “bottom-up”, whereas the same is not true about the “top-down” (3b). As will be clear immediately, however, this is immaterial to the topology of the resulting phrasal objects. To see that, we need to consider a phrasal entity that boils down to the *maximal expansion* of procedures as in (3), literally to infinity because they are (in principle) recursive on all branches.

Lindenmayer [18] explored the elimination of Chomsky’s Traffic Convention, so that in Lindenmayer (L-)systems each rewritable symbol must rewrite, at whatever derivational line the system is running. This is the reason L-systems, while good at modeling some fractal structures, do not generate formal languages, even though they describe the overall topological space of a certain natural language expansion. For instance, Boeckx, Carnie, & Medeiros [19] show how the customary syntactic X’-theory, presumed by Speas [4], yields maximally expanding structures as in (26) (where the “slash” notation is meant to represent an unbounded expansion):



Sending this schema to infinity (imagining all relevant projections that become possible as the structure unfolds), produces an L-tree following the rule-schema in (27) (where k is a constant):

- (5) a. $+ \rightarrow +- \text{ (equivalently } +, - \text{ merge to form } \{+, -\} \text{)}$
- b. $- \rightarrow k+ \text{ (equivalently } k, + \text{ merge to form } \{k, +\} \text{)}$



More importantly for our purposes, in this Lindenmayer (maximally expanded) fashion, there is a useful matrix representation for the ensuing phrasal object and its corresponding topology.

Medeiros [20] expands on an idea dating back to King [21], for an L-system like (5), over n specific symbols (here three: $+$, $-$, and k). The idea is to associate these symbols to entries in square matrix columns 1 to n , next to represent each rule expansion as corresponding square matrix rows, utilizing a zero for whichever symbol a rule expansion lacks. Uriagereka [22] represents constants as their own column, albeit with a corresponding row of only zeroes. For (5):

(6)		TYPES
a. $+ \rightarrow +- (equivalently +, - \text{ merge to form } \{+, -\})$	\swarrow \rightarrow	$\begin{matrix} + & - & k \\ + & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix}$
b. $- \rightarrow k+ (equivalently, k, + \text{ merge to form } \{k, +\})$	\rightarrow \swarrow	$\begin{matrix} + & - & k \\ - & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix}$
c. No expansion for the constant/terminal	\rightarrow	$\begin{matrix} + & - & k \\ k & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix}$

All that (6) encodes is that process (5a) (merging + and −, projecting +) is presented as one token + and one token − in the “+ row” (with zero ks in that row); and process (5b) (merging k and +, projecting −) is coded as 1 token + and one token k in the “− row” (with zero − s in that row); while the k row is all zeroes because k doesn’t expand, as it is “its own projection”. As such, this is just a compact way to pack the relevant information into a square matrix. One such matrix is a particularly elegant mathematical object, with an associated characteristic polynomial calculated from scalars associated to the matrix diagonal (it’s so-called *trace* and *determinant*).

While readers can find details about matrices in any standard introduction to linear algebra (even in the Wikipedia entry for “square matrix”), it is worth bearing in mind that the characteristic polynomial for a square matrix is a fundamental invariant across different bases (which keep relations unchanged) for the matrix. Recall that a polynomial is a collection of monomial terms K^x , where constant K is the term’s coefficient and a root of the polynomial turns it to an equation; the root of polynomial $P(z)$ is the number z_j such that $P(z_j) = 0$. $P(z)$ is of degree n if it has n roots, its degrees of freedom. Thus, the characteristic polynomial of a matrix can be thought of as its numerical description. The polynomial roots (solutions to the polynomial when construed as equating zero) constitute fundamental elements in the matrix diagonal, its eigenvalues. From those, one can compute the matrix *eigenvectors*, or characteristic vectors of the linear transformation the matrix represents, which changes by a scalar factor when the linear transformation is applied—the factor in point being the eigenvalue. Aside from having the eigenvalues as roots, among the polynomial coefficients are the matrix *trace* (sum of the elements in the diagonal) and determinant (here, a product of the elements in the diagonal minus the product of those in the off-diagonal, with further caveats for higher-dimensional matrices). It is interesting to note for our purposes that, with Medeiros’s method in place, we can compute how an L-system grows as it expands. The degree n of the characteristic polynomial tells us the dimension $n \times n$ of its corresponding square matrix and, therefore, the number of rewrite or Merge type applications that correspond to it in an L-system. In turn, one matrix eigenvalue, often called its *spectral radius*, can be defined as in (7a), a quantity related to the topological derivational entropy h_T of a matrix A, (7b):

- (7) *Derivational entropy*
 - a. For *spectral radius* $\rho(A)$, square matrix A’s largest absolute value λ_{max} of A’s eigenvalues,
 - b. A’s topological derivational entropy $h_T = \log_2 \lambda_{max}$

So relevant to this calculation of “how the system grows” is the magnitude of the highest eigenvalue of its associated matrix. The systems that interest us are “fractal” in that the recursive application of their constituent rules, when maximally applied, correspond to square matrices whose repetitive effect on the system is tracked by self-multiplication (a matrix power), as shown for related systems in Ott [23]. We can see this idea at work by simply running the successive powers of an L-system interpreted through its Medeiros-matrix (henceforth M-matrix) as in (6):

$$(8) \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^1 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}; \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}; \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}; \dots$$

With each successive power we obtain, in the matrix top row, the number of symbols of a given kind in a new derivational line in the L-tree, as readers can check by simply counting them in (5) (one can also try as an exercise the next power, to see how the next top row will

be 5, 3, 2, etc.).⁴ Slightly more technically, for any two successive matrix powers A^n and A^{n+1} , the ratio between the spectral radii $\rho(A^{n+1})/\rho(A^n)$ is the initial spectral radius $\rho(A^1)$ for the seed matrix.⁵

Note: the claim is not that an object like (4), its abstract version in (5), or its square matrix representation in (6), are anything like “syntactic trees” that users parse in the performance of their grammatical knowledge. It is, instead, that the underlying *topological representation* of the (X' -theoretic) phrasal scheme has the “shape” implied in (6), whose derivational entropy, in the sense in (7), can be simply calculated—this quantity, again, telling us “how the system grows”. Here is not the place to demonstrate this, but actually the X' -theoretic representation in (4) is one among infinitely many that branch, albeit in binary fashion (or presuming binary Merge) and involving “self-projecting” constants (or absolute terminals) like k there. Indeed, among such binary structures, it is provably the most elegant in terms of its potential to pack the largest amount of structure (see Uriagereka [24]). One can think of the general topology of a phrasal system as the equivalent of a terrain over which to ride a bike. Thus, Tour de France stages are characterized by a “profile”, which can be “flat”, “mountainous”, “hilly”, “potentially windy”, and so on—characteristics that obviously affect how fast and safely cyclist ride, in a peloton or otherwise, with likely breakaways or a final sprint, etc. So too, the phrasal topology is basically determining what sorts of parse-trees are possible within a given (grammatical) system, at this point merely worrying about constituency conditions (note also that the matrix representation ignores linear order). In the present paper, I will be concentrating on only head-complement relations (see Note 1).

3. The Fundamental Assumption and its Fundamental Corollary

After having completed that excursus into phrasal topologies represented as M -matrices, we can now return to the issue of Chomsky’s 1974 system as presented in (1) and (2) above. One could soundly express the substantive intuitions we discussed in the following formal fashion:

(9) *Fundamental Assumption*

The V dimension is a (mathematical) transformation over an orthogonal N dimension.

If we instantiate (5) in the complex plane, we could conclude:

(10) *Fundamental Corollary*

The N dimension has unit value 1; the V dimension, unit value i ; $[\pm N, \pm V] = [\pm 1, \pm i]$.

Before being tempted to ask “what that even means”, in substantive terms, note an immediate consequence of the Fundamental Corollary: it allows us to operate with Chomsky’s matrices. Clearly, $[\pm N, \pm V]$ makes no claim different from $[\pm V, \pm N]$, in that alternative order, whereas $[\pm 1, \pm i]$ can be entirely different from $[\pm i, \pm 1]$. That may matter much, formally, which we need to delve into, so as to seriously (attempt to) answer what (10) might imply, entail, and in the end mean.

Chomsky [2], in the same section quoted in Section 1 above, raised the following issue:

There will also be subsidiary features, that are necessary to distinguish auxiliaries from main verbs and adverbials from adjectives for example. So there will be a hierarchy of categories, super-lexical categories, super-super-lexical categories . . . In other words it ought to be in essence the case that the structure of NP’s, S’s, and AP’s should be about the same . . . and this will be true of higher order endocentric categories.

Now: in what sense are lexical projections, like noun phrases, and grammatical projections, like sentences, “about the same”? And what are “higher order endocentric categories”? One can always introduce more dimensions, for instance calling A/P dimensions $\pm A$ and $\pm P$, playing the substantive game already alluded to for color and cones, VOT, and so on. But a direct alternative is to argue that further formal conditions emerge by algebraically operating with the fundamental ones in (10)/(11) only—that is, by taking our formalism seriously.

Once again, on p. 2 (his number (38)), Chomsky orders his features in the familiar fashion of $[\pm N, \pm V]$. This, as such, may be quite different from expressing the mere presence of $\pm N$ and $\pm V$ features, if the brackets imply order—one different from that in $[\pm V, \pm N]$. The theorist needs then to decide whether that is a notational variant, a matter that arises with ± 1 and $\pm i$, as well as expressions combining these: $[\pm 1, \pm i]$ and $[\pm i, \pm 1]$ could be equivalent ways of expressing a complex number, indicating some value in the x axis (for ± 1) and the y axis (for $\pm i$); but these two could also be different vectors. Indeed, Orús et al. [25], presuming (5)/(6), argue for the relevance of representing these notions in terms of the diagonal of 2×2 square matrices as follows:

$$(11) \quad \text{a. } \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm i \end{bmatrix} \quad \text{b. } \begin{bmatrix} \pm i & 0 \\ 0 & \pm 1 \end{bmatrix}$$

It is entirely possible that these square matrices have little to do with the Medeiros square matrices discussed in Section 2 to represent L-trees. But because one should try bold hypotheses to test their validity, I will be working towards a system in which, in the end, the ontology of the objects in (11) is actually the same one as the ones for L-trees already discussed. I will not argue that point, however, until Section 7 of this paper.

Once again, the advantage of matrices is that we can treat them as linear operators with standard properties,⁶ for traces and determinants, characteristic polynomials, or eigenvalues:⁷

(12)

a. $\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$	b. $\begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}$	c. $\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$	d. $\begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}$
tr. $1 - i$	tr. $-1 + i$	tr. $1 + i$	tr. $-1 - i$
det. $-i$	det. $-i$	det. i	det. i
$\lambda^2 - (1 - i)\lambda - i$	$\lambda^2 + (1 - i)\lambda - i$	$\lambda^2 - (1 + i)\lambda + i$	$\lambda^2 + (1 + i)\lambda + i$
$\lambda_1 = -i; \lambda_2 = 1$	$\lambda_1 = -1; \lambda_2 = i$	$\lambda_1 = i; \lambda_2 = 1$	$\lambda_1 = -1; \lambda_2 = -i$

(13)

a. $\begin{bmatrix} -i & 0 \\ 0 & 1 \end{bmatrix}$	b. $\begin{bmatrix} i & 0 \\ 0 & -1 \end{bmatrix}$	c. $\begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}$	d. $\begin{bmatrix} -i & 0 \\ 0 & -1 \end{bmatrix}$
tr. $1 - i$	tr. $-1 + i$	tr. $1 + i$	tr. $-1 - i$
det. $-i$	det. $-i$	det. i	det. i
$\lambda^2 - (1 - i)\lambda - i$	$\lambda^2 + (1 - i)\lambda - i$	$\lambda^2 - (1 + i)\lambda + i$	$\lambda^2 + (1 + i)\lambda + i$
$\lambda_1 = -i; \lambda_2 = 1$	$\lambda_1 = -1; \lambda_2 = i$	$\lambda_1 = i; \lambda_2 = 1$	$\lambda_1 = -1; \lambda_2 = -i$

Just by examining the algebraic characteristics of these Chomsky objects (through the lens of (10)), we know the categories stand in specific formal relations with regards to one another. In fact—because multiplying matrices with identical real or inverse complex entries result in the identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and multiplying any of the objects in (8)/(9) among themselves, or among the eight possible results of multiplication among the combinations as in Table 1—one can prove that the Chomsky matrices align into an Abelian (commutative) group for matrix multiplication (see Note 7).

Table 1. Chomsky group.

Inverse, Adjoint	$\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix} \leftrightarrow \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}$	$\begin{bmatrix} -i & 0 \\ 0 & 1 \end{bmatrix} \leftrightarrow \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} i & 0 \\ 0 & -1 \end{bmatrix} \leftrightarrow \begin{bmatrix} -i & 0 \\ 0 & -1 \end{bmatrix}$	
Multiplication within (13)	$\begin{bmatrix} \pm 1 & 0 \\ 0 & +i \end{bmatrix} \begin{bmatrix} \pm 1 & 0 \\ 0 & +i \end{bmatrix} = \begin{bmatrix} \pm 1 & 0 \\ 0 & +1 \end{bmatrix}$		Multiplication within (14)	$\begin{bmatrix} \pm i & 0 \\ 0 & +1 \end{bmatrix} \begin{bmatrix} \pm i & 0 \\ 0 & +1 \end{bmatrix} = \begin{bmatrix} \pm 1 & 0 \\ 0 & +1 \end{bmatrix}$	
Multiplication from (13) to (14)	$\begin{bmatrix} \pm 1 & 0 \\ 0 & +i \end{bmatrix} \begin{bmatrix} \pm i & 0 \\ 0 & +1 \end{bmatrix} = \begin{bmatrix} \pm i & 0 \\ 0 & +i \end{bmatrix}$		Multiplication from (14) to (13)	$\begin{bmatrix} \pm i & 0 \\ 0 & +1 \end{bmatrix} \begin{bmatrix} \pm 1 & 0 \\ 0 & +i \end{bmatrix} = \begin{bmatrix} \pm i & 0 \\ 0 & +i \end{bmatrix}$	

That basically means the multiplication in point is a very stable scaling, always *staying within the algebraic space* that could be characterized as involving objects of the following form:

$$(14) \begin{bmatrix} \pm 1, \pm i & 0 \\ 0 & \pm 1, \pm i \end{bmatrix}$$

One way to interpret such matrices as linear operators (see Note 6) is through column vectors:

$$(15) \text{ a. } \begin{bmatrix} \pm 1 \\ 0 \end{bmatrix}, \text{ b. } \begin{bmatrix} 0 \\ \pm 1 \end{bmatrix}$$

Matrix (14) operates on whatever vector space it transforms by carrying its “horizontal” or x unit vector to (15a) and its “vertical” or y unit vector to (15b), to scale that space from those initial conditions. This may be harder to visualize for the complex entries, with these column vectors:

$$(16) \begin{bmatrix} \pm i \\ 0 \end{bmatrix}, \text{ b. } \begin{bmatrix} 0 \\ \pm i \end{bmatrix}$$

But we just have to imagine a “rotation” of the plane where the x and y unit vectors exist onto a more abstract hyper-plane, as it were “popping out of the page”, to intuit further dimensions of transformation that we allow ourselves to contemplate, when considering the effect of operating by way of the combinations that (16) implies. To be sure, not having given any “meaning” (of any kind) to these numbers, such considerations are purely abstract—but I return to that shortly.

4. The Jarret Graph

To relate those abstract ideas to grammaticality, consider self-products in the Chomsky group (Table 2):

Table 2. Self-products for the Chomsky group.

$\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \pm i & 0 \\ 0 & \pm 1 \end{bmatrix}^2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \pm i & 0 \\ 0 & \pm i \end{bmatrix}^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$
--	---	--	---

For matrices with just real entries, the result is the identity matrix, while for those with complex entries only, the result is its negative counterpart, $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$, also in the group. The result of self-multiplying matrices with mixed entries is $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, depending on whether the complex entries are. This is a well-known matrix, called $\pm\sigma_Z$, first systematically studied by Wolfgang Pauli while analyzing electron spin, around 1924 (see https://en.wikipedia.org/wiki/Pauli_matrices (accessed on 28 August 2022)). This matrix is remarkably elegant, as shown through relevant formal characteristics:

(17)

a. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	b. $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$	c. $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	d. $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
tr. 2	tr. -2	tr. 0	tr. 0
det. 1	det. 1	det. -1	det. -1
$\lambda^2 - 2\lambda + 1$	$\lambda^2 + 2\lambda + 1$	$\lambda^2 - 1$	$\lambda^2 - 1$
$\lambda_1 = 1; \lambda_2 = 1$	$\lambda_1 = -1; \lambda_2 = -1$	$\lambda_1 = -1; \lambda_2 = 1$	$\lambda_1 = -1; \lambda_2 = 1$

Note how the matrix trace (sum of entries in the matrix diagonal) is zero for both (17c) and (17d) (the negative counterpart of (17c)), as a consequence of which the characteristic polynomial results in the elimination of the λ term and, therefore, an identical polynomial for σ_Z and $-\sigma_Z$. This is unlike what we see for the identity (17a) and its negative counterpart (17b), each with its characteristic polynomial. Consequently, the eigenvalues (roots of the polynomial) are the same for σ_Z and $-\sigma_Z$ (again, unlike for the identity and its negative). There is no other pair of matrices in the Chomsky group for which this is true; so suppose we build on this elegant formal fact.

The reason for that is not algebraic but informational, since what we should be seeking, in any linguistic organization, is a formal system for which syntactic computations can be grounded with reasonable priors—then to build semantic and phonological nuances “on top”. If nature’s job is to decide on initial conditions for a computational system deploying the Chomsky matrices, there is no better formal solution, in terms of matrix elegance, than to have the base step in the system land on “the most elegant” (symmetric, orthonormal) matrix. Curiously, as we see from Table 2, precisely all the original Chomsky matrices (in the sense of literally interpreting his 1974 $[\pm N, \pm V]$ as in (1), with the items listed, in that order, in the matrix diagonal) land on Pauli’s σ_Z after squaring them. This is arguably significant too, under certain assumptions about Merge.

As mentioned already in Section 1, (external 1st) Merge, EM, is inherently asymmetrical for a head and the projection of its complement. That works directly for the operation after it has started. But it is computationally impossible to merge a head *to a projection* if we do not have, yet, any projection formed in a given derivation—in the pristine moment in which just two heads are selected for assembly. A natural way to address this is the idea in Guimarães [26] of allowing the self-merger of heads, at the base of a derivation. Then we can characterize EM as an operation resulting in an anti-symmetrical relation (i.e.: asymmetrical in all instances but the relation of an element to itself). But once again note that the results in Table 3 obtain for *all* Chomsky matrices:

Table 3. Self-products for the original (reinterpreted) Chomsky matrices.

$\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
---	---	--	--

In other words, self-merging (here understood as self-multiplication) of the nouns, verbs, adjectives, and prepositional representations results in σ_Z , obviously collapsing the formal results into identical representations. While algebraically this is entirely fine, for a symbolic system it is the equivalent of Paul Revere’s famous code having been the senseless “One if by land, one if by sea!”⁸ The whole point of a simple code is to have different signals representing *different* events, which is precisely what the self-multiplication in Table 3 denies, since matrix multiplication is not a structure-preserving operation. Is there a solution to this information impasse?

Actually, there is a simple one: phrasal axioms. We routinely take syntactic representations to map to some meaning, which we customarily do by axioms to establish relevant representations; e.g., mapping noun phrases to entities or sentences to truth values. So suppose we assume this:

(18) *Anchoring Axiom*

Only nouns self-merge.

While obviously (18) is not algebraic, it is as straightforward, as such, as stipulating that NPs map to entities. It is, of course, cognitively interesting why (18) should be, and one could imagine that the alien *heptapodes* in the movie *Arrival*, let’s say, had a different anchoring axiom in their cyclic view of time—perhaps for them only verbs self-merge. The point is that any information system with associated meaning must presume axioms of this sort, and this, as we see shortly, does work.

For initiating the computation in the self-merger of nouns—as the first product in Table 3, in Chomsky’s terms—limits the next steps in direct ways. This is best seen by considering the following graph, helpfully proposed to me by physicist Michael Jarret. In (19), the notation is algebraic, linguistic codes tracking the presumed substance. The Chomsky matrices come with a “hat” $\hat{}$, indicating their being taken as operators on the vector space of the Chomsky group. By the Anchoring axiom, this means a self-merging noun $\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$ (**boldfaced** in (19)) plays a dual role: as an operator on itself, with an argument of identical formal characteristics.

$$\begin{array}{c}
 (19) \\
 \begin{array}{ccccc}
 \begin{array}{c} \hat{V} \\ \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix} \end{array} & & \begin{array}{c} \hat{P} \\ \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix} \end{array} & & \begin{array}{c} \hat{A} \\ \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \end{array} \\
 \begin{array}{c} \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \\ \text{VP} \end{array} \leftarrow \begin{array}{c} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ \text{NP} \end{array} \begin{array}{c} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \\ \hat{P} \end{array} \begin{array}{c} \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \\ \text{PP} \end{array} \rightarrow \begin{array}{c} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \text{AP} \end{array} \\
 \begin{array}{c} \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \\ \hat{N} \end{array}
 \end{array}
 \end{array}$$

Note, also, how the Jarret graph only relates the categories within the Chomsky group that involve real entries in the top row, and how the Chomsky operators take one of two “twin” matrices—in the sense that they (NP, VP, PP, AP) share the *same eigenvalues and determinant* (−1 for NP, *i* for VP, −*i* for PP, 1 for AP projections). This is a consequence of the Anchoring axiom and “semiotic” assumptions: the idea that the computational system seeks to *maximize the connections among categories* in the Chomsky group, and thus even if other mappings are algebraically viable, they lead to disconnected graphs, or graphs without the recursive condition that (19) presupposes.

$$\begin{array}{c}
 (20) \\
 \begin{array}{ccccccc}
 \begin{array}{c} \hat{V} \\ \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix} \end{array} & & \begin{array}{c} \hat{P} \\ \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix} \end{array} & & \begin{array}{c} \hat{A} \\ \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \end{array} \\
 \text{determinant } i \leftarrow \begin{array}{c} \text{determinant } -1 \\ \text{VP} \end{array} \begin{array}{c} \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix} \\ \hat{N} \end{array} \begin{array}{c} \text{determinant } -i \\ \text{NP} \end{array} \begin{array}{c} \text{determinant } -i \rightarrow \text{determinant } 1 \\ \text{PP} \end{array} \begin{array}{c} \text{determinant } 1 \\ \text{AP} \end{array} \\
 \begin{array}{c} \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix} \\ \hat{N} \end{array}
 \end{array}
 \end{array}$$

The original Chomsky matrices can be seen as operators on “projections” within the Chomsky group, whose “labels” are the determinants listed in (20) (there are no other possible determinants for objects in the Chomsky group). I am using those words in quotes because, algebraically, the “projections” are simply matrices within the group that are not being treated as operators, while their “label” is nothing but their determinant (which is equivalent for “twin” matrices with the same trace and eigenvalues). The graph makes the following categorial generalization:

- (21) “Selection” conditions for operators in the Jarret Graph:
- a. Nouns may either self-merge or take PPs to NPs
 - b. Verbs take NPs to VPs
 - c. Prepositions take NPs to PPs
 - d. Adjectives take PPs to APs

What is remarkable about (19) is that the statements in (21), substantive though they may seem, have no semantic cause to them: they follow from the algebra, once the Anchoring Axiom is established and “semiotic” conditions (e.g., about involving all categories into a connected, recursive, graph) are presumed. These may map to the semantics, but not *because* of meaning.⁹

5. Non-Algebraic Assumptions in the Jarret Graph

To see the basic algebraic possibilities presumed in the Jarret Graph, consider the determinant products in Figure 1, where the graph, expressed in terms of the matrix determinants as in (20), correspond to the highlighted possibilities to the right of the table. Of course, one needs to consider why the other algebraic possibilities in Figure 1 (covering the entire space of combinations) do not result in a graph equivalent to the Jarret Graph, with grammatical reality.

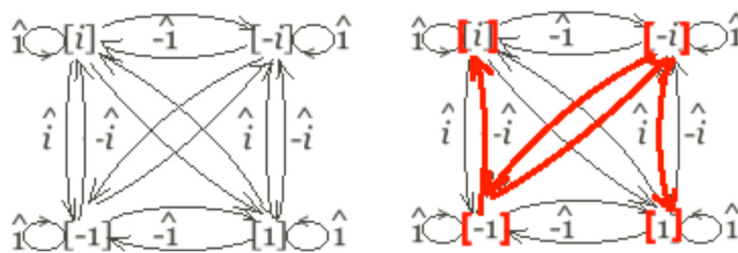


Figure 1. Determinant products for the matrices in the Chomsky group.

The answer is substantive: the Anchoring Axiom—plus semiotic/information assumptions about to be specified—yields the Jarret Graph as what the language faculty found within this algebraic space. A different species (anchoring the system on verbs, prepositions, or adjectives) would yield a different solution, as could a system that is not meant to be used to encode and share information.

To ponder why other possibilities do not arise, consider (22),¹⁰ which would wreak havoc on the “selection” generalizations in (21) (vis-à-vis the descriptively adequate graph in (19)):

$$\begin{array}{c}
 \hat{V} \quad \hat{A} \quad \hat{P} \\
 \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \quad \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix} \\
 \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}_{VP}, \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}_{NP} \xleftarrow{\quad} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_{NP} \xrightarrow{\quad} \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}_{PP}, \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}_{PP} \xrightarrow{\quad} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{AP}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}_{AP} \\
 \hat{N} \\
 \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}_N
 \end{array}$$

Here we have swapped the A in the Jarret Graph for the P element, preserving the rest. Note, however, that (22) cannot be obtained under the assumption of the Anchoring Axiom and the fact that the Chomsky category for nouns is $\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$. What we obtain under those presuppositions is (23):

$$\begin{array}{c}
 \hat{V} \quad \hat{A} \quad \hat{P} \\
 \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \quad \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix} \\
 \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}_{VP} \xleftarrow{\quad} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_{NP} \xrightarrow{\quad} \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}_{PP} \xrightarrow{\quad} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}_{AP} \\
 \hat{N} \\
 \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}_N
 \end{array}$$

This is because, after Chomsky’s $\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$ self-merges, it yields Pauli’s $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. In the present exercise, we are swapping the next operator (using an A $\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$, instead of a P $\begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}$); then the ensuing product is not the “twin” category $\begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}$ but, rather, again the initial $\begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$. This graph is as algebraically fine as the Jarret Graph, but it does not cover all eight categories in the Abelian subset of the Chomsky group: only six of them are involved. If the point of these combinatorics is to find ways to maximally involve them all, then (23) is simply invalid.

One could also consider a related problem as in (24), where we swap the V and the N in the Jarret graph. Although algebraically there is nothing wrong with this either, (24) directly violates the Anchoring Axiom, for the only way in which the system could

involve NP direct objects—all of which the graph covers. Evidently, the issue is to ponder in/di-transitives or clausal dependents, which is not a matter of the graph overgenerating.¹¹

Ultimately, more important than such considerations are further consequences the present approach affords. Note that only two matrix pairs in the Jarret Graph involve only real entries: those corresponding to $\pm\sigma_Z$ and $\pm I$. Matrices of that kind—with real values in the diagonal and also real or conjugate symmetrical values elsewhere, here zero—are called *Hermitian* and present particularly elegant properties, vis-à-vis non-Hermitian counterparts. In comparable situations in particle physics, these same operators, equipped with positive inner products (the central metric in a vector space), have real eigenvalues. This is key to representing physical quantities, since measurements correspond to real quantities. Interestingly, this separates what “exists” (is mathematically necessary for a system to work) from what is “observable” (definitely identified for measurement). If we suppose the way for the abstract syntax in that mathematical existence to interface with concrete external systems, or definite internal thought, is through equal certainty, then only Hermitian points in the Jarret Graph yield “interpretations”:

(29) *Interpretive Axiom*

Only Hermitian matrices are legible at the interfaces.

While not (logically) necessary, (29) makes sense and has architectural consequences: primitive semantic types—like “entity” or “predicate”—correspond to the “legible” (read: real) matrix projections.¹² None of the Chomsky matrices (lexical operators) are “legible” in that sense, so their meaning is entirely contextual: it becomes viable in the domain of sentence grammar. Interpretive Axiom (29) should be seen in the same light as Anchoring Axiom (18): neither is a piecemeal statement about the syntax/semantics interface, though they both have radical consequences.

6. Beyond the Jarret Graph

The discussion in the previous section pertains only to the four Chomsky matrices and their “projections” via matrix multiplication, yielding the Abelian group that the Jarret Graph lives on. As noted, we also have a larger group, involving extensions of the Chomsky matrices, obtained by flipping around the attributes. Indeed, once Pauli’s σ_Z is invoked, it is natural to ask whether the other Pauli matrices ($\sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$) should also play a role in the system. Multiplication by either one of those yields an interesting group of 32 matrices, 16 positive and their negative counterparts, that Orús et al. [25] dubbed the Chomsky-Pauli group G_{CP} . Readers will be easily able to spot here the Abelian group arrayed into the Jarret Graph, as well as the entire Chomsky group (all of them using diagonal matrices only; Z and I , the identity, are two of Pauli’s matrices and the mnemonic C is a shorthand for the Chomsky objects). In addition to that, the G_{CP} contains also non-diagonal matrices; inasmuch as these are symmetrical with the Chomsky objects, they are notated in (30) with an S , and they align with Pauli’s X and Y .

(30) Chomsky-Pauli Group G_{CP}

$$\begin{array}{llll}
 I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; & -I = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}; & iI = \begin{bmatrix} i & 0 \\ 0 & i \end{bmatrix}; & -iI = \begin{bmatrix} -i & 0 \\ 0 & -i \end{bmatrix} \\
 X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; & -X = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}; & iX = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}; & -iX = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} \\
 Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}; & -Y = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}; & iY = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}; & -iY = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
 Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; & -Z = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}; & iZ = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}; & -iZ = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} \\
 C1 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}; & -C1 = \begin{bmatrix} -1 & 0 \\ 0 & i \end{bmatrix}; & iC1 = \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}; & -iC1 = \begin{bmatrix} -i & 0 \\ 0 & -1 \end{bmatrix} \\
 C2 = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}; & -C2 = \begin{bmatrix} -1 & 0 \\ 0 & -i \end{bmatrix}; & iC2 = \begin{bmatrix} i & 0 \\ 0 & -1 \end{bmatrix}; & -iC2 = \begin{bmatrix} -i & 0 \\ 0 & 1 \end{bmatrix} \\
 S1 = \begin{bmatrix} 0 & 1 \\ -i & 0 \end{bmatrix}; & -S1 = \begin{bmatrix} 0 & -1 \\ i & 0 \end{bmatrix}; & iS1 = \begin{bmatrix} 0 & i \\ 1 & 0 \end{bmatrix}; & -iS1 = \begin{bmatrix} 0 & -i \\ -1 & 0 \end{bmatrix} \\
 S2 = \begin{bmatrix} 0 & 1 \\ i & 0 \end{bmatrix}; & -S2 = \begin{bmatrix} 0 & -1 \\ -i & 0 \end{bmatrix}; & iS2 = \begin{bmatrix} 0 & i \\ -1 & 0 \end{bmatrix}; & -iS2 = \begin{bmatrix} 0 & -i \\ 1 & 0 \end{bmatrix}
 \end{array}$$

What I am calling “the Chomsky group” contains only diagonal matrices with mixed (real and complex) entries (the portion arrayed into the Jarret graph also being a group, indeed an Abelian one); the entire set of matrices with mixed entries is actually not a group. However, readers can check (see Note 7) how the extension of the vector space in (14) as in (31) does yield a multiplicative group:

$$(31) \quad a. \begin{bmatrix} \pm 1, \pm i & 0 \\ 0 & \pm 1, \pm i \end{bmatrix} \quad b. \begin{bmatrix} 0 & \pm 1, \pm i \\ \pm 1, \pm i & 0 \end{bmatrix}$$

Although the present exercise is not the place to try this, one needs to determine how, by the same methods and results, we can use the matrices in the G_{CP} as operators to cover some “grammatical category” space (for T, Det, C, and so on). Bear in mind that the “grammatical structure” never starts a derivation, it is presumed to be added on “lexical structure”—which the Jarret graph grounds. Ideally, the way to connect the Jarret graph to a Super Jarret Graph is by applying the extended-Chomsky categories (“grammatical” elements) to the Jarret Graph, in the process yielding a related object with “more structure”, targeting each of the projections in (16). We should not need more than that, the super-graph preserving the basic symmetries of the Jarret Graph, including the determinants/labels we have established for basic “projection”—which, in this sense, could be seen as an “extended projection” into a layer of functional structure. One should also not need further phrasal axioms beyond the already stipulated Anchoring and Interpretive axioms, so that whatever semantics ensues follows from, rather than determine, the algebraic fate of the graph. Figure 2 shows a cartoon version of the SJG from work in progress.

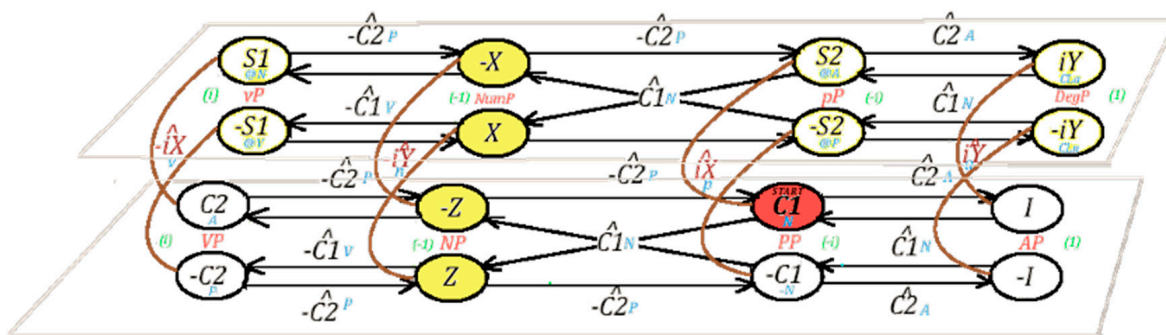


Figure 2. A (tentative) version of the SJG.

The Super Jarret Graph contains the Jarret Graph as a sub-graph (in the lower tier), so that the information flow can start in the emphasized node and go from any of the nodes in the more basic “lexical” graph to this extension, where the observable/interpretable

categories (all Hermitian) are highlighted. Bear in mind that the super-graph contains only sixteen of the categories in the G_{CP} , presupposing further nuances if the entire group has grammatical significance. The super-graph should be seen as an empirical claim based on the formalism deployed, presuming the Jarret graph with all the implications above. The goal, as a result of the various dynamics, should be to obtain a “periodic table” of categorial elements of this elementary sort, ideally without having to invoke new dimensions or higher orders. Figure 3 below is based on the Jarret Graph as already discussed, as well as Table 3 and extensions from the presuppositions behind it: semiotic and phrasal axioms, together with “churning calculations” in the way any “categorial grammar” of this sort may function, applying “type changing” rules—here, matrix multiplication. The question marks in the table are meant to suggest that we haven’t discussed here any specificities for those categories or associated projections. In general, with any “periodic table” of this sort, one either needs to predict why a given gap exists in the paradigm or otherwise argue for a given underlying category “fitting the bill”.

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

Figure 3. “Periodic Table” of syntactic categories based on the G_{CP} .

Note that superposing objects in the G_{CP} and normalizing the results (to entries $\pm 1, \pm i$) has a consequence for what one may think of as “phrasalization” of the categories under discussion:

$$(32) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

If this sort of matrix-superposition is grammatically relevant (through the M-matrices in Section 2), it may have a dramatic effect on the interactions of the ensuing system.

Let’s start by recalling the self-products in Table 3, generalizing them for the matrices in (32), as follows:

$$\begin{aligned}
 (33) \quad & \text{a. } \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^4 = \dots = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
 & \text{b. } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^5 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \dots \\
 & \text{c. } \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}^2 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^4 = \begin{bmatrix} 8 & 8 \\ 8 & 8 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^5 = \begin{bmatrix} 16 & 16 \\ 16 & 16 \end{bmatrix}; \dots
 \end{aligned}$$

Observe how only the successive powers for matrices as in (33c) “grow”, not those as in either (33a) or (33b). The difference between each condition is easy to state through the spectral radius $\rho(A)$ defined in (7a), from which the system’s derivational entropy is calculated, as repeated now:

(34.) *Derivational entropy*

- a. For spectral radius $\rho(A)$, square matrix A ’s largest absolute value λ_{\max} of A ’s eigenvalues,
- b. A ’s topological derivational entropy $h_T = \log_2 \lambda_{\max}$.

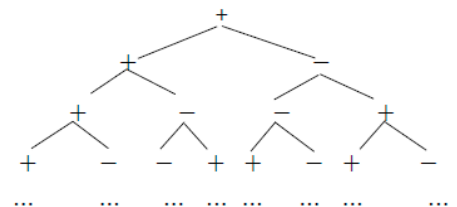
For either (33a), (33b)—or any of the matrices in the G_{CP} —for A matrix in the sequence of powers, $\rho(A)$ is always 1. In contrast, some of the superposed matrices (arising from summing matrices within G_{CP}) have $\rho(A)$ larger than 1.¹³ Consequently, the topological derivational entropy h_T of the matrices in (33a) or (33b) is $\log_2 \lambda_{\max} = \log_2 1 = 0$; in contrast, for h_T of the matrices in (33c) $\log_2 \lambda_{\max} = \log_2 2 = 1$. This topological entropy matters for interpreting systems’ growth as they (try to) expand in L-system fashion, via rewrite or EM merge processes.

7. Categories & Interactions Redux

At the end of Section 3 I promised to consider the bold hypothesis that the ontology of the Chomsky/Pauli 2×2 objects is ultimately related, if not equatable, to the Medeiros interpretation of L-trees topologies in square matrix fashion. Intuitively, just as it is obviously relevant for a power sequence of the sort in (33c) to grow, one may take it to be equally relevant for the power sequence of a related matrix to stay “cyclic”, as in (33a) or (33b). The latter condition arises for a matrix A with spectral radius $\rho(A) = 1$, whose topological derivational entropy is zero and, thus, the subsequent powers of A fall into a multiplicative group. Computationally, the group in point could be seen as Markovian, “stable” in that very repetition, perhaps like “earworms” in the stuck song syndrome (<https://en.wikipedia.org/wiki/Earworm> (accessed on 28 August 2022)). While this is meant just as a metaphor for now, the idea is that the formal representation of certain matrices arising from multiplication of a computational seed disallows growth into a phrasal topology that could permit an associated phrasing. Again, all matrices within the G_{CP} have the formal condition of their power sequences being “cyclic” in this sense, while some of their (normalized) superpositions can be shown to allow for the fractal growth the standard L-systems imply—starting with the simple (33c), which corresponds to a trivial Lindenmayer expansion as in (35):

(35)

- a. $+ \rightarrow ++$ (equivalently $+, -$ merge to form $\{+, -\}$)
- b. $- \rightarrow -+$ (equivalently $-, +$ merge to form $\{-, +\}$)



From this perspective, while elements in the G_{CP} need to be necessarily categorial in the Markovian cyclicity of corresponding powers, the mere superposition of its elements is enough, under the right conditions, to allow for the phrasalization of corresponding categories—via the same algebra.

For concreteness, consider the sum of three Pauli matrices in (36a), normalized to (36b):¹⁴

in a text like this (and see Note 16). To be sure, that alone will not distinguish, let's say, *write* from *read*, *page* from *noun*, or any of the thousands of instantiations one carries in one's head—worse: anyone *could* carry in one's head, in principle, and for any language in the past or the future. But just as assigning categorial features [+N, -V] does not, in itself, distinguish all possible nouns there could be, so too the implied algebra is meant to combine with whatever system may be in place for the further nuances one may want to add to the relevant vector space.¹⁷ Once we establish the vectorial units of said space as priors, whatever substantive dimensions are added to the foundations are as immaterial (to the structure itself) as the notes of a melody one can deploy within a given harmony. Importantly—while there are grammatical processes that depend on the implementation of the features this system presupposes (by changing values and attributes as relevant, to distinguish nouns from verbs, etc.)—there is no syntactic process that depends on the exact implementation of the relevant vector spaces. Thus, there are syntactic operations that target nouns *or* verbs, but none that target the noun *page* as opposed to any other or the verb *write*, but not the verb *read*, etc.¹⁸ What we are attempting to establish now is just the algebraic foundation of the vector space where syntax lives.¹⁹

8. Conclusions

Verbs and, especially, nouns are fundamental. These elements are not just the cognitive back-bone of the grammatical system, they also feature prominently in the main processes the grammar instantiates: θ -dependencies, Case, agreement, or construal relations. Here I have made nouns axiomatically start a derivational system based on matrix multiplication, instantiating a reflexive base condition. That algebraic system can be accessed by formally pursuing an intuition dating back to antiquity, which has an abstract “nominal” dimension “at right angles” to a maximally distinct “verbal” dimension. In the system proposed, operating on nouns yields a measurable/observable (Hermitian matrix), which in turn limits other potential combinations with abstract lexical categories; this is what I have been calling the Jarret Graph, which was argued to underlie complement interactions for lexical heads.

There are two ways in which that decision is not enough. First, there are obvious categories in grammar that do not reduce to the basic lexical four; second, equally obviously, there are relations beyond head-complement dependencies. The first issue can be dealt with within our formalism by exploring it to its fullest: literally turning entries around; this is how we went from the Abelian group of categories within the Jarret Graph, first, into the Chomsky group (flipping the entries in the diagonal) and, next, into the G_{CP} , which includes all of the Pauli group and extends it (by mixing real and complex entities, or vertically flipping the elements in the Chomsky group into their symmetrical counterparts). Although this short paper cannot fulfill the promise of distributing the formal entities in that group into a full-fledged “periodic table” of grammatical categories, mapping them within the formal system amounts to a dissertation exercise, with familiar categorial-grammar methods (we know the input, suspect the output, have an excellent sense of the phonetics and an informed idea of the semantics, so it is all a matter of “aligning the Rubik cube”). A harder—more interesting—question is how to go beyond head-complement relations, in particular into head-specifier relations and the long-range correlations they afford.

That one shouldn't trivially align head-complement and head-specifier relations ought to need not argument, in syntactic, semantic, and phonological terms. In minimalist parlance, head-complement relations define the so-called *domain* D of a phase, the rest being its *edge* E. As emphasized in Uriagereka [30] (chapter 5), there are several grammatical conditions taking place within Ds, from theme selection with projection consequences to agreement probing, including partitive case and head-to-head dependencies (incorporation, affixation, light verbs, etc.). In contrast, Es determine everything from external argument taking to structural Case, together with all sorts of (successive) displacements and entanglements (see [27]). Such E-related super-configurational conditions are decidedly more complex than the corresponding D ones, of a (phrasal) configurational sort. These

distinctions have dramatic consequences, also, for the mapping to the interfaces, from intonational units to event-participation and the convoluted gymnastics of generalized quantification, or intertwined construals and ellipses too numerous to mention here, with consequences for pragmatic conditions and the articulation of Familiar/Novel information. In this exercise, we just looked at the configurational relations, which is only a stepping stone towards understanding the more elaborate ones that presuppose them.

A novel idea explored here is that the presumed difference between categories and interactions—in a context of the head-complement sort—are less dramatic than it may seem, before analyzing the matter algebraically. The formal intuition is no different from why the powers for any number k^n growing exponentially with the size of n do not for the subcase $k = 1$. We are dealing with matrices, but they too have powers, and their effect on their spectral radius (per (7)) does depend on whether it is 1 or more, thereby affecting the system's topological derivational entropy. In the former instance the power sequence “cycles back” to the origin, after hitting the identity matrix, while in the latter the power series yields fractal growth instead.²⁰ We have capitalized on that distinction to separate “cyclic/Markovian” categories from derivational topologies that may carry phrasal combinations. To go on with the Tour metaphor in Section 2, the “earworm” stages would have to be designed by M.C. Escher, to cycle back after a topological trick involving a Möbius Strip! In contrast, all other stages would shoot out to infinitude, creating a space for exploration of the relevant terrain that can take arbitrarily many twists and turns, within the restricted topology. Algebraically, the difference is small: mere symmetry within the underlying square matrices vs. breaking that symmetry into some result that could be chaotic.

Although I have not sketched the idea beyond the programmatic, it reduces to linguistic knowledge being accumulated in a (mental?) lexicon by way of the presuppositional effect of depleting words. Although this has a Hebbian ring to it, which connectionist networks have taken to new depths through state-of-the-art algorithms, the issue is what could the complex entries be that I have crucially introduced. These cannot relate to “neurons that fire together wiring together”, if that is being presupposed as some sort of (distant) underlying brain physiology to carry some version of the system . . . That said, by now there are non-obvious conditions neurons are known to perform, which have little to do with the usual firing in the direction of the axon. Neurons, for instance, “fire backwards” while mammals sleep (for reasons that might relate to the consolidation of memories—see [31]). Also, there appear to be physiological dependencies “at right angles” with standard neuronal firing. At a neuronal level, one can imagine modeling the balance between excitatory postsynaptic potentials vs. inhibitory postsynaptic potentials in terms of negative vs. positive scalars in a network; at the same time, Hebbian plasticity is known to amplify correlations in neural circuits, creating a positive feedback loop that renders circuits unstable, which needs to be constrained. Heterosynaptic plasticity apparently serves a central homeostatic role to cause pathway unspecific synaptic changes in the opposite direction as Hebbian plasticity (see, e.g., [32]).²¹ The scope of these changes appear to be global in the dendrites, suggesting that, aside from positive/negative interactions among neurons, to yield standard Hebbian learning and some inhibitory correlate, something else entirely, and orthogonal to such interactions (though still numerical in their own right) may also be at play.

Those comments are meant within the relatively new, but entirely classical, domain of neuroscience. Even more radically, Fisher [33], Straub et al. [34], speculate with quantum processing with nuclear spins in the brain, identifying phosphorus as “the unique biological element with a nuclear spin that can serve as a qubit for such putative quantum processing—a neural qubit—while the phosphate ion is the only possible qubit-transporter.” This is presupposing that central quantum-processing requires quantum entanglement, which Fisher argues can happen as a result of an enzyme-catalyzed reaction that breaks a pyrophosphate ion into two phosphate ions. The admittedly drastic idea is that multiple entangled relevant molecules “triggering non-local quantum correlations of neuron firing

rates, would provide the key mechanism for neural quantum processing". Evidently, if anything this dramatic were at stake, much of what is known about neuroscience in general would have to be rethought. I mention it here simply as an indication that all bets probably should be off when it comes to interpreting the relatively simple observables discussed here, which need not go into such old chestnuts as awareness, consciousness, morality, and the like. Ordinary syntax poses related and more tractable questions.

Be that as it may, an individual "use" of language by any speaker may be seen as relevant to convey some truths (or lies!) about the world. However, the social effect seems to be to create an ensemble of weighted concepts that get harbored in categories transmitted from a generation to the next. In this sense, while (obviously) there is a portion of that repository that I happen to store in my head, you in yours, etc., there is also a societal entity that persists, probably the way economies or social systems may, drifting away as the case may be. For those structures, it is immaterial what I or you may be saying or how it affects our lives, as it is the fact *that* we are using it. This even allows for the possibility of some of the complex algebraic possibilities explored here having nothing to do with what gets literally "recorded" in actual brains, possibly existing in a reality that one may wish to call Platonic—but with relatively little added value to such a claim. Unless, of course, economies and social systems "evolving" in history are also Platonic.

My concrete claims boil down to two. One, that while syntactic deployment requires an open-ended topology (with non-zero topological derivational entropy) to allow for recursive thoughts, lexical storage demands, instead, a cyclical scaffolding (with zero entropy) to let ideas repetitively cohere into entities stable enough to last one's lifetime and, ideally, get passed on, in some simple Markovian fashion. Here we have only studied what amounts to the hypothesized priors of such a system: its underlying scaffolding—therein the second claim. For the symmetries and asymmetries studied here to make any sense, that scaffolding must involve complex (non-observable) dimensions. While these are not obvious to relate to brain interactions (even with hand-waving), they are one of the most interesting aspects of the proposal.

The final ace in the sleeve for this sort of proposal is how it relates to ideas in Smolensky & Legendre [35], starting from (vastly) different presuppositions, since the latter work is coming from a connectionist tradition, while the present one is explored within symbolic conventions. Interestingly, in some regards the present approach, upon tensorizing the G_{CP} in ways sketched in Orús et al. [25], is entirely compatible with that model, and thus can preserve its virtues, including a way to directly map the relevant representations to actual phrase-markers (not just L-trees). More importantly, both systems present the ability to (formally) state entanglements upon relevant superpositions—what syntacticians customarily call "chains of occurrences". While this idea is only offered as a proof of concept in the connectionist instance, and without obvious connections to actual observables across languages, I intend to show in a sequel to this piece how the present formulation has the same mathematical result as a direct consequence, moreover tracking grammatical facts in a more direct fashion (e.g., allowing us to state differences between voice and information questions). It is interesting that one should converge in a similar math from very different assumptions (in the present instance, asking how to generalize from the most modest formal objects to matrices combining them, step by step). While data-driven analyses involve the same math that can be deployed over the group we have begun to formally analyze, we "got there" from first principles that we still teach in our introductory linguistics classes, dating back to the dawn of linguistics. While this could certainly be mere coincidence, it may also be an indication that some progress may be at hand as these matters are further explored.

Funding: This research received no external funding.

Acknowledgments: I want to thank editor Peter Kosta for his interest in this work, as well as two anonymous reviewers. This piece is part of a larger project in collaboration with Román Orús, Michael Jarret, and the late Roger Martin. I want to also thank Naomi Feldman for the most fun seminar we

jointly put together for a great group of graduate students, whose questions certainly helped me focus many of these ideas. All errors and misinterpretations remain my responsibility.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- 1 Here I concentrate on head-complement relations only. One of the reviewers reasonably wonders “how this approach works with complex syntactic operations such as movement”; while this is very much part of the program, that matter cannot be seriously addressed within a chapter like this. As the reviewer also notes, “the functional structure above lexical items is related to extended projections . . . (tentatively presented in Section 6)”. S/he wonders whether “restrictions on the kind and quantity of functional items/structure or some parallelisms among the extended projection of lexical items”; the answer is: Yes, as such formal symmetries are crucial to this approach. That said, I need to ask readers to be patient with the exposition, which cannot be rushed—or it will not make algebraic sense. It is my hope that the separate pieces that articulate this ensemble will come out together in the form of chapters in a single monograph.
- 2 The “*” represents the Kleene star operator or closure, named in honor of Stephen Cole Kleene. For a given set like T, T* is defined as the smallest superset of T that contains the empty string and is closed under concatenation. See https://en.wikipedia.org/wiki/Kleene_star (accessed on 28 August 2022).
- 3 One of the reviewers takes me to assume “the verb/noun distinction as . . . radical”, noting how others see the distinction less dramatically. Careful readers will see that this entire exercise is meaningless if, in point of fact, the distinction is not dramatic to the extreme of (relevant) orthogonalities. This can, of course, mean that the project is wrong; if so, so be it. But formally it would make no sense to “weaken” the claim, so I will stick to the presumptions. Perhaps it should be clear that the notion of “noun” and “verb” that I am after has nothing to do with lexical instantiations (which can be identical); what matters in the present context is the (I think radical) fact that, for example, event semantics is articulated in a way different—as a “sentence spine”—than arguments thereof; the latter being foundational nouns. Of course, one can verbalize nouns or nominalize verbs, but what is being sought in the present context are the underlying dimensions.
- 4 If we consider the power series, the ensuing aggregation relates to the total number of nodes in the L-tree.
- 5 And note also that A^0 is the identity matrix, whose $\rho(A^0) = 1$, so $\rho(A^1)/\rho(A^0)$ is of course also $\rho(A^1)$.
- 6 Lest this be confusing, “linear operator” has nothing to do with “linearization” in the sense used in syntax (to order terminals in the speech signal). I say this because one of the reviewers takes me to be alluding to “the issue of linearization (and, implicitly, labelling)”. My use of the notion is in the *linear algebra* sense (so as a linear map within a vector space), to denote a very tight operation that keeps basic relations unchanged. I know that, as the reviewer points out, for some authors “linearization” (in the syntactic sense) relates to “labeling”, which I have nothing to contribute to (indeed, as mentioned at the end of Section 2, the M-matrices code no “linear ordering” in that syntactic sense among terminals, which is left as a separate problem at right angles with phrasal topology). This is not to say that the present system does not care about “labels”: here such abstract (non-terminal) representations are matrix *determinants* (nothing to do with “determiners” in the syntactic sense); these are algebraically related to matrix *traces* as in Section 2 (no relation to syntactic “traces” either). The terminological nightmare is what it is, but the truth is the terms in point have a much older tradition in algebraic systems, which syntacticians are not always careful in distinguishing; alas, even the term *linear transformation* is used in math, and it turns out to be an interesting (and difficult) question whether this relates to what syntacticians call a “transformation”, with a structural description and a structural change—I believe it does, but this must be argued, which I will not go into here. Unfortunately, there is nothing much I can do about any of this, beyond noting the issue and proceeding with the relevant caution.
- 7 Interested readers can check these expressions in the Matlab platform or the popular [Wolframalpha.com](https://www.wolframalpha.com) (accessed on 28 August 2022).
- 8 For those unfamiliar with US independence, Revere prepared a code for the colonists of Charlestown about troop movements, in terms of their route being signaled by the number of lanterns in a church steeple: “one if by land, two if by sea”. See https://en.wikipedia.org/wiki/Paul_Revere%27s_Midnight_Ride (accessed on 28 August 2022).
- 9 A reviewer asks the question, about generalization (21), of whether all N/V/A/P elements follow it. I discuss this further at the end of Section 5, but I must also raise a deeper concern. That generalization is what the Jarret Graph gives us, the graph itself following from algebraic assumptions coupled with reasonable grounding conditions (of the semiotic sort because that is what grounds the system). We may choose to ignore this and continue to search for other answers or may, instead, be moved by whatever portion of the “empirical cake” we get to describe this way, with absolutely no semantic assumptions (beyond the Ur Anchoring Axiom). This is an aesthetic decision that I have little to contribute to. In a similar vein, the reviewer asks whether “a more detailed mention of several categories . . . would be helpful”. We certainly all agree that there are more “parts of speech”, and even that, as Chomsky [2] suggested already in 1974, that we could add further dimensions to the system to capture those—one can always add more dimensions. But the issue is how much mileage one can get out of the smallest amount of formal machinery. At the end of the chapter, I suggest some natural extensions that do not entail expanding the dimensions of the system or its underlying algebra. But it should be clear that no formal system ever describes a “totality of the natural phenomena” it is attempting to describe. Even physics, the obvious model to follow, cannot describe the universe with the same math, which

happens to also be (an extension of) linear algebra. It would seem unreasonable (and not very illuminating) to demand more of linguistics than we do of any of the other natural sciences.

10 I thank MG Hirsch for the question and discussion of its significance.

11 The linguist may note ditransitive verbs are bi-clausal in some languages, differences in verbal (periphrastic or synthetic) instantiations, or many idiosyncrasies. Worrying about differences now, however, is the equivalent of preventing Galileo from tabling the behavior of balloons while studying the falling of objects: the foundations of a scientific theory are built on patterns, not exceptions—to better understand the exceptions and beyond.

12 Primitive types differ depending on the semantic theory assumed; a larger issue (particularly if the Interpretive Axiom is assumed) is whether only certain portions of the syntax are, in fact, interpretable.

13 Some of the superpositions also yield the zero matrix; e.g., any matrix plus its negative.

14 A normalization can mean very different things in different formal contexts. Here we could take it as a way to achieve an interpretation with matrix entries that are not larger in magnitude than 1, which can have a probabilistic interpretation. Note, for concreteness that: (i) $\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. So the second element in the sum in (i) can be seen as a “normalizer”. This of course is stipulative, so one must explore under what algebraic circumstances one can naturally go from one of the matrices in the C_{CP} group to an M-matrix associated to a phrasal topology, like the one in the right-hand side equation (i). I will not study this here, but I can anticipate that this is behind the idea of “chain” formation in the present system, involving the supra-phrasal linkage of specifiers, through Internal Merge.

15 None of that is to say that anything presumed here is computationally trivial or even obvious. This starts with (37b), which is easy to state in “projection” terms, but much less so in Merge (or any BPS) terms. I am setting aside discussion of this important nuances simply to make the larger points.

16 The Pauli group is algebraically equivalent to the 4D quaternions postulated by William Hamilton in 1843, which are behind the virtual reality graphics in vogue these days. Going from the Pauli group to the G_{CP} can be seen as a rotation on the Pauli group, of the sort Hamilton discovered vis-à-vis the complex plane—so it is possible that the syntactic system (if it does live in the G_{CP}) is really an 8D representation.

17 This is meant very literally. As Naomi Feldman observes, the implied vectorial algebra can combine through a tensor product with any vectorial system (e.g., of a visual sort or any other), without this changing the basic underlying “syntactic” structure presumed here. I will not develop the point now, but it is mathematically rather direct, given that tensor products are structure-preserving.

18 This insight was first presented in Bresnan [28], p. 200, when noting how no language has a rule extraposing phrases involving the concept redness. This is generally true for any such process and any such feature. I thank Howard Lasnik for the reference and discussion relating to this topic.

19 This touches upon an issue Peter Kosta raises, regarding language acquisition. The short answer is that the present system is too abstract to bear on that. But a long answer is more interesting: the system does predict that nouns are fundamental in anchoring syntax, which can be tested empirically. Alison Brooks mentioned to me a fascinating counter-example: San children apparently start linguistically acquiring verbs before nouns, which she associates to the (itself surprising) facts that nouns, unlike verbs, exhibit (difficult) click phonemes. Brooks also insightfully notes, however, that the toddlers use pointing in place of nouns! If this is ultimately correct, it is grist for a mill Jackendoff [29] emphasized: that the linguistic system crucially connects to vision. My (modest) contribution to this would be that linear algebra can help with that connection, since one needs it for the visual system—and the present project argues that it is needed in language too.

20 The issue relates to a comparison between Fourier and Taylor series, both function decompositions represented as linear combinations of countable sets of functions, thereby specified by a coefficient sequence. The intuition is “categories” correspond to periodic functions while interactions “live” in the topological space of fractal L-systems that relate to the infinite sum of powers in a Taylor series. While abstractly similar, these differ in that the computation of a Taylor series is “local”, unlike the computation of a Fourier series.

21 The balance across synaptic conditions in a normal brain presumes homeostatic mechanisms and, in particular, the goal of maintaining excitation/inhibition balance and total activity at the network level, affecting synaptic weights. Although this is speculation, the thought is that these global forms of balance presume orthogonal conditions that one would hope to model in terms of complex vectors as invoked here.

References

1. Lyons, J. *Introduction to Theoretical Linguistics*; Cambridge University Press: Cambridge, MA, USA, 1968.
2. Chomsky, N. *The Amherst Lectures*; University of Paris VII: Paris, France, 1974.
3. Jackendoff, R. *X'-Syntax: A Study of Phrase Structure*; MIT Press: Cambridge, MA, USA, 1977.
4. Speas, M. *Phrase Structure in Natural Language*; Kluwer: Dordrecht, The Netherlands, 1990.
5. Muysken, P. Parametrizing the notion Head. *J. Linguist. Res.* **1982**, *2*, 57–76.
6. Kayne, R. *The Anti-Symmetry of Syntax*; MIT Press: Cambridge, MA, USA, 1994.

7. Chomsky, N. Bare Phrase Structure. In *Evolution and Revolution in Linguistic Theory: Essays in Honor of Carlos Otero*; Campos, H., Kempchinsky, P., Eds.; Georgetown University Press: Washington, DC, USA, 1994.
8. Jakobson, R.; Fant, G.; Halle, M. *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*; MIT Press: Cambridge, MA, USA, 1952.
9. Jakobson, R.; Halle, M. *Fundamentals of Language*; Mouton: The Hague, The Netherlands, 1971.
10. Kent, R. *De Lingua Latina*; Varro, M.T., Translator; Heinemann: London, UK, 1938.
11. Palmer, S. *1999 Vision Science*; MIT Press: Cambridge, MA, USA, 1999.
12. Poeppel, D. The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* **2003**, *41*, 245–255. [[CrossRef](#)]
13. Eiseley, L. *Darwin’s Century*; Doubleday Anchor Book: Garden City, NY, USA, 1961.
14. Chomsky, N. The Logical Structure of Linguistic Theory. Harvard/MIT mimeograph, partly submitted as UPenn. Ph.D. Thesis, 1955. [Published in part, 1975, Plenum: New York, USA].
15. Stowell, T. Origins of Phrase Structure. Ph.D. Thesis, MIT, Cambridge, MA, USA, 1981.
16. Chomsky, N. *Lectures on Government and Binding*; Foris: Dordrecht, The Netherlands, 1981.
17. Lasnik, H.; Kupin, J. A restrictive theory of transformational grammar. *Theor. Linguist.* **1977**, *4*, 173–196. [[CrossRef](#)]
18. Lindenmayer, A. Mathematical models for cellular interactions in development II. Simple and branching filaments with two-sided inputs. *J. Theor. Biol.* **1968**, *18*, 300–315. [[CrossRef](#)]
19. Boeckx, C.; Carnie, A.; Medeiros, D. Some Consequences of Natural Law in Syntactic Structure. Master’s Thesis, University of Arizona, Tucson, AZ, USA, Harvard University, Cambridge, MA, USA, 2005.
20. Medeiros, D. Economy of Command. Ph.D. Thesis, University of Arizona, Tucson, AZ, USA, 2012.
21. King, C. Some Properties of the Fibonacci Numbers. Master’s Thesis, San José State College, San Jose, CA, USA, 1960.
22. Uriagereka, J. *Biolinguistic Investigations and the Formal Language Hierarchy*; Routledge: London, UK, 2018.
23. Ott, E. *Chaos in Dynamical Systems*; Cambridge University Press: New York, NY, USA, 1993.
24. Uriagereka, J. Forthcoming. “Minimalist Goals”. In *The Cambridge Handbook of Minimalism*; Grohmann, K., Leivada, E., Eds.; CUP: Cambridge, MA, USA.
25. Orús, R.; Martin, R.; Uriagereka, J. Mathematical Foundations of Matrix Syntax. *arXiv* **2017**, arXiv:1710.00372.
26. Guimarães, M. In Defense of Vacuous Projections in Bare Phrase Structure. In *University of Maryland Working Papers in Linguistics*; The University of Maryland: College Park, MD, USA, 2000; Volume 9.
27. Lasnik, H.; Uriagereka, J. *Structure: Concepts, Consequences, Interactions*; MIT Press: Cambridge, MA, USA, 2022.
28. Bresnan, J. Theory of Complementation in English Syntax. Ph.D. Thesis, MIT, Cambridge, MA, USA, 1972.
29. Jackendoff, R. *Foundations of Language*; Oxford University Press: Oxford, UK, 2002.
30. Uriagereka, J. *Spell-Out and the Minimalist Program*; Oxford University Press: Oxford, UK, 2012.
31. Bukalo, O.; Campanac, E.; Hoffman, D.A.; Fields, R.D. Synaptic plasticity by antidromic firing during hippocampal network oscillations. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5175–5180. [[CrossRef](#)] [[PubMed](#)]
32. Chistiakova, M.; Bannon, N.; Chen, J.-Y.; Bazhenov, M.; Volgushev, M. Homeostatic role of heterosynaptic plasticity: Models and experiments. *Front. Comput. Neurosci.* **2015**, *9*, 89. [[CrossRef](#)] [[PubMed](#)]
33. Fisher, M.P. Quantum cognition: The possibility of processing with nuclear spins in the brain. *Ann. Phys.* **2015**, *362*, 593–602. [[CrossRef](#)]
34. Straub, J.; Nowotarski, M.; Lu, J.; Sheth, T.; Fisher, M.; Helgeson, M.; Jerschow, A.; Han, S. Phosphates form spectroscopically dark state assemblies in common aqueous solutions. *PNAS* **2021**, *20*, 1–8.
35. Smolensky, P.; Legendre, G. *The Harmonic Mind*; MIT Press: Cambridge, MA, USA, 2005.