Matrix Syntax
Roger Martin, Román Orús & Juan Uriagereka

1. *Preliminaries focusing on the Trouble with Chains*

While it may not be necessary for any analytical science to be quantitative in order to be taken seriously, gaining quantitative traction—if natural within a discipline's subject matter—can be an advantage. This is because of the rigor one can associate to calculations, but more generally because the level of predictions and accuracy of testing can gain a different scope. Our project can be seen, in practical terms, as a way to implement that desideratum within well-known parameters.

Our project stems from generative grammar and the Computational Theory of Mind, and as such it is deeply concerned with the nature of lexical categories, phrases, various sorts of merge, Agree, displacement, chains, control, ellipsis, rules of construal and other such notions that have arisen from a long tradition of theoretical investigations into the structure of the human language faculty. All the machinery that, in particular, the Minimalist Program uses constitutes our basic repository.

At the same time, within a specifically minimalist approach, we worry about long-range correlations and how best to tackle them. Perhaps the most obvious such concern is the very notion *grammatical transformation*, which we presuppose. Paramount among the issues that these devices create is the fact that their interpretation is *distributed* (both in phonetic and semantic terms). Much discussion over the years—as well as posturing—has gone over how to interpret that. In short, we don't know how to, not in classical computational terms (and see Colins & Stabler 2016 for essentially the same admission).

Consider for instance the situation in (1):

(1) Friends of each other$_i$ seemed to the Obamas t_i to appear to the Bushes t_i to have shown up unannounced at the White House.

A. Friends of Barack seemed to Michelle and friends of Michelle seemed to Barack to appear to the Bushes to have shown up unannounced at the White House.
B. Friends of George W. seemed to the Obamas to appear to Laura, and friends of Laura seemed to the Obamas to appear to George W., to have shown up unannounced at the White House.

Both of the interpretations of (1) in A and B are possible, which is represented by assuming that "movement" (of *friends of each other*) creates "copies", in bold as in (2):

(2) **Friends of each other** seemed to the Obamas **friends of each other** to appear to the Bushes **friends of each other** to have shown up unannounced.

The different interpretations in (1) can be said to correspond to the interpretation of different "copies" in (2), as in (3), where here the "copy" that is interpreted is still in bold and the others are in strikethrough. (3a) yields interpretation (1A) and (3b) that in (1B):

(3) a. ~~Friends of each other~~ seemed to the Obamas **friends of each other** to appear to the Bushes ~~friends of each other~~ to have shown up unannounced.
b. ~~Friends of each other~~ seemed to the Obamas ~~friends of each other~~ to appear to the Bushes **friends of each other** to have shown up unannounced.
c. **Friends of each other** seemed to the Obamas ~~friends of each other~~ to appear to the Bushes ~~friends of each other~~ to have shown up unannounced.

However, many questions now arise. First, it is not clear why only one "copy" can survive at PF, and furthermore why this should correspond to the representation in (3c)—we cannot have PFs like (3a-b), even when the LF takes one of these guises—where the bold "copy" is pronounced and the strikethroughs necessarily unpronounced. What about on the LF side? Is there a valid interpretation for the chain if *no* strikethrough is involved? If so, arguably the unacceptable (4a) should be possible with an LF like (4b):

(4) a. * Friends of each other seemed to themselves to have shown up unannounced.
b. $[_{IP}$ [friends of each other]$_i$ seemed to themselves$_i$ $[_{IP}$ [friends of each other$_i$]$_i$ to have shown up unannounced]]

Thus, to prevent (4), we must make the further assumption that *only one* of the "copies" can be interpreted at LF. But that isn't any more obvious than why only one of the "copies" can be pronounced at PF. One can of course stipulate all of that—but the question is *why the chain behaves that way* and not in other equally rational ways.

Although there is useful terminology that distinguishes lexical types and their tokens, from *occurrences* thereof, where technically a chain is a *set of occurrences* spanning over two or more grammatical contexts, no formalism we know of yields that as a straightforward consequence. A context-free grammar is good at capturing that type/token distinction, by way of non-terminals such as P, or equivalently a labeling mechanism as in $[_P$ *from*], shorthand for $[_{from}$ *from*], where the sub-index label denotes the type and the italicized expression presents the token to be inserted in grammatical contexts. Unfortunately, however we choose to make this precise, it doesn't help us understand what *occurrences* amount to.

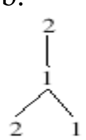
We argue that *chains are non-classical objects*, of the sort commonly assumed in physics, exhibiting conditions that are often described as "spooky". We are not the first to bring such notions into the discussion of language. For example, Paul Smolensky (Smolensky 1990; Smolensky & Legendre 2006) has argued for something along these lines for phonology and other parts of language—although within connectionist presuppositions that we do not find necessary. In other domains of cognitive science too, researchers of various orientations have suggested "spooky" connections.

We assume linguistic information, to use Randy Gallistel's felicitous phrase, is *carried forward in derivational time*, represented as digital signals of some sort (Gallistel 2006). However, those representations, in our view, are what PF and LF tokens are all about, *not the way things exist at a more elementary level.* That bedrock is, we think, syntactic. We will be taking syntax to act on some Hilbert space, by way of linear operations. Moreover, projecting syntactic stuff into interface observables is what "collapses it" into a classical reality, in which entities present reference and quantification, truth values, or for that matter the very signals of speech or writing are linearized one right after the other. That is our project in a nutshell.

None of this really makes sense without quantitative assumptions, or at least elaborate logical assumptions. Some researchers have attempted the latter, which we sympathize with. We think there may be a simpler way of proceeding, stemming from a fact that is familiar to most linguists: *we operate on feature matrices.* We would like to show next that it is easy to translate back-and-forth between familiar phrase-markers and matrices, and moreover that connatural to the latter, if their values are numerical, are very interesting quantities that turn out to be central, both, to a project attempting to construct relevant Hilbert spaces and, more generally, turn quantitative.

2. *Medeiros Matrices*

Medeiros (2012) shows how to map Lindenmayer-systems to a matrix. For instance, the Fibonacci system in (5a) with the tree in (5b) maps to the matrix in (5c) with the characteristic polynomial in (5d), whose roots (the matrix eigenvalues λ_1 and λ_2) indicate the overall distribution of the symbols with respect to one another:

(5) a. 2→ 1, 1 → 2 1 b. 

c. $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ d. $x^2 - x - 1$

$$\lambda_1 = \frac{1}{2}\left(1 + \sqrt{5}\right)$$
$$\lambda_2 = \frac{1}{2}\left(1 - \sqrt{5}\right)$$

The general method for translating between rewrite rule systems as in (5a) and matrices is as follows. For a rewrite system over n symbols, one first chooses a way to associate the n symbols to integers 1 through n. (In (5), those very symbols, *1* through *2*, are used to be the rewrite symbols.) Next one records the system as a matrix, where the number in the i^{th} row and j^{th} column counts the number of symbols of type j occurring to the right side of the rewrite rule for type i. Then we record a derivational line as a vector, with number of symbols of type i in the i^{th} place. To get the subsequent derivational line, we *multiply the vector by the matrix*; the resulting vector counts the number of symbols of each type.

The key is really to list in each row the number of 1's, 2's, or whatever symbols we are using. In (5c) above, for instance, the first row (which expands symbol *1*) contains one *one* and one *two*, while the second row (expanding symbol *2*) contains one *one* and no *two*'s. This is the general approach for any L-system of any complexity (with a finite number of rules). So for a system with rules (6i) through (6n) we create the matrix in (6) (focus on the boldfaced expressions in (6), intended to be numerical):

(6)

i. α --> a α's b β's c γ's ...
ii. β --> d α's e β's f γ's ...
iii. γ --> g α's h β's i γ's ...
⋮
n. χ --> x α's y β's z γ's

$$\begin{bmatrix} a & b & c & \cdots \\ d & e & f & \cdots \\ g & h & i & \cdots \\ \vdots & & & \\ x & v & z & \cdots \end{bmatrix}$$

Matrices can be seen to act as *linear operators on spaces*, uniformly distorting them: turning the space upside down, around, stretching it, folding it, and more—the sorts of twists and turns, for example, that proteins take as they fold into the characteristic forms that distinguish one protein from the next. If linear operators are used to model protein folding, it is not crazy to consider how such operators might be implicated in the shaping of syntactic objects, such as chains, which obey c-command and locality conditions, or even the apparently mundane distinction between what acts as a noun phrase with reference or a verb phrase without.

Why matrices? First, *because they're there!* Whereas it took a revolution in physics, coming from the world of motion and waves, to re-conceptualize matters in terms of matrices, modern linguistics basically *starts* in matrices, due to seminal work on phonological feature matrices, which Chomsky (1974) extended to syntax. Of course, one need not interpret matrices the way a mathematician does; one could think of them as mere lists of attributes with values, as in computer science. Then again, the tool exists for us to use, if we so desire. Second, because some matrices, like the ones Medeiros related to familiar trees, are not just *ipso-facto* numerical, but indeed very well behaved.

Medeiros's numbers come from a simple assumption related to the fact that a rewrite rule (the inverse of the standard merge procedure generalized to *n*-ary conditions) involves different tokens and types. Each such token can be counted, which yields a number. What is remarkable is that with such a simple assumption one can express a tree like (5b), as opposed to any other imaginable tree of arbitrary complexity. Moreover, note that Medeiros' method yields a *square matrix*, as it contains the *n* symbols in the system, and for every such symbol whether the rewrite contains *x* many of its token instances. Square matrices are particularly symmetrical and thus easy to operate with.

Note that a matrix like (5c) has a *characteristic polynomial*—as shown in (5d). All square matrices have characteristic polynomials, invariant for matrices under different *bases* (rotations of the matrix that keep their basic relations unchanged). Readers may remember 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 $P(z)_j = 0$. We say that $P(z)$ is of degree n if it has n roots, which can be thought of as its degrees of freedom. The characteristic polynomial of a matrix can be thought of as essentially its ID number. When the polynomial is turned into an equation, the solutions to that equation constitute key elements in the matrix diagonal, called *eigenvalues*—which is again what (5d) shows. Note that the eigenvalues in (5d) are ϕ, the golden ratio, and its negative inverse.

There are many important calculations one can run with eigenvalues. Consider, for instance, what is customarily called in symbolic dynamics the *topological entropy* of a dynamical system h_t (note that here we are referring to the mathematical concept of topological entropy, as opposed to topological entropy in physics; see Ott 1993). This a real number measuring the system's "complexity", in the sense that periodic orbits in a dynamical system are in one-to-one correspondence with characteristic derivational cycles in a corresponding symbolic system, so the two are topologically equivalent. For symbolic systems where a characteristic recursion is present (like the Fibonacci one in (5)), the topological entropy numerically captures the *exponential growth of that periodicity* within the system. The standard way to calculate a system's derivational entropy h_t, for a transition matrix M, is as in (7):

(7) $\boldsymbol{h_t = \log_2 \lambda_{max}}$, for λ_{max} the largest eigenvalue of M.

The largest eigenvalue of the matrix is chosen because it tracks the repeated multiplication (power) of the matrix, thus its *iterative properties* leading to its exponential growth. The logarithm is used just to get a rate. In our case, the transition matrix for our phrasal system is the Medeiros matrix in each instance—concretely, (5c) in the case we are now analyzing, pertaining to the Fibonacci tree in (5b). It should not be hard to see that this matrix's largest eigenvalue (root of the polynomial) is ϕ, so the binary logarithm of that is the system's topological entropy.

Again, the Medeiros matrices represent the *token number of symbols of a given type* in a given iteration of an L-system (*set* of derivational lines that permit the characteristic recursion in the L-system to occur). One can call such an iteration in the actual derivation a *cycle*, although in other fields the term *orbit* is used. So, from this perspective, a Medeiros matrix tells us *the essential type/token ratio for a derivational cycle/orbit*, whose topological entropy is a real number like $log_2\ \phi$. It goes without saying that we, then, need to establish what the topological entry of a derivational cycle entails for the derivation itself, clarifying issues like why it is a cycle to begin with, what sorts of objects are stable within that domain, or what to do with the domain, for instance, if the entropy becomes for some reason unstable—whether to transfer it out of the computation or remedy the derivational crisis in some operational way.

We submit that clarifying such familiar matters (the nature of phases, conditions of phase stability, phase transfer, or what impels phase components to abandon a phase) are hardly trivial or resolved matters. Moreover, it is clear to us that unearthing a tool that *is already there* can only be helpful. Researchers may of course choose to ignore the tool. But we should be thankful to Medeiros for having found it, without doing any violence to the objects we normally operate with—he simply bothered to count.

3. *The Fundamental Assumption and Anti-symmetrical Merge*

We believe the topological entropy that Medeiros's method allows us to calculate constitutes nothing short of a *quantum number* within the sort of system we are attempting to build—clarifying for us fundamental aspects of Cinque and Rizzi's (2008) cartographic program. Yet we think we need other such quantum numbers.

Consider familiar objects as in (8), from Chomsky 1974.

(8) a. noun: [+N, -V] b. verb: [-N, +V]
c. adjective: [+N, +V] d. adposition: [-N, -V]

(8) capitalizes on a semantic intuition that "nouniness" is conceptually orthogonal to "verbiness", and those two separate lexical dimensions articulate all of the conceptual space the lexicon needs. N and V features were postulated by Chomsky so as to rationalize the distribution of lexical categories. He could, of course, have called those features A and B, or *1* and *i*, and retain the system we customarily teach our students. Now, in the latter instance there would be one more level of precision: when we say that "intuitively N and V are cognitively orthogonal", we could state that orthogonality in precise mathematical terms, inasmuch as *1* is mathematically orthogonal (maximally different) from $i = \sqrt{-1}$. There are, to be sure, many other mathematical orthogonalities one could postulate, but the one between *1* vs. *i* has the added advantage that it is easy to operate with such terms arithmetically.

That being said, let's make the following Fundamental Assumption:

(9) *Fundamental Assumption:* N = *1* and $V = i = \sqrt{-1}$.

(10) a. noun: $[1, -i]$ b. verb: $[-1, i]$
c. adjective: $[1, i]$ d. adposition: $[-1, -i]$

Written as in (10), representations as in (9) may be seen as vectors. It is, moreover, convenient to translate the vectorial representations in (10) further—simply so as to operate with them more easily—to the square matrices in (11).

(11) a. noun: $\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$ b. verb: $\begin{pmatrix} -1 & 0 \\ 0 & i \end{pmatrix}$
c. adjective: $\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$ d. adposition: $\begin{pmatrix} -1 & 0 \\ 0 & -i \end{pmatrix}$

Notice that the objects in (10) are placed *in the matrix diagonal* in (11). In this format, the matrices, which we call *the Chomsky matrices*, are said to be "diagonal" and "unitary", which means they have the property of their *inverse* being identical to what is called their *adjoint*. Explaining that technically would take us too far afield, but suffice it to say that these are *extremely elegant matrices*, with mathematical properties that are well known. There are many other interesting properties that our matrices have, as they integrate into a curious mathematical group that will be discussed later on below.

To ponder the value of the Chomsky matrices, consider major syntactic dependencies:

(12) a. Nouns select PPs. b. Verbs select NPs.
c. Adjectives select PPs. d. Prepositions select NPs.

As generic statements, (12c) and (12d) are virtual universals, while (12a) and (12b) are statistically overwhelming. Of course, verbs also select other categories, which complicates a system that needs to invoke functional categories, but in science typically one starts by trying to predict the most basic interactions. In the case of language, we suppose that (12) is *cognitively prior* in that it is what learners get from Universal Grammar in the absence of experience, further complications being learnt. While the facts in (12) —together with the additional fact that, of all the major categories, it is *nouns that appear in bare guise*, without dependents, in the form of names and pronouns, etc.—are stipulated in many ways, we have never seen them explained.

To provide an explanation for (12), we start with Chomsky's matrices, plus the assumption that First Merge—the relation between a head and a complement—is *matrix multiplication*. Once one goes through the trouble of postulating the N vs. V distinction in numerical terms, it is natural to ask whether it may buy us more than formalizing the relevant orthogonality among the features. We can claim that such a multiplication is really a deformation of a given (conceptual) space by way of a linear operator, but we are sure that such an abstract reflection does not help the putative question of "Why matrix multiplication?" If readers are wondering just that, we cannot answer the question *a priori*; we can only show the results of taking such a step.

We further assume the following about merge:

(13) Merge is antisymmetrical.

This is actually presupposed in an often-cited idea of Chomsky's, namely, that Merge is literally the successor function in mathematics, represented as in (14):

(14) a. $\emptyset = 0$ b. $\{\emptyset\} = 1$ c. $\{\{\emptyset\}\} = 2$ d. $\{\{\emptyset\}\} = 3$ ... n. $\underbrace{\{\{\{\ldots\{\{\{}_{n} \emptyset \}\}\}\ldots\}\}\} = n$

In order to go from $\{\emptyset\}$ to $\{\{\emptyset\}\}$, we need to merge $\{\emptyset\}$ with itself, in the process yielding $\{\{\emptyset\},\{\emptyset\}\} = \{\{\emptyset\}\}$. Although it seems that this kind of symmetrical operation is generally disfavored in syntax, where typically atomic elements from the lexicon, compose with *complex objects* that have been previously assembled in the derivation, we can allow for this particular symmetrical situation, *if we restrict it to self-merge.* A relation that is asymmetrical except when holding with itself is called *antisymmetrical*.

Again, anti-symmetry does allow for a symmetry condition for self-merge, which turns out to be important at the point of launching a derivation. This is because when the derivation starts, there are no "complex objects assembled" in it *yet*. One way to break this inevitable symmetry is to allow self-merger (of heads), an idea first proposed by Max Guimarães (2000) and later picked up by Kayne (2009). When considered from the point of view of the Chomsky matrices and Merge as multiplication, the result of self-merging any of the Chomsky categories is surprisingly the same:

(15) $\begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix}\cdot\begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix}\cdot\begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix} = \begin{pmatrix}-1 & 0\\ 0 & i\end{pmatrix}\cdot\begin{pmatrix}-1 & 0\\ 0 & i\end{pmatrix} = \begin{pmatrix}-1 & 0\\ 0 & -i\end{pmatrix}\cdot\begin{pmatrix}-1 & 0\\ 0 & -i\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = Z$

(To multiply such matrices as in (15) we multiply the entries, in entry-wise fashion.) *Z* is one of the famous Pauli matrices, which has been put to use to predict properties of an electron's angular momentum, in quantum terms. The reasons for that are not important now, but they boil down to the fact that *Z* is what physicists call a *Hermitian matrix*.

Hermitian matrices are to matrices what real numbers are to numbers. Both can be measured. If one wishes to get philosophical about it, Hermitian stuff is what one can "pin down". We will get a feel for that as we get our hands into computations, but we can point out the obvious already: the elements in the diagonal in *Z* are both real numbers. These are key in understanding the essence of a matrix, its eigenvalues. The eigenvalues of the Chomsky matrices are combinations of ± 1 and $\pm i$. It is different for *Z*, as a result of which the matrix has other elegant properties. We can think of *Z* as a welcome encounter arising from the self-merger of the Chomsky objects. At the same time, it is also interesting to ponder what we should make of that, especially within in a semiotic system that in some sense *carries thought*, and even allows us to communicate it.

A linguistic system that is trying to start in a self-merger with the math in (15) has to resolve that "multiguity", so that instead of all possible self-mergers leading to *Z*, the system chooses one, *any* one, to the exclusion of the others. One may think of this choice as the core *Saussurean arbitrariness* in the system, as the choice of any such mapping is in principle as good as any other. Thus the *right* choice is:

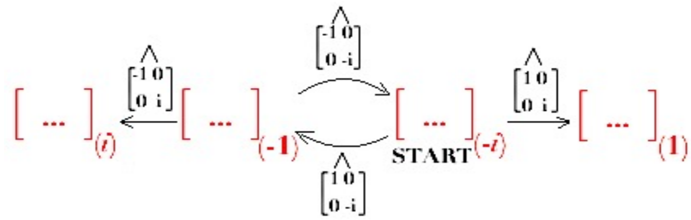
(16) N (understood as Chomsky's $\begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix}$) self-merges as $\begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix}\cdot\begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix} = Z.$

This is a *cognitive anchor* that we do not (seek to) explain. We make the choice in (17) for empirical reasons: we know derivations bottom out as nouns, the one category class that can project without dependents. Guimarães proposed self-merger *for nouns*, not surprisingly; the insight was the self-merger, not that it was for nouns.

4. *Projecting from the Bottom and Selection Restrictions*

Once the "human language anchor" in (16) is assumed, things start falling into place, in a form that can be summarized in terms of a diagram proposed to us by Michael Jarret, which we refer to as the *Jarret graph*, presented in its abstract version in (17). In this graph we need to distinguish operational ***edges*** (the Chomsky matrices, all of which are presented with a "hat" ^ to signal their operator status) and argumental ***nodes***. Both of these are matrices, since these are linear operators that can be operated on, also. But the emphasis in each instance is different: while the Chomsky objects with a "hat" are very specific, what they operate on is more open-ended: a matrix with the *determinant* signaled in parenthesis, these ranging over ±1 and ±*i*.

(17)



A matrix determinant is another invariant scalar obtained, for simple square matrices, by *multiplying the items in the diagonal and subtracting from that the product of the items in the off diagonal.* We propose that the matrix determinant determines what linguists call a "category label", which for the projections we will be operating with are *the fundamental orthogonal features ±1 and ±i only.* In effect, this is a second quantum number, together with topological entropy as discussed. We assume that the interpretation of determinants as labels are relevant for arguments only, not the operators (with a hat in the Jarret graph). The specific labeling system we argue for is:

(18) a. N projections: label/determinant -1 b. V projections: label/determinant *i*
c. A projections: label/determinant 1 d. P projections: label/determinant -*i*

The Jarret graph is basically saying that N heads select (multiply with) matrices of type -*i* (the prepositional projections) to yield -1 projections, while P heads, in turn, select matrices of type -1 (the nominal projections) to yield -*i* projections—that being the *recursive core* of the system. In addition, the graph also says that V heads select matrices of type -1 to yield *i* projections, while A heads select matrices of type -*i* to yield 1 projections—that being the *non-recursive periphery* of the first-merge system. This is all done in "categorial-grammar fashion", with the Chomsky operators being (label-less) matrices that take matrix arguments to project given results, as in (17).

In addition, the graph has a START point, explicitly signaled in (17). This is the *anchoring assumption* we have argued for. It would be silly for a graph as in (18) to start at the peripheral edges, since then the computation has nowhere to go; the core is a more useful place to start. But the core itself has *two* different sites: one labeled -*i* and the other one labeled *1* (that number being a matrix determinant). It is sound to argue, on formal grounds alone, that it is natural for the system to start at a state that *carries the computation to the very elegant Pauli matrix Z*, with determinant/label -*1*. We have already shown above how all instances of self-merge, for any of the Chomsky categories, yield this result. That being the case, the *only* matrix that carries the system to the *Z* configuration with determinant -1 is precisely $\begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$, which we call Chomsky1 (or C1), for this very reason. So it is sound for the Jarret graph to start in C1 in formal grounds, which we return to. (It is still a substantive claim to postulate that C1 corresponds *to nouns*, which we are adapting from Chomsky 1974 by way of our Fundamental Assumption; in other words, the formal system could have just as naturally started in C1 with us having assigned that matrix to verbs, prepositions or adjectives…)

The other formal properties of the Jarret graph—why the -*1* and -*i* projections are at the core, others at the periphery—*follow from the results of matrix multiplication over the*

Chomsky matrices. Specifically, only the following *eight* results are mathematically possible, via multiplication. We have mentioned how, starting on the self-merge of Chomsky's *C1* (19e), we obtain Pauli's *Z* (19a), with label/determinant -1. We can then proceed with the specific options in the Jarret graph. *Z* can multiply into -*C1* (19g) with label/determinant -*i* by -*C2* (19h) (staying at the core of the graph), *or* multiply into -*C2* (19h) with label/determinant *i* by -*C1* (19g) (going into the left periphery of the graph).

(19) a. *Z* b. *I* c. -*Z* d. -*I* e. *C1* f. *C2* g. -*C1* h. -*C2*

$$\boxed{\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}},\ \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\ \boxed{\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}},\ \boxed{\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}},\ \begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix},\ \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix},\ \begin{pmatrix}-1 & 0\\ 0 & i\end{pmatrix},\ \begin{pmatrix}-1 & 0\\ 0 & -i\end{pmatrix}.$$

The same reasoning obtains for the ensuing matrices. For example, the -*C1* obtained in the previous instance can multiply into -*Z* (19c) with label/determinant -1 by *C1* (19e) (staying at the core of the graph), or it can multiply into -*I* (19d) with label/determinant 1 by *C2* (19f) (going into the right periphery of the graph). Readers can try this as an exercise for other states in the graph, and it will become apparent that *all the results fall within the "1st Merge" Abelian group* in (19)—which is commutative for multiplication.

Note that, for each of the lexical projections, there are *two equivalent matrix variants* with the same label/determinant; we call them "twin" projections. The projections are equivalent in that *the twin matrices share the same determinant*, understood syntactically as a label. For example, *Z* and -*Z* have label/determinant *-1* because the determinant is the product of the items in the diagonal minus the product of those in the off diagonal—so *-1* in both instances. Readers can check that this is true for all other twin categories in (20).

(20) a. NP, label -1 b. AP, label 1 c. VP, label *i* d. PP, label -*i*

$$\begin{array}{cc} Z & -Z \\ \boxed{\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}}, & \boxed{\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}} \end{array}\quad \begin{array}{cc} I & -I \\ \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, & \boxed{\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}} \end{array}\quad \begin{array}{cc} C2 & -C2 \\ \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix}, & \begin{pmatrix}-1 & 0\\ 0 & -i\end{pmatrix}, \end{array}\quad \begin{array}{cc} C1 & -C1 \\ \begin{pmatrix}1 & 0\\ 0 & -i\end{pmatrix}, & \begin{pmatrix}-1 & 0\\ 0 & i\end{pmatrix}. \end{array}$$

Other matrix multiplications are possible among the eight items in (20); but *only those expressed in the Jarret graph present this kind of symmetry*. Thus, we could have multiplied, say, a preposition understood as -*C1* (19g) times a verb phrase understood as *C2* (19f); the result is -*I* (19d) with label/determinant is *1*; but that simply cannot be a projection from a prepositional head -*C2* with label -*i*. Readers can try similar multiplications off the edges of the Jarret graph, to see how only the connections made explicit within it preserve endocentricity/selection in the sense described.

That is what predicts the facts in (12), together with the formal fact that multiplication only allows certain results. Had we asked whether we could obtain a projected *Z* (20a) from the last matrix multiplication mentioned in the previous paragraph (-*C1* (19g) times *C2* (19f)), the answer would be *no*. That is not for substantive reasons as presupposed in (18); it follows from assuming a numerical base and elementary multiplications—one cannot obtain *i* from *1x1*. Thus there is an important consequence of the numerical assumptions we made to substantiate Chomsky's intuition about the cognitive orthogonality of N and V attributes, as well as his general approach to treating categories as feature matrices, together with interpreting these and their hypothesized elements in a mathematical sense: we are now able to *predict certain elementary combinations in syntax without having to invoke external considerations about other cognitive interfaces.*

4. *Topological Entropy within the Jarret graph*

We have so far presented two numbers that we think may be relevant to syntax: the topological entropy of a phrase maker (analyzed as a matrix), and category labels understood as the determinant of the Chomsky matrices, also scalars. Next, we consider how to relate such numbers and, in the process, reflect on how they may *affect the cartography of phrases.*

We first map our labels to natural numbers, e.g. as in (4) (or any other combination, this bit being arbitrary; readers may wish to try out alternative mappings as an exercise):

(21) a. [1] = 1 b. [-1] = 2 c. [*i*] = 3 d. [-*i*] = 4

Initial conditions in syntax (1st merge) exist in the possible multiplications of the four types of elements, times any of them (including themselves). In other words, we have, in terms of the mapping in (21) so as to get the Medeiros results, the following:

(22) a. 1 x 1 = 1 b. 1 x 2 = 2 c. 1 x 3 = 3 d. 1 x 4 = 4
e. 2 x 1 = 2 f. 2 x 2 = 1 g. 2 x 3 = 4 h. 2 x 4 = 3
i. 3 x 1 = 3 $\underline{\text{j. } 3 \text{ x } 2 = 4}$ k. 3 x 3 = 2 $\underline{\text{l. } 3 \text{ x } 4 = 1}$
m. 4 x 1 = 4 $\underline{\text{n. } 4 \text{ x } 2 = 3}$ o. 4 x 3 = 1 $\underline{\text{p. } 4 \text{ x } 4 = 2}$

We may now represent the information in the Jarret graph in (17) as in the following rules, each *uniquely rewriting a category with the relevant label* (matrix determinant), as corresponding to the underlined multiplications in (22):

(23) a. [1] → [*i*] [-*i*] a'. 1 → 3 4
b. [-1] → [-*i*] [-*i*] c'. 2 → 4 4
c. [*i*] →[-*i*] [-1] d'. 3 → 4 2
d. [-*i*] → [*i*] [-1] b'. 4 → 3 2

The rule system in (23) translates into the matrix in (24), per Medeiros's method. Bear in mind that the numbers in the rewrites in (23) correspond, in order, to a matrix column, just as the numbers to the left of the arrow correspond to token symbols in a matrix row. We have also provided in (24), first, the characteristic polynomial of the matrix, followed by its eigenvalues, as well as the matrix determinant, in addition to the entropy h_t. The reason this quantity is *1* here is because we are applying definition (7), and the largest eigenvalue of the matrix (root of the polynomial) is *2*, whose binary logarithm is *1*.

(24) $\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$ $x^4 - 3x^2 - 2x; \lambda_1 = 2, \lambda_2 = -1, \lambda_3 = 0, det.: 0,$ $\boldsymbol{h_t = 1}$

Now, this is a holistic statement for the entire rule system in (23), signaling the space that the Jarret graph covers. When one uses grammatical rules in given sentences, one wants to be able to think of each of them separately, to use them in concrete combinations.

Consider, next, the fact that matrix (23) is the sum (in entry-wise fashion) of four separate matrices, each of which we can represent separately as in (25):

(25) a. [1] → [*i*] [-*i*] or 1 → 3 4 $x^4, \quad \lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**

$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

b. [-1] → [-*i*] [-*i*] or 2 → 4 4 $x^4; \lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**

$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

c. [*i*] →[-*i*] [-1] or 3 → 4 2 $x^4; \lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**

$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

d. [-*i*] → [*i*] [-1] or 4 → 3 2 $x^4 \quad \lambda_1 = 0; \lambda_2 = 0, \lambda_3 = 0, \quad det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**

$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$

Thus seen, these separate matrices (which collectively aggregate to (24)) are in a sense degenerate, as their repeated zero eigenvalues show us. It is instructive to examine the topological entropy in each of these matrices. For that, bear in mind that log 0 is undefined (one cannot get zero by raising anything to the power of anything). At best, zero can only be *approached* using an infinitely large and negative power. This means that, for the separate Medeiros matrices in (25), either there is no way to define $\boldsymbol{h_t}$ or it hits some absurd infinitude. This arguably corresponds to the fact that there is *no growth* in the systems as represented in (25): a single rewrite terminates the process.

But we can also consider the rules in (25) in binary combinations, as follows (where the a-a' rules and b-b' rules are applied in combination in any given system):

(26) a. [1] → [*i*] [-*i*] or 1 → 3 4 $\quad x^4, \ \lambda_1 = 0, \lambda_2 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**
a'. [-1] → [-*i*] [-*i*] or 2 → 4 4

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

b. [*i*] →[-*i*] [-1] or 3 → 4 2 $\quad x^4\text{-}x^2; \lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 0, \lambda_4 = 0 \ ; det.: 0,$ $\boldsymbol{h_t = 0}$
b'. [-*i*] → [*i*] [-1] or 4 → 3 2

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

(27) a. [1] → [*i*] [-*i*] or 1 → 3 4 $\quad x^4, \ \lambda_1 = 0, \lambda_2 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**
a'. [-*i*] → [*i*] [-1] or 4 → 3 2

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

b. [-1] → [-*i*] [-*i*] or 2 → 4 4 $\quad x^4, \ \lambda_1 = 0, \lambda_2 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**
b'. [*i*] →[-*i*] [-1] or 3 → 4 2

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(28) a. [1] → [*i*] [-*i*] or 1 → 3 4 $\quad x^4, \ \lambda_1 = 0, \lambda_2 = 0; det.: 0,$ $\boldsymbol{h_t = -\infty}$ **??**
a'. [*i*] →[-*i*] [-1] or 3 → 4 2

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

b. [-1] → [-*i*] [-*i*] or 2 → 4 4 $\quad x^4\text{-}2x^2; \lambda_1 = -\sqrt{2}, \lambda_2 = \sqrt{2}, \lambda_3 = 0, \lambda_4 = 0 \ ; det.:0,$ $\boldsymbol{h_t = 1/2}$
b'. [-*i*] → [*i*] [-1] or 4 → 3 2

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Most of these combinations still yield undefined results. However, *two rule combinations produce sensible topological entropies*: (26b/b') and (28b/b'), repeated below:

(29) a. [i] →[-*i*] [-1] $\quad x^4\text{-}x^2; \lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 0, \lambda_4 = 0 \ ; det.: 0,$ $\boldsymbol{h_t = 0}$
b. [-*i*] → [*i*] [-1] **[i.e. VP → V NP, PP → P NP]**

(30) a. [-1] → [-*i*] [-*i*] x^4-$2x^2$; $\lambda_1 = -\sqrt{2}$, $\lambda_2 = \sqrt{2}$, $\lambda_3 = 0$, $\lambda_4 = 0$; *det.:0*, $\boldsymbol{h_t} = \frac{1}{2}$
b. [-*i*] → [*i*] [-1]   **[i.e. NP → N PP, PP → P NP]**

What could those topological entropies actually mean for our derivations?

5. *Building Cartographies based on our Quantum Numbers*

The fact that the rule combinations in (29) and (30) yield meaningful entropies ought to indicate that these particular rules *act well in tandem*. (Note: for polynomial $x^4 - x^2$, whose highest eigenvalue is *1*, the log is zero, while for polynomial $x^4 - 2x^2$, whose highest eigenvalue is $\sqrt{2}$, the binary log is ½.) This is certainly the case for the rule system in (30), which corresponds to the *iterative core* of the Jarret graph. This is what allows even first-merge to recur, into "tail recursion" sequences of the sort in (3`):

(31) … *stories from rumors about pictures of NYC*

As for rule system (29), it contains the "skeletal" rule VP → V NP, which allows us to *articulate clauses*. If we had to choose our rules intuitively, one would have precisely chosen (29) as the most important in terms of "what a sentence boils down to," and (30) as "what allows the system (of first merge) to recur." Again intuitively, one would expect the rule AP → A PP to be secondary in some sense, merely intended to add qualifiers to a noun phrase (or perhaps a verb phrase in the form of adverbs). It is interesting that, in all rule combinations in (26)/(28), instances involving that rule (namely, (26a), (27a) and (28a)) yield *undefined topological entropy*, as if such a rule didn't participate at all in the overall growth of this system. None of those seem like accidental results.

Whether those intuitions can be translated into a formal system depends on precisely how we use topological entropy, substantively. In its original conception within symbolic dynamics, this measure seeks to identify whether types of symbols exist within a system, and what that tells us about its periodicities. We are pushing the boundaries of the notion when we apply it to the Medeiros translation, which keeps track not just of types, but actually also token uses of particular symbols in given contexts—this is how we got our *2'*s, indicating the *repetition of a given token symbol*. It is not accidental that the rule combination above that has the highest topological entropy is rewriting a symbol as two tokens of a type. Note: that rule, in itself, does not yield a meaningful entropy when combining with any of the others—just when it combines *with the counterpart that closes an iterative cycle* in the first-merge system. That is what gives the system its growth, so it is well that the topological entropy should be sensitive to such a nuance.

It is probably also not accidental that the rule that results in inconsequential or no growth happens to be the one that, in matrix terms, involves the identity matrix as a result (now speaking from the point of view of the determinant/labels). We do, of course, want to be able to introduce adjectives and adverbs into the system, but it is also intuitively clear that they do not constitute a structural portion that adds structural weight in any sense: they are optional, a zero-sum gain in most instances, as a consequence of which one can also add them indefinitely without essential grammatical functions being affected. It is arguable that the undefined topological entropy may be tapping into just these matters: adding these "adjuncts" simply has no consequence for the system, so their topological entropy is not defined.

As for the skeletal rule that combines verbs and noun-phrases yielding zero entropy, this too makes sense in that *we don't want recursion in the system to be in the skeleton*, or its growth would be unwieldy. That component of the structure is surely central, and obviously not a zero-sum gain as in the adjunct instance. Then again, while adding more and more prepositions and nouns to the "tail recursion" in (31) doesn't change the fundamental character of, say, a transitive structure (i.e. embedding successively longer versions of (31) under some verb or another, we still have a simplex verb phrase), adding more verbal structure to any simple transitive situation, like *eat beans*, drastically changes the structure in familiar grammatical terms (e.g. as in *try to eat beans*).

Nothing in what we have said formally has these intuitions encoded into a system where, for example, the cartographic idea that *sentences are built around verb phrases* is captured. That, in turn, presupposes, in the broad terms of Hale & Keyser, that derivations *start* in noun phrases (directly as verb complements, indirectly as prepositional complements that end up integrating some verb phrase). Given topological entropies as in (29)/(30) (the only meaningful ones for instantiations of (25) into non-holistic rule combinations), this may be a consequence of assuming "information gravity", which recalls the idea that most of the mass in a physical atom is in its nucleus:

(32) The derivation starts at its most (topological) entropic state.

This, if true, is a substantive use of the notion of topological entropy that is not formally necessary—and a way to relate that scalar to our other scalar (the label understood as a determinant). Evidently if a derivation has to choose among the rule combinations corresponding to the ensemble in (24) (the Jarret graph in matrix form) that have a valid topological entropy, (32) forces it to pick a start point in the rule combination in (30), the one with the highest topological entropy.

That is consistent with the START point in the Jarret graph in (17), although it doesn't, in itself, remove the anchoring stipulation of associating the ideal initiation of the derivation with the self-merger of nouns. This is because what is driving the higher entropy of the system is the use of a rule system that rewrites a symbol as two identical tokens of the same type (our *2* in the matrix). Now just as we have chosen to associate that formal state of affairs with the NP, we could have associated it, equally arbitrarily, to any other substantive category. To be sure, then all familiar combinatorics would fall apart—but that is an empirical argument, not a formal one. It could be that an alien language is built around the VP in just those terms, just so as to articulate denotations around actions, with recursion of those in first merge terms vis-à-vis some applicative element, and all of that terminating around a skeletal nominal. Not human language, but a language that could use every formal nuance we have proposed.

Conceptual (stipulative) anchorings aside, (32) is interesting in its own right. First, it is a general statement, which one can falsify elsewhere in the world of derivations. That part is not blamed on an interface with the rest of cognition, which boldly entails, if true, that different creatures with a different cognition—if they abide by the same sorts of formal entities (and these map per the Medeiros method of encoding L-systems in general)—would also have to start their derivations where the Jarret graph indicates.

Second, if (32) is meant literally, the only way to make sense of the quantifier within it is in numerical terms, like those provided by h_t. Note that we couldn't have applied h_t to the Chomsky matrices. This is because those matrices have imaginary values, so they may not have a highest eigenvalues at all. But h_t works just fine for the Medeiros matrices. Just as the "label" quantum number allows us to discuss selection issues as above—sieving through the multiplications in (22) to converge in those grammatical options that the Jarret graph instantiates—so too the "entropy" quantum number allows us to prevent certain grammatical combinations. It already did in the form in (32), situating the "nucleus" of a derivation in its verb phrase, built around its nominal complement. But we can push that idea forward if we strengthen (32) with (33):

(33) The derivation ends at its least (topological) entropic state.

Of course, we need to beef (33) up by introducing functional categories, as well as the effect of movement in the system (which ought to distribute entropy through chain links). But it should be clear that doing so ought to help us in specifying familiar tendencies within the cartographic program, like "heavy" verbs existing configurationally lower than "light" verbs, etc. In a sense, the joint action of (17)/(33) should be seen as a way to seek *dynamical equilibrium in the derivation*, which ought to relate to its phases.

The latter point is at the core of topological entropy, a notion that works in any dynamical system, highlighting its (quasi-)periodicities related to its derivational dynamics. If topological entropy is the way to articulate the Cartographic Program, it is a consequence that it should, in the process, signal the system's cycles/orbits. Again, this is not a specifically linguistic requirement, which connects it to other system's presenting Dynamical Frustration, in Binder's (2008) sense: in *any* such system (quasi-)periodicities are the function of the system's topological entropy.

6. *The Explosion Problem with Specifiers and the Need for Matrix Compression*

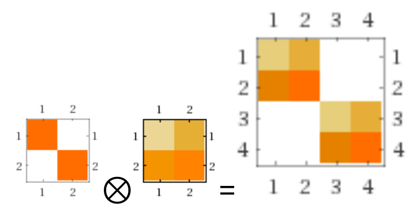
We came to these matters through an attempt to crack the chain nut in a rigorous way, which we want to sketch also. Just as we have proposed matrix multiplication for first merge, we propose another kind of product for other forms of merge beyond the initial conditions—those in which *both* seeking-to-merge items are complex, having lived a rich derivational life (instead of coming from the lexicon). *Tensor* products have the effect of *concatenating two matrices* into a larger one. This is useful in "building structure" for that very reason. Whereas regular matrix multiplications do not preserve structure (once modified, a linearly altered structure could have come from different multiplications), *tensor products are structure-preserving*: by looking at a tensor product, we know what went into it. For this reason, while matrix multiplication retains the dimensionality of its factors, tensor products generally have *a dimensionality that grows upon obtaining.*

The dimensionality of the matrix is its number of rows and columns—the information that takes to specify it. The inner dimensionalities of matrices determine what sorts of operations are allowed among them. For example, only matrices of identical inner dimensionalities can be added/subtracted, and only a matrix A with the same number of rows as the number of columns as a matrix B can enter into a matrix multiplication A B. For the objects in the Abelian group in (20), *multiplying its members times any other does not change the dimensionality of the factors:* the result is of the same dimensionality of each of the factors—otherwise (20) would not be a group. But this does *not* happen when we invoke a tensor product concatenating two matrices. The dimensionality of a matrix thus originating is *the product of the dimensionalities of the factors.*

(34) a.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 3 & 4 \end{pmatrix}$$

b.



Had (34a) been matrix multiplication involving the identity matrix *I*, the result would be identical (in dimensionality and everything else). Because this is a tensor product, even if it involves the identity matrix, the result is quite different. It preserves the *shape* of the second factor, precisely because it involves *I*. That said, it is obvious that the output is a 4x4 matrix. Moreover, by looking at the output we know that it must have originated in the product to the left, in this sense the tensor product being structure preserving.

We take the grammar to use structure-preserving tensor products to generate genuine phrase-to-phrase mergers (as opposed to "lexically more drastic" head-to-phrase conditions). This has vast consequences. To see this issue right away and more concretely, let's say that the way we generate (35a) is by the tensor product of *children's* and *pictures of NYC*. This should be possible regardless of whether the genitive is complex as in *relatives of children's*, as seen in (35b), since a phrase like that would fall into the characterization in (35a). But what about *women's children's pictures of NYC*?

(35) a. *Children's pictures of NYC*.
b. *Relatives of children's pictures of NYC*.
c. *Women's children's pictures of NYC*.
d. *London's women's children's pictures of NYC*.

In (35c) we have a specifier (*women's*) within a specifier (*children's*). So if each elsewhere merger, going beyond the initial head-complement relations, is supposed to invoke tensor products, *and* the tensor products' dimensionalities are the products of the dimensionalities of its factors, the dimensionality of *women's children's pictures…*

should be equal to that of *children's pictures* **times** that of *women's*. This can then go on into *London's women's children's pictures…* as in (35d) and so on—indefinitely. We call this the Explosion Problem, the solution of which tells us something extremely interesting about the nature of specifiers.

The general approach to such problems of dimensionality is matrix compression, based on *dimensional reduction*. What we seek for that purpose is matrix results where *entire rows or entire columns reduce to zero*, thus could be eliminated. These are matrices with zero eigenvalues, thus presenting a kind of inner symmetry. For our "Magnificent Eight" objects in (20) (the Chomsky matrices and the Pauli matrix $\pm Z$ and the identity matrix and its inverse $\pm I$), the following statement is always formally true:

(36) *The diagonal elements are the polynomial roots and matrix eigenvalues.*

Consider these formal conditions for those Magnificent Eight:

Matrices: Properties	$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$-Z = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$-I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
Char. polynomial	$x^2 - 1$	$x^2 - 1$	$x^2 - 2x + 1$	$x^2 + 2x + 1$
eigenvalues	1, -1	-1, 1	1, 1	-1, -1
determinant	-1	-1	1	1
trace	0	0	2	-2

Table 1: Algebraic properties of the Pauli matrices within the Magnificent Eight.

Matrices: Properties	$C1 = \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$	$-C1 = \begin{pmatrix} -1 & 0 \\ 0 & i \end{pmatrix}$	$C2 = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$	$-C2 = \begin{pmatrix} -1 & 0 \\ 0 & -i \end{pmatrix}$
Char. polynomial	$x^2 - (1 - i)x - i$	$x^2 - (-1 + i)x - i$	$x^2 - (1 + i)x + i$	$x^2 - (-1 - i)x + i$
eigenvalues	1, $-i$	-1, i	1, i	-1, $-i$
determinant	$-i$	$-i$	i	i
trace	$1 - i$	$-1 + i$	$1 + i$	$-1 - i$

Table 2: Algebraic properties of the Chomsky matrices within the Magnificent Eight.

The notion *trace* in these tables has nothing to do with syntactic traces: a matrix trace is just the *sum of the elements in its diagonal*, another number. We can now speak rigorously about these matrices. For example, the Pauli matrices are different from the Chomsky ones in that *all the eigenvalues of the first are real*—not all the eigenvalues of the second. We call Hermitian those matrices whose eigenvalues are real, so now we know that *Chomsky's matrices are not Hermitian*. Observe, also, putative unifications across categories. We have already observed how the positive and negative versions of the "twin" categories share the same determinant. But there are more generalizations of interest. Note that only $\pm Z$ presents the *same characteristic polynomial* $x^2 - 1$ (all other matrices in the Magnificent Eight have different characteristic polynomials). Now that is a specific sense in which $\pm Z$ is the most elegant among the Magnificent Eight: aside from being Hermitian, it has a unified characteristic polynomial, a unified trace, and a unified determinant—which no other matrix in the group does.

Considerations about characteristic polynomials, eigenvalues, and so on, obtain for all square matrices, not just our Magnificent Eight. This is important when considering these architectural issues from a broader perspective. To begin with, Pauli's $\pm Z$ is only one of three Hermitian matrices within Pauli's Group of matrices, which includes also $\pm I$ and imaginary versions of all these matrices. It turns out that when we multiply any of the non-diagonal matrices in Pauli's Group by any of the Chomsky matrices, we end up in another group of 32 matrices that we have called *the Pauli/Chomsky Group*. This group is extremely interesting, since it allows us to systematically explore a corresponding Hilbert space where specifier relations live and can be superposed into chains. Moreover, the group gives us nothing short *of a periodic table of syntactic elements*: the Magnificent Eight (lexical projections) and *corresponding functional projections* with well-behaved characteristics of the sort studied in the previous section (some will have unified polynomials, not all; some will be Hermitian, not all; some will be unitary, not all, etc.).

The research program is then clear:

(37) A. To find out how the functional categories (Infl, Comp, *v*, etc.) relate to their lexical categories and to one another in a principled fashion.

B. To determine how this group of 16 twin projections constitute the basis for standard syntax, in terms of their multiplications and tensor products.

C. To understand which of the tensor products among the categories in the periodic table lead to compressible results.

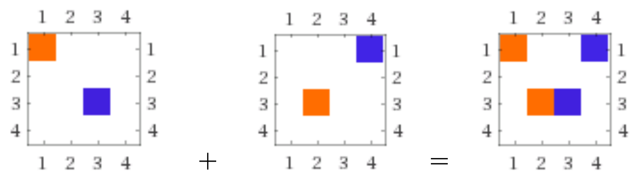
D. To figure out how the tensor products sum with one another into chain dependencies, and which among those present separable contexts.

Now, just as the formal tools above show us in what sense the Pauli/Chomsky matrices are elegant, contributing to their distribution in syntax, or to what degree they are measurable, they also allow us to understand how dimensions can be reduced after they have grown due to a tensor product. In this regard, it is useful to emphasize that *our matrices have a dimensionality equal to their non-zero eigenvalues.* So a non-compressible 4x4 matrix has four substantive eigenvalues, whereas a compressible 4x4 matrix has as many zero eigenvalues as matrix dimensions are irrelevant to it.

To cut to the chase, Hermitian matrices like $\pm Z$ have a different effect on what they operate on than the other matrices in the Pauli/Chomsky group. In a nutshell, tensor products involving $\pm Z$ are *reducible in dimension in a way that is not generalizable to the other matrices.* When externally multiplied by any of the other (non-Hermitian) matrices through a tensor product, the overwhelming majority of the matrices in the Pauli/Chomsky group yield results that are "allover the place". Basically, *only the Hermitian matrices with a unified characteristic polynomial like $\pm Z$ are well behaved.*

While tensor products involving the Hermitian matrices have the neat effects just described, in-and-of-itself that doesn't result in the required reduction of matrix dimensionality. At the same time, because of their intrinsic neatness, it is not difficult to find situations in which we can *add* matrices—as we saw already when we discussed the Medeiros matrices in (24)/(25)—of a specific sort to others that are "complementary", in the sense of allowing us to eliminate some of these eigenvalues. In (39), the specifications corresponding to the sum in (38), we clearly have a dimensional reduction, for *the ensuing matrix has two zero eigenvalues.* What took us to the dimensional reduction is the complementary neatness that went into the summed matrices, which manifests itself on the fact that the determinant and trace (in the matrix sense) of the sum is zero, as a consequence of which the characteristic polynomial is simplified to $x^4 + 2i\,x^2$. (For diagonal square matrices, the determinant amounts to the product of the eigenvalues, while the trace boils down to the sum of the eigenvalues.)

(38) a. $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -i \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -i & 0 \\ 0 & 1 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & i & -i & 0 \\ 0 & 1 & -1 & 0 \\ i & 0 & 0 & -i \end{pmatrix}$

b. 

(39) *dt.: 0, tr.: 0; char. pol.:* $x^4 + 2i\,x^2$*; eigenvalues: (-1+ i), (1 - i), 0, 0.*

7. Chains and Beyond…

The importance of the foregoing exercise is to prepare the ground for the operations that, in conditions of superposition (sums) as in (38)/(39), may lead to different chain collapses. This is the crux of the idea: *chains exist, prior to being observed, in superposed states.* At the observation point, if at all possible, they materialize, with some probability, in one of those states, which thus becomes observable.

There are very well understood properties of superposed states that, in principle, allow for their separability, for instance when they are *orthogonal* to start with (with regards to some orthonormal basis). The situation is all or nothing: if the states are orthogonal, the separation, in the right conditions, is inevitable; if they are not orthogonal, the separation is impossible. Moreover, there is no such thing as being observable in multiple states *at the same time*, much as there is meaning to the states all existing simultaneously. This is what moves us in this formalism.

With Chomsky (1995), we take a chain to be an {{α, K}, {α, Λ}} object, where a specifier α moves from context Λ to context K. Since we are modeling specifiers by tensor products, we can then take the chain to be the sum:

(40) a. $[\alpha \otimes K] + [\alpha \otimes \Lambda] = \alpha \otimes [K + \Lambda]$
b. $[K \otimes \alpha] + [\Lambda \otimes \alpha] = [K + \Lambda] \otimes \alpha$

To say α *separates* from these superpositions is to say one can “factor out specifier α” from the relevant tensor products, as in the right-hand side of the equations in (36). So the chain, in a deep sense, *links the contexts of each of its occurrences*, K and Λ. After "factoring out" the separable element α, what remains is the superposition [K + Λ].

Here is the key, now: if the superposed contexts are mutually *orthogonal*, we can apply to such *complementary* conditions the standard logic in quantum mechanics. Basically, when the relevant system is measured, it has 50% probability of being observed in the K configuration and 50% probability of being observed in the Λ configuration. If we suppose that the linguistic way of observing its abstract representations is by sending them to relevant interfaces, *within those representations* we can say that chain {{α, K}, {α, Λ}} *collapses* at either configuration K or configuration Λ, with equal probability. That is our approach to "reconstruction" effects.

Of course, we have to make precise what we mean by "orthogonal", or "maximally different" within an orthonormal basis. The following is the standard approach:

(41) Two vectors x and y in vector space V are orthogonal if their inner (scalar) product is zero.

A convenient way to define the scalar product between two matrices is as in (42), where *tr.* again represents a matrix trace—our third scalar discussed above:

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

Where, for *ket* |A>, A's conjugate adjoint $A^{\dagger}$ is the *bra* <A|.

Here we are using a vector notation introduced by Paul Dirac for notions discussed above already. What (41)/(42) boil down to is that we take two matrices A and B, understood as vectors, to be orthogonal if and only if the trace of multiplying *A's adjoint* $A^{\dagger}$ times B is zero. Because we have the Pauli/Chomsky group to work with, determining this, which in Dirac's shorthand is <A|B>, is relatively simple: we just need to churn the calculations.

The points to take home are straightforward. First, this is supposed to work with the very same types of conditions and reasoning as it does in quantum physics. The issue is not really whether the computations are wrong (they aren't), but rather whether they are meaningful. To decide on that depends on whether we have alternative theories of chain reconstruction effects and the like, and if so, whether such alternatives fare better on empirical grounds. Our attempt here is simply to show how things work in our terms.

Second, there is very little leeway for messing with the formalism. If "collapses" are meant seriously, they take place in a Hilbert space along the lines of what is guaranteed by (41)/(42) in the context of something like the Chomsky/Pauli group. In particular, if two matrices come out as orthogonal by the definitions we are introducing, they cannot be "quasi-orthogonal" or "orthogonal up to speakers' intuitions", and so on. One could, of course, change the definition of the inner/scalar product in (42), and then different things would be orthogonal. Or reject the Pauli/Chomsky group as the locus for all of this, and then perhaps in a different realm other things would be orthogonal. But in the scenario we are presenting there are *no* alternatives.

Third broad point to bear in mind: let's not loose track of the fact that we are attempting to kill several birds with the same… *chain*. At the very least we want to address the Compression Problem for specifiers. This is to say that we are not just after "reconstruction effects" for chain occurrences. While that is what has motivated the program, once we invoke matrices, groups, Hilbert spaces and so forth, one hopes that all of that doesn't amount to mere paraphernalia to address the technical problem of occurrences. For us, chain occurrences are interesting inasmuch as they touch on all these other issues, taking us from humble phrases to complex long-range correlations.

To be sure, chains are not the only long-range correlations that grammars present: there is obligatory and non-obligatory control, ellipsis of various kinds, binding and obviation effects, and much more. We have the sense that treating these matters within a Hilbert space of relations is promising, particularly when, beyond the superpositions just discussed, such system a fortiori involve entanglements. Basically, *whatever doesn't separate is entangled*, so there is plenty of room to explore what happens beyond the core situation in (37) and (38). Space/time considerations prevent us from doing so here and now, though we admit we plan to get all entangled on such a task in the immediate future.

One last point is worth emphasizing: much of what we have said above would not make (non-metaphorical) sense without the use we have made of scalars of different kinds. We have shown the role played by both topological entropy and label/determinant scalars. We have just alluded to the important role of matrix traces—another scalar—in determining the inner product of our Hilbert space. (We could also show how traces in the Pauli/Chomsky matrices help us separate substantive categories from grammatical ones.) Moreover, the logic of chain collapses as sketched ultimately depends on Heisenberg's Uncertainty Principle and the wave/particle duality that it formulates. That very logic requires a "lower boundary", usually expressed in terms of Planck's famous constant—at any rate, *a non-zero real number*. That apparatus has to be numerical, indeed real in the technical sense. No real numbers, no syntax as we have examined it. We could of course be wrong in our analyses, but if we are not, they provide *bona-fide* arguments that "mind phenomena" require real quantities as they materialize, enough at least to show up with coherent patterns as examined here.

We are not the first to have argued that the human lexicon is a Hilbert space, or that it is best to treat minimalist and other grammars as vector spaces. We have the feeling, however, that we are the first to have "gone quantum" by way of taking very seriously our linguistic fundamentals (the division into nouns, verbs, adjectives and adpositions, the role of structure, selection and endocentricity, within phrases, standard cartographies and cycles/phases, etc.). This is a sense in which our approach is as conservative as it is radical. We have shown how a Hilbert space can be constructed from assumptions that our undergraduates are exposed to. The only twist we have added is to interpret familiar conceptual orthogonalities in mathematical terms, which we have found worth studying.

References

Binder, P.-M. 2008. “Frustration in Complexity.” *Science*, Vol. 320, Issue 5874, pp. 322-323.

Chomsky, N. 1974 The Amherst Lectures, delivered at the 1974 Linguistic Institute, University of Massachusetts, Amherst: Université de Paris VII.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Cinque, G. & Rizzi, L., 2008. “The cartography of syntactic structures,” in V. Moscati, (ed.), *CISCL Working Papers on Language and Cognition*, 2, 43-59.

Gallistel, C. R. 2006. “The nature of learning and the functional architecture of the brain,” in Q. Jing & et al. (eds.), *Psychological science around the world: Proceedings of the 28th international congress of psychology*.

Guimaraes, M. 2000. “In Defense of Vacuous Projections in Bare Phrase Structure,”, in Guimaraes, M., L. Meroni, C. Rodrigues & I. San Martin (eds.), *University of Maryland Working Papers in Linguistics*, Vol. 9, pp. 90-115.

Hale, K. & J. Keyser. 2002. *Prolegomena to a Theory of Argument Structure*. Cambridge: MIT Press.

Kayne, R. 2009. “Antisymmetry and the Lexicon.” *Linguistic Variation Yearbook 2008*, pp. 1-32.

Medeiros, D.-P. 2012. *Economy of Command*. Doctoral Dissertation, University of Arizona.

Ott, E. 1993. *Chaos in Dynamical Systems*. New York: Cambridge University Press, 1993.

Smolensky, P. 1990. “Tensor product variable binding and the representation of symbolic structures in connectionist networks.” *Artificial Intelligence*, 46, 159-216.

Smolensky, P. & G Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar* (Vols. 1-2). Cambridge: MIT Press.