biolinguistics

XP

ZP

...

γ

α

XP

Human Uniqueness, Cognition by Description, and Procedural Memory

John Bolender, Burak Erdeniz & Cemil Kerimoğlu

Evidence will be reviewed suggesting a fairly direct link between the human ability to think about entities which one has never perceived — here called 'cognition by description' — and procedural memory. Cognition by description is a uniquely hominid trait which makes religion, science, and history possible. It is hypothesized that cognition by description (in the manner of Bertrand Russell's 'knowledge by description') requires variable binding, which in turn utilizes quantifier raising. Quantifier raising plausibly depends upon the computational core of language, specifically the element of it which Noam Chomsky calls 'internal Merge'. Internal Merge produces hierarchical structures by means of a memory of derivational steps, a process plausibly involving procedural memory. The hypothesis is testable, predicting that procedural memory deficits will be accompanied by impairments in cognition by description. We also discuss neural mechanisms plausibly underlying procedural memory and also, by our hypothesis, cognition by description.

Keywords: basal ganglia; cerebellum; cognition by description; knowledge by description; language evolution; procedural memory; theory of descriptions

1.   Introduction

We review evidence suggesting a fairly direct link between the human ability to think about entities which one has never perceived, what we call 'cognition by description',[1] and procedural memory. The discussion is exploratory and its conclusions are tentative, but we submit that evident links between specific forms of uniquely human cognition and procedural memory merit attention.

---

[1]   We use the term 'cognition by description' — instead of Russell's (1910) term 'knowledge by description' — because knowledge is often taken to involve justification and hence to have normative implications (Kim 1993: chap. 12). Since our aim is scientific, we choose a purely descriptive term.

2.   Cognition by Description and Merge

Non-humans give no evidence of being able to think about fictional characters or entities which they have never perceived, such as electrons or Poseidon or the children of one's unborn children. Instead, one finds a one-to-one mapping between a brain process and some property in the environment (Gallistel 1990). In spoken remarks (Chomsky 2006), Noam Chomsky has reflected on how this differs from the semantics of human mental systems:

> The concepts and words that [humans] have do not pick out entities in the world that, say, a physicist could identify, that can be found by mind-external investigation. But, rather, they are basically products of the imagination. […] The formal concept of reference that you study in logic, that's developed by Frege, Peirce, Tarski, Carnap, and so on -— it just doesn't apply to natural language. […] Human language and thought just don't have […] terms that pick out pieces of the world. This appears to be a respect in which humans differ sharply from any other organism. As far as is known, animal communication systems are based on […] an isomorphism between some internal symbol and some identifiable aspect of the world. […] Looking at the vervet monkey anthropomorphically, the way we look at everything, we say that the monkey is giving this call because there's an eagle there, and it's telling the rest of the crew of monkeys to run away. What appears to be happening [however] is that there's a reflexive reaction to some, say, motion of leaves or something like that, and then the call comes out. There's apparently an identifiable isomorphic relation between the call and a physically identifiable aspect of the environment. […] That appears to be the case for all animal communication systems, as far as anyone's discovered. If that's the case, then one fundamental difference between humans and the rest of the organic world is that we have concepts that do not pick out mind-independent entities.

How can humans use symbols, ultimately mental symbols, to indicate things in one's 'subjective universe', as it is sometimes called?

This ability should not be equated with displaced reference. A typical definition of displaced reference would be "the ability to refer to information that is spatially and temporally displaced from the location of the speaker and the listener" (Morford & Goldin-Meadow 1997: 420). But this definition doesn't distinguish referring to entities that one perceived earlier, but which are not currently present, from the designation of entities that one has *never* perceived. There is an important difference between seeing a red ball, leaving the room so that one no longer sees it, and then referring to the red ball; as opposed to referring to the children of one's unborn children. There is also the ability to talk and think about entities which do not even exist. Strictly speaking, this is not *reference* at all, but it is a sophisticated semantic ability nonetheless.

In this article we are concerned with what appears to be a uniquely human ability, namely the ability to talk and think about entities which one has never perceived and which, in some cases, do not exist. This is what we call 'cognition by description', a naturalized variant of Bertrand Russell's knowledge by description (Russell 1910, 1959). Chomsky has suggested that a computational operation could explain this uniquely human ability (Hauser *et al.* 2002, Chomsky

2005, 2006, 2007). In other words, the evolution of a recursive procedure, in conjunction with whatever mental apparatus was already there,[2] resulted in cognition by description.

Derek Bickerton, however, insists that the mere addition of recursion in hominid evolution is not enough to explain this sophisticated capacity.

> Those [semantic] properties, as [Chomsky] quite correctly states, are precisely the properties that distinguish human concepts from the concepts of other species — they refer to mental constructs rather than natural objects. But if concepts with such properties are unique to human language, *how could recursion have applied to them when language did not yet exist*? Either those concepts (and probably the words with which they were linked) already existed, implying some kind of system intermediate between animal communication and true language, or recursion could not […] have applied to anything. Since syntactic language now exists, it is a logically unavoidable conclusion that there must have been some kind of protolanguage before recursion.   (Bickerton 2005: 3, emphasis in the original)

Not surprisingly, Bickerton takes the following two issues in language evolution to be fundamentally distinct:

(1)   How did symbolic units (words or manual signs) evolve?
(2)   How did syntax evolve?

> Symbolic units and syntax are the only real novelties in human communication, and are therefore the most salient (as well as the most difficult) of the things any adequate theory of language evolution must account for. There is no reason to believe that the emergence of the two was either simultaneous or due to similar causes, and some good reasons for supposing the contrary. Chomsky (1980) has made a clear distinction between the conceptual and the computational aspects of language. The conceptual element, especially if we take it to include conceptual structure as well as the lexical instantiation of the latter, must be phylogenetically far older than any computational mechanism.   (Bickerton 2007: 511)[3]

Contra Bickerton, we wish to show how the addition of a specific recursive operation could transform a system of mental symbols which are referential in the manner of, say, vervet communication systems, into a symbol system suitable for cognition by description. It is not necessary to posit an a-grammatic protolanguage, the concepts of which made possible cognition by description, prior to the evolution of uniquely human computational abilities. In other words, Bickerton's (1) and (2) may be far more tightly intertwined than he recognizes, where the "symbols" in his (1) designate entities in one's 'subjective universe'. Contrary to Bickerton, adding the right sort of computations to a purely referen-

---

[2]   "In conjunction with whatever mental apparatus was already there" is an important qualification. In discussing this topic, people sometimes forget that it is a *necessary* condition that is at issue, not a sufficient one. This is why a condition, such as Williams syndrome (Karmiloff–Smith 1992, Smith 2004), in which mental retardation is accompanied by sophisticated syntax is not a counterexample to the hypothesis of this paper.

[3]   One can recognize the distinction between conceptual and computational systems, while acknowledging that one system uses symbols generated by the other.

tialist system can yield a mind capable of cognition by description.[4] All the really heavy lifting here, in fact, was done by Russell (1905, 1919) in his theory of descriptions and his closely related work on knowledge by description (Russell 1910, 1959), and also by Chomsky (1976) in his work on trace theory. We try to show here how Russell's insights can be extended to questions of human cognitive evolution.[5]

In doing so, we assume that the computational core of language plays a large role in uniquely human cognition generally, a simpler hypothesis than positing one recursive system for language and a separate one for belief-forming systems. This working hypothesis suggests a working methodology, namely to investigate syntactic computations as a means of understanding uniquely human concepts. It wouldn't hurt to emphasize this working methodology, and to write it out as a principle:

> So far as possible, seek explanations of uniquely human concepts in terms of syntactic computations.

A classic example of the application of this methodology would be Chomsky's attempt to understand unbounded counting in terms of syntax (Hauser *et al.* 2002, Chomsky 2005), to which we return later.

Russell's theory of descriptions is often understood as a theory about the logical form of sentences containing determiner phrases (Neale 1990).[6] Logical form is here defined as "whatever features of sentence structure (1) enter directly into semantic interpretation of sentences and (2) are strictly determined by properties of (sentence-) grammar" (Chomsky 1976: 305-306).[7]

Russell was especially concerned with the logical form of sentences containing determiner phrases which designate at most one thing; for example, such phrases as *the inventor of the telegraph*, *the author of* De Legibus, *Whistler's mother*, and *my favorite book*. This class of determiner phrases also includes phrases which designate at most one type of thing, such as *the element with atomic number 1*. These phrases are known as 'definite descriptions'. Russell (1905, 1919) was concerned to show that sentences containing definite descriptions, which we

---

[4]   This would imply that humans and non-human primates both utilize symbols, 'primitives', with referentialist semantics. This disagrees with Chomsky's (2007: 20) remark that "even the simplest words and concepts of human language and thought lack the relation to mind-independent entities that has been reported for animal communication". But it is not clear how Chomsky means to defend such a sweeping claim. Our proposal is that, while humans and other primates share a referentialist semantic system, humans alone enjoy a recursively generated semantics which draws its primitives from that referentialist system.

[5]   In doing so, we take no stand on the metaphysical issues which concerned Russell, such as sense-datum theory, logical data, etc.

[6]   An alternative would be to interpret Russell's theory as concerned with an ideal or perfect language, as opposed to natural language. Russell was not always consistent on this point, but he sometimes explicitly denied any interest in natural language: "My theory of descriptions was never intended as an analysis of the state of mind of those who utter sentences containing descriptions" (Russell 1957: 388). But if the theory of descriptions can be given a testable formulation and is shown to have explanatory power, e.g., in accounting for human uniqueness, than it deserves to be taken seriously as a scientific hypothesis about the human mind.

[7]   This is Chomsky's definition of 'LF'. How precisely Russell's own notion of logical form overlaps and contrasts with LF will not be discussed.

will call 'definites', implicitly feature bound variables. Using a simpli-fied example, the logical form of the definite *The author of* De Legibus *was Roman* would be [*There is a unique* x, *such that* x *wrote* De Legibus] *and* [*For all* x, *if* x *wrote* De Legibus *then* x *was Roman*]. The logical form more abstractly stated would be [*There is a unique* x, *such that Lx*] *and* [*for all* x *if Lx then Rx*], where *L* and *R* are predicates. The semantic interpretation could be expressed as *Whatever unique thing wrote* De Legibus, *that thing was Roman*.

According to Russell's (1910, 1959) theory of knowledge by description, the ability to have thoughts of the form [*There is a unique* x, *such that Lx*] *and* [*for all* x *if Lx then Rx*] makes it possible to think about things one has never perceived. For example, one can think about the author of *De Legibus*, namely Cicero, by forming quantified mental representations even if one had never met him. This is because the use of quantification can restrict the extension of a predicate to at most one individual thus forming a predicate which can do much of the semantic work that would otherwise be done by a proper name. Russell was, at least implicitly, following our working methodology, since he submitted that the logical form of definites also enters into knowledge by description. In other words, the operator-variable structures which figure into language also figure in the formation of thoughts about unobserved things. Analogously, we suggest that bound variables play a crucial role in cognition by description, such as one's beliefs about Cicero. This is simpler than supposing that the mind reinvents the wheel, so to speak, by generating operator-variable structures for language and then generating them separately, and all over again, for the belief systems.

To illustrate, suppose that an early hominid discovers an artifact in the forest, say, a stone tool. Let's suppose that this hominid uses the sort of mental symbol system that plausibly characterizes vervet monkeys, so that the system's semantics is limited to objects of immediate experience. The hominid thinks about the tool only by reason of "a reflexive reaction", to use Chomsky's phrase. In other words, there is a fairly direct causal link between the presence of the tool and the tokening of the relevant mental symbol. Now let us also suppose that the maker of the tool died long before its discoverer was born, and no bodily remains of that maker are anywhere to be found nearby. Given a system of mental representation limited to currently existing entities — essentially a reflex — the discoverer might be able to think about the tool but not about its maker.

But we know that that's not what actually happens, at least not for Homo sapiens. The Homo sapiens can think about the maker of the tool without needing to perceive that maker. How is this possible? Russell suggested that one uses definite descriptions. Since a definite description is implicitly a quantifier phrase, the quantifier operator restricts the extension of the relevant predicate so that it is satisfied by at most one individual or set of individuals singled out by their shared properties, the resulting expression being equivalent, for practical purposes, to a symbol for an unperceived entity or type of thing. To return to the example of the hominid encountering the stone tool; s/he could designate the long dead maker of the tool by forming a mental representation with the logical form *there is a unique* x, *such that* x *made this tool*.

Note that Russell was suggesting that one uses the product of language, namely a definite description, in mental operations that one would not intuitively

think of as linguistic. To use the terminology of modularity theory (Fodor 1983), one uses representations generated by the language faculty to form representations in the belief systems. This explanation is more economical than proposing two distinct systems for generating operator–variable structures, and thus agrees with our working methodology.[8]

From a Russellian perspective, adding variable binding to an otherwise referentialist symbol system would suffice for definite descriptions and hence knowledge by description. Therefore, *if the addition of a recursive procedure can make possible variable binding*, then one would have to disagree with Bickerton when he says that one cannot account for the evolution of cognition by description by appealing to the evolution of a recursive procedure. So can the mere addition of a recursive procedure explain variable binding? To answer the question, let's consider what sort of recursive procedure Chomsky has in mind in the first place.

In this procedure, known as *Merge*, two objects are combined such that one alone is the 'head', determining the resultant object's combinatorial properties (Chomsky 2005, 2007). The resulting compound can then be merged with another object to yield a more complex compound with a new head. And so on. This makes possible the recursive embedding of an object within an object of the same type, such as a verb phrase within a verb phrase.

For example, *for* can be merged with *mercy* to yield a PP:

(1)   [$_{PP}$ *for mercy*]

For is the head, since it determines that the phrase is a Preposition Phrase, rather than being a Noun Phrase. The grammatical category of the phrase determines how it can combine with other objects. If it were a Noun Phrase, it would appear in different grammatical contexts. Note that this result of Merge can itself be merged with *plead* to form the Verb Phrase:

(2)   [$_{VP}$ *plead* [$_{PP}$ *for mercy*]]

This VP can be merged with *to* thus forming the Infinitive Phrase:

(3)   [$_{IP}$ *to* [$_{VP}$ *plead* [$_{PP}$ *for mercy*]]]

This IP can further be merged to the verb *refuse* yielding the next VP:

(4)   [$_{VP}$ *refuse* [$_{IP}$ *to* [$_{VP}$ *plead* [$_{PP}$ *for mercy*]]]]

And so on.

---

[8]   An anonymous referee suggested that our hypothesis is committed to there being an implausible isomorphism between language and the systems of belief, specifically that language produces belief structures and that the belief systems produce syntactic structures. To the contrary, we posit a division of labor and hence a crucial difference between these two faculties. One would be crippled without the other precisely because they are not isomorphic.

We see here an illustration of recursion, in this case the inclusion of a Verb Phrase in a Verb Phrase. The result of this repeated use of Merge is a hierarchical structure in which one phrase dominates another which dominates another, in this case terminating with the domination of *for* and *mercy*.[9] There is evidence that language has hierarchical phrase structure (Miller 1962, Miller & Isard 1963, Epstein 1961a, 1961b, Fodor & Bever 1965, Johnson 1965, Graf & Torrey 1966, Mehler *et al.* 1967, Anglin & Miller 1968, Levelt 1970). Furthermore, positing a recursive operation is unavoidable in explaining how finite resources can potentially generate an infinite number of hierarchically ordered structures composed of discrete elements (Turing 1950, Boolos *et al.* 2002).

Merge can only take two forms: Either an object O is merged to an object which is a constituent of O or O is merged to an object which is not a constituent of O — internal Merge and external Merge, respectively (Chomsky 2005). The earlier example of Merge, illustrated in *refuse to plead for mercy*, was external. What about internal Merge? Consider the phrase *Socrates thought what*. Merging *what* with *Socrates thought what* to yield *what Socrates thought what* is an example of internal Merge, because *what* was already a constituent of *Socrates thought what*. The resulting four-part phrase *what Socrates thought what* would not be fully pronounced. But semantically, all four elements are interpreted, the first *what* as an operator and the second *what* as the variable it binds.[10] In other words, the semantic interpretation would be *for which thing* x, *Socrates thought* x, pronounced in English as 'What Socrates thought'.

> [A]s a simple matter of logic, there are two kinds of Merge, external and internal. External Merge takes two objects, say *eat* and *apples*, and forms the new object that corresponds to *eat apples*. Internal Merge — often called Move — is the same, except that one of the objects is internal to the other. So applying internal Merge to *John ate what*, we form the new object corresponding to *what John ate what*, … . [A]t the semantic interface, both occurrences of *what* are interpreted: The first occurrence as an operator and the second as the variable over which it ranges, so that the expression means something like *for which thing x, John ate the thing x*. At the sensorimotor side, only one of the two identical syntactic objects is pronounced, typically the structurally most salient occurrence.   (Chomsky 2007: 21)

Internal Merge generally creates operator–variable structures (Chomsky 1976).

Chomsky (2005) argues for Merge by noting that the arrangement of discrete elements into a potential infinitude of hierarchical structures requires a combinatorial operation. Given that Merge as such is a recursive operation and that internal Merge accounts for bound variables, one can begin to see how recursion could figure into 'definites' (i.e. sentences containing definite descriptions) and hence cognition by description, as per our working metho-dology.

More needs to be said as to exactly how internal Merge figures into

---

9   This process is indifferent between synthesis and analysis. Syntax merely distinguishes grammatical from ungrammatical structures, it is not specifically a sentence producing process. For some early (pre-Merge) discussion of this, see Chomsky (2002: 48).

10   By copying *what* into a superordinate position, it binds the unspoken *what* in the subordinate position, according to the c-command condition on binding (Reinhart 1976).

definites, but before doing so, we need to reflect further on the nature of definites. Here we follow Larson & Segal's (1998: 247f.) analysis of quantified sentences including definites. A quantified sentence is analyzable into three elements:

(5)   a.   A **quantification** stating how many are involved, e.g., *all*. In the case of a definite, the quantifier states that at most one is involved, e.g., *the*.

      b.   A **restriction** stating the class of entities involved, e.g., *mice*.

      c.   A **scope** stating what is true of the individuals in the restriction, e.g., *being mortal*.

In *All mice are mortal*, *all* expresses quantification, *mice* expresses the restriction, and *are mortal* expresses the scope. It will be seen that internal Merge plays a role in distinguishing restriction from scope in sentences containing determiner phrases, including definites.

Many linguists today agree with Russell that the use of definites involves operator–variable structures at the level of logical form,[11] although the Russellian approach to such structures has been amended somewhat (Keenan & Stavi 1986, Neale 1990, Larson & Segal 1998). While Russell used the unary quantifiers introduced by Frege, it is more plausible that determiners in natural languages are binary; that *all*, for example, is a relational predicate whose arguments are two sets. For example, the *all* in *All mice are mortal* expresses a relation between the set of mice and the set of mortals. But this approach to determiners is still Russellian in an important respect; namely, it still involves bound variables. Furthermore, the bound variables that figure into sentences using determiners are plausibly due to a sub-case of internal Merge known as 'quantifier raising' (see Larson & Segal 1998, Hornstein & Uriagereka 1999 for recent discussion). Given our working methodology, this means that it is a plausible hypothesis that quantifier raising plays a crucial role in cognition by description.

According to recent semantic theory (*supra*), two different sorts of Merge must occur so as to distinguish restriction from scope. Consider the logical form of *All mice are mortal* and how it is produced. *All* is externally Merged to *mice* so that *mice* serves as the restriction. But an internal Merge operation must be performed so that *being mortal* will serve as scope. Internal Merge establishes relations of scope, in the case of quantification, by producing bound variables. More specifically, *all mice* initially occurs as the complement of *mortal* and is then internally merged in a higher position (i.e. 'raised') leaving an unpronounced variable as the complement of *mortal*, namely *mortal* x. The result of this quanti-

---

[11]   Donnellan (1966, 1968) argues that there is a 'referential use' of definite descriptions such that a definite need not contain a bound variable (see also Devitt 2004). This is often taken as "an attack on Russell", but Donnellan is only saying that Russell's theory of descriptions doesn't fully generalize. Donnellan never denied that definites sometimes conceal operator-variable structures as revealed on a semantic analysis. This would be the case for the 'attributive use' of definites. So, even if Donnellan is right, it does not mean that the semantic analyses of this paper are wrong. It would just mean that they are only true of a specific use of definites. That 'attributive use' would employ the same computational procedures (e.g., internal Merge) which enter into cognition by description.

fier raising is that the set of mortal things becomes the scope of the expression. Using Neale's (1990) notation, '[all x: mice x] (mortal x)' is a good expression of the logical form of *All mice are mortal*; 'all x' must be raised to a superordinate position in order to bind the variable in 'mortal x'.[12] Let's consider an example of a definite, namely, *The maker of this arrowhead was skilled*. Ignoring tense, the logical form is more revealing notated as follows: [the x: make this arrowhead x] (skilled x). The variable appearing in '(skilled x)' is bound by the operator as a result of quantifier raising, the set of skilled things thus serving as the scope.

If internal Merge is necessary for cognition by description does it follow that internal Merge is also necessary for formulating counterfactuals or questions at the level of thought? Not necessarily. In the case of counterfactuals and questions, all that is needed is the ability to combine meaningful units in various different ways. Assume that Socrates, Thrasymachus, and kissing are objects of immediate awareness. One should be able to form the thought *Socrates kissed Thasymachus*, even if it is a false thought, simply by combining the relevant meaningful units so as to yield the representation *Socrates kissed Thasymachus*. This would not be cognition by description, but it would be counterfactual.[13] One should also be able to formulate the query *Did Socrates kiss Thrasymachus?* by forming the representation *Socrates kissed Thasymachus* and then adding the conceptual element of interrogation.[14]

As briefly noted earlier, there is a precedent for attempting to explain uniquely human cognitive abilities in syntactic terms in Chomsky's (2005) suggestion that unbounded counting results from Merge. As Noam Chomsky (p.c.) puts it, "[t]here are a number of ways of deriving the number system from Merge. To take one, assume that the lexicon has a single member, call it 1, and accept the convention that {X} = X. Then 1 = {1}. Internal Merge yields {1, {1}}. Call it 2. Etc. Addition and other operations follow pretty simply". This agrees with our working methodology.

3.   Unique to Humans?

In the past few years, there has been much discussion as to whether recursion in cognitive processes is unique to humans (Hauser *et al.* 2002, Pinker & Jackendoff 2005, Parker 2006) or whether a specific recursive procedure, such as Merge, is unique to humans (Chomsky 2005). The debate is relevant here because we want to know which is the more plausible hypothesis: Did the evolution of Merge as such usher in cognition by description, or was it specifically the evolution of internal Merge? Or maybe even just quantifier raising? If Merge-like procedures are found in other species, but without evidence of internal Merge, this would be

---

12   Raising is necessary for binding because of the c-command condition.

13   Note that Russell's knowledge by description, even though it involves knowledge of some truths, still counts as knowledge of things. For Russell (1959: 46f.), I have knowledge by description of, say, Socrates and Thrasymachus, but I do not have knowledge by description, say, that Socrates pitied Thrasymachus.

14   This may not be the same as forming the sentence *Did Socrates kiss Thrasymachus?* which evidently does require internal Merge, at least in English, with the unspoken trace of *did* following *Socrates*.

relevant. The debate concerning whether or not recursion is unique to humans, and the closely related question of whether or not hierarchically structured mental representations are unique to humans, remain very much alive (Gibson 1993, Byrne & Russon 1998, Spinozzi *et al.* 1999, Bergman *et al.* 2003, McGonigle *et al.* 2003, Fitch & Hauser 2004, Suzuki *et al.* 2006).

Here is one example of the debate: Some scientists have argued that the European starling can parse recursively center-embedded structures. Starlings can be trained to behave as though they have internalized rules of the form $a^n b^n$ as applied to the chirps and warbles they are familiar with from their own songs, at least when $n=2$ (Gentner *et al.* 2006). Does this mean that they are parsing structures of the form [*A*[*AB*]*B*] in which there is an *AB* recursively nested in another token of *AB*? In other words, does it mean that we find here the computational power minimally required for a context-free grammar as Timothy Gentner concludes? Not necessarily. According to Chomsky, the conclusion of Gentner and his colleagues "is based on an elementary mathematical error" (quoted in Goudarzi 2006). He adds that the birds' behavior "has nothing remotely to do with language; probably just with short-term memory". In other words, the starlings could be employing a non-recursive device for counting chirps and warbles. The bird could be counting two chirps, storing the result in memory, and then checking to see if the warbles also equal two.[15] This need only bestow on them the computational power of a finite-state automaton with counters (Chomsky 1959: 151), a non-recursive machine.

We do not take a position as to whether recursion is unique to humans. But we do hypothesize that (at least) internal Merge is unique to humans, and that this explains why cognition by description is only found among them. In fact, the limitation of cognition by description to humans is evidence for the limitation of internal Merge to humans. (When we say 'humans', we are not excluding other extinct hominid species; we take no stand on whether, say, Neanderthals utilized internal Merge.)

Hypothesizing that internal Merge is unique to humans leaves open the question of whether or not external Merge is as well. Fitch *et al.* (2005: 186-187) have considered the possibility that navigation in some nonhuman species employs a combinatorial computational procedure which is very much like external Merge. To give an example, an animal may be able to remember the location of its home by means of a mental representation that would be well expressed in English as [[[[*the hole*] *in the ground*] *near the tree*] *by the lake*]', exhibiting a nested structure analogous to [*refuse* [*to* [*plead* [*for mercy*]]]] and also exhibiting compositionality, an important feature of Merge. But note that internal Merge is not required to form this specific mental representation. There is a tendency for linguists working in the minimalist paradigm to treat internal and external forms of Merge as necessarily both being utilizable by a mind if either is (Berwick 1998). But, given that internal Merge requires a more developed procedural memory system than does external Merge alone, as we will discuss in the next section, it should come as no surprise that a mind may utilize the external form only.

---

[15]   See Pinker & Jackendoff (2005, fn. 10) for a similar criticism, albeit directed against Fitch & Hauser (2004).

4.   Internal Merge and Types of Memory

Internal Merge is tantamount to what is often called 'syntactic movement' or just 'movement'. This is because internal Merge does look like the rearrangement of parts, if one focuses on phonology alone. For example, in the case of internally merging *what* with *Socrates thought what*, it looks as though *what* moves from the end of the structure to its beginning, a transformation of a more basic structure. Syntactic movement respects parts of speech and phrase structure; i.e. it is 'structure-dependent' meaning that part of speech and phrasal location are crucial in determining which object is moved. In *The dog who dug there was growling*, it is possible to move *was* to the front yielding *Was the dog who dug there growling*?. But *dug* cannot be moved to the front. So **Dug the dog who there was growling?*, despite its lovely poetic meter, is ungrammatical. Not only is it the case that one can only move a verb in English question formation, it also matters which clause the verb appears in prior to movement. It is the auxiliary verb in the main clause which moves. Given poverty-of-the-stimulus evidence collected by Stromswold (1999), the structure dependency of movement seems to be innate, and hence an invariant feature of language.

To know how a sentence is divided into phrases, and the parts of speech of its elements, is to remember something about how it was constructed, i.e. how objects were merged together to form this complex object, this sentence. To know, for example, that *The dog who dug there was growling* contains a sub-clause, and where that sub-clause begins and ends, is to remember that *The dog who dug there was growling* was put together out of simpler parts and to remember something about what the parts of speech of those parts were at each step in the derivation. Also to know that *was* is here an auxiliary verb is to know something about how the parts of the sentence were put together. A mapping from one derivational step to the next, when movement is involved, "rearranges the elements of the string to which it applies, and it requires considerable information about the constituent structure of this string" (Chomsky 1956: 121). This information is tantamount to a memory of derivational steps, what is sometimes called 'derivational memory'. In other words, internal Merge requires derivational memory (Piattelli–Palmarini & Uriagereka 2005: 53f.). *If cognition by description involves quantifier raising and quantifier raising is a sub-case of internal Merge, then cognition by description requires derivational memory.*

We need to reflect on some more general features of memory before returning to the discussion of the specific memory demands of internal Merge. When the word 'memory' is used in everyday language, it is usually declarative memory that is meant; i.e. the conscious recollection of facts and events. There are also unconscious, evidently, non-declarative memory systems too (Squire 2004). The form of non-declarative memory of special interest in understanding the structure-dependency of internal Merge is procedural memory, namely the sort of memory implicated

> in the learning of the new, and the control of long-established, motor and cognitive 'skills', 'habits', and other procedures, such as typing, riding a bicycle, and skilled game playing […]. The [procedural] system underlies aspects of rule-learning […], *and is particularly important for acquiring and*

performing skills involving sequences — whether the sequences are serial or
abstract, or sensori-motor or cognitive […]. It is commonly referred to as an
implicit memory system because both the learning of the procedural
memories and the memories themselves are generally not available to
conscious access.
                                        (Ullman & Pierpont 2005: 401; emphasis added — JB, BE & CK)

How might this relate to language? Linguistic mappings of sounds onto
meanings can be divided into the idiosyncratic and the principled. For example,
refuse to plead for mercy is mapped onto its semantic content in a principled way
because its meaning is a function of the meanings of its parts and their manner of
combination. That's compositionality. The same cannot be said for *kick the bucket*
when used as an idiom. One must memorize the meaning of the latter, rather
than constructing it from its parts.[16] Michael Ullman and his colleagues
hypothesize that principled mappings utilize the procedural memory system,
while idiosyncratic mappings utilize declarative memory, what is known as the
'declarative/procedure model' (Ullman & Gopnik 1994, Pinker & Ullman 2002,
Ullman 2004, Ullman & Pierpont 2005, Newman *et al.* 2007). In terms of
Chomsky's linguistics, this would mean that Merge requires procedural memory,
whereas lexical pairings of sound and meaning utilize declarative memory.

Part of what recommends Ullman's hypothesis is its accounting for the
otherwise mysterious range of disabilities associated with Specific Language
Impairment (SLI), "a developmental disorder of language in the absence of frank
neurological damage, hearing deficits, severe environmental deprivation, or
mental retardation" (Ullman & Pierpont 2005: 399). The authors note that, in
addition to difficulties in grammar, those with SLI exhibit impairments in motor
skills, working memory, and word retrieval. This cluster of symptoms could be
explained in terms of a deficit in procedural memory. The hypothesis is further
recommended by the fact that disorders involving impairment of procedural
memory are accompanied by grammatical difficulties, while disorders involving
declarative memory are accompanied by lexical difficulties (Ullman 2004). It has
also been hypothesized that a role for the *FOXP2* gene in procedural memory
may explain why a defect in that gene results in grammatical difficulties (Ullman
& Pierpont 2005, Piattelli–Palmarini & Uriagereka 2005), a basal ganglia abnor-
mality being implicated (Watkins *et al.* 2002).

The procedural/declarative hypothesis is controversial. Some have argued
that the evidence favors a single-mechanism model (Bird *et al.* 2003, Joanisse &
Seidenberg 1999, McClelland & Patterson 2002, Longworth *et al.* 2005). Newman
et al. (2007: 436) conclude that "the issue is still open, and further evidence is
necessary to help constrain the range of possible theoretical interpretations". The
procedural/declarative model is assumed here for the sake of developing a
hypothesis to test.

The structure of a phrase involves an abstract sequencing insofar as it
exhibits hierarchical relations, as illustrated earlier by the example *refuse to plead
for mercy*. So, on Ullman's model, Merge requires procedural memory. This point

---

16   Although it is principled to the extent that one can say *kicks the bucket* or *kicked the bucket*,
      and so on.

is essentially the same as that made by Ullman & Pierpont (2005) in their discussion of 'rule governed' syntax. But note that Piattelli–Palmarini & Uriagereka (2005) point out that derivational memory, discussed above, is also plausibly procedural since it is a kind of sequence memory. Memory of derivational steps is a memory of the order in which objects were merged together and which parts of speech those objects were. *So internal Merge should place even greater demands on procedural memory than external Merge alone*. External Merge alone involves hierarchical relations, but internal Merge also requires a memory of the steps taken in forming such relations (Chomsky 2002: 37). The role of procedural memory in internal Merge means that cognition by description places a heavy demand on procedural memory.

But what of the remark one sometimes hears in linguistics that internal Merge 'comes for free'? Does this contradict the point just made? What does it mean to say that internal Merge 'comes for free'? Let's turn to some pertinent literature.

Joseph Aoun and colleagues have argued that the potential use of internal Merge in grammatical derivations is inevitable, given the presence of external Merge and given the distinction between derivations and the lexicon.[17] To quote from them:

> We believe that Copy is […] conceptually necessary, in the sense of following from a very uncontroversial design feature of Universal Grammar. It rests on the fact that there is a (virtually unanimously held) distinction between the lexicon and the computational system and that words are accessed from the lexicon. How does Copy follow from this fact? It is universally assumed that the atoms manipulated by the computational system come from the lexicon. How does the computational system access the lexicon? It does so by copying elements from the lexicon to the computational system. That accessing the lexicon involves copying is clear from the fact that the lexicon gets no smaller when it is accessed and words are obtained for manipulation by the syntax. If this is correct, then grammars that distinguish the lexicon from the computational system conceptually presuppose an operation like Copy. As virtually every approach to grammar assumes something like a distinction between lexicon and grammar, Copy is a 'virtually conceptually necessary' operation […].
>
> (Aoun *et al.* 2001; cf. Hornstein 2001: 211f.)

Given that "copies are conceptually costless" (Hornstein 2001: 22, n. 10), then the presence of external Merge gives us internal Merge for free. Why? Because internal Merge is Copy combined with (what would otherwise be) external Merge as illustrated earlier by the example of *what Socrates thought what* (Hornstein 2001).

But internal Merge comes for free only as a potential. Internal Merge may exist in the system simply as the existence of Copy and the existence of external Merge. But this alone would be a matter of competence, not execution. In other words, it does not follow that the two would be executed together. The system may not be able to execute internal Merge *in performance* until the procedural

---

[17] Unless there is a rule forbidding internal Merge, but Ockhamist considerations militate against supposing so.

system has become powerful enough to support a robust derivational memory. Hence, while Aoun *et al.* have made a case for internal Merge in competence, the conclusion of their argument is still compatible with there having been an earlier period of human existence in which external Merge was in use but without internal Merge, due simply to a less developed procedural system.

Merge creates hierarchical structures and hence plausibly relies upon procedural memory, as do all the thought processes which utilize Merge. Internal Merge takes advantage of an especially sophisticated form of procedural memory insofar as it requires memory of derivational steps. Given the importance of quantifier raising in cognition by description, we can speculate that uniquely human procedural memory plays an especially important role in cognition by description, and hence in all the cultural achievements which plausibly depend upon it: awareness of history, religion, and science.

5.   The Neuroscience of Procedural Memory

We can make some plausible conjectures about some of the brain mechanisms which underpin cognition by description by considering the neuroscience of procedural memory.

The procedural memory system consists of parallel closed loops between the cortex and the basal ganglia (the corticostriatal circuits), and between the cortex and cerebellum (the corticocerebellar circuits). The corticostriatal circuits consist of parallel and closed loops that project from the cortex to the striatum. Subsequently each circuit splits into two kinds of pathways — direct and indirect — and projects back to the same region of the cortex from which it originated, via the thalamus. The direct pathway projects from the striatum to the globus pallidus interna (GPi), and from there to the substantia nigra and from there to the thalamus. The indirect pathway in turn projects to the globus pallidus externa (GPe), then to the subthalamic nucleus and from there to GPi and then to the thalamus. The different basal ganglia-thalamocortical loops project to different areas of the cortex (e.g., primary motor cortex, premotor cortex and prefrontal cortex) and hence subserve different functions. Indeed, different channels enjoy a similar synaptic organization; this indicates that diverse functions served by the procedural memory system depend on *similar mechanisms*. Each channel is involved in those functions that are carried out by the cortical area to which it projects. The circuits projecting to the primary motor cortex or premotor cortex are involved in motor functions, whereas circuits projecting to the prefrontal cortex are involved in cognitive functions (for review, see Ullman 2004, Ullman & Pierpont 2005).

The cerebellum is also considered very important in procedural memory. The connections between the cerebellum and the cortex are also mostly parallel and functionally segregated. The projections from the cortical areas reach the pontine nuclei; from there the neurons project to the cerebellar cortex. The projections continue into the deep cerebellar nuclei especially the dentate nucleus, and from there to the thalamus and finally again to the cortical area of origin (Kelly & Strick 2003, Middleton & Strick 2000, 2001, Ramnani & Miall 2001,

Ramnani 2006). The corticocerebellar connections are also organized into parallel closed loops. The cerebellum participates in motor learning through sensing and correcting 'motor errors' — i.e. differences between intended movements and those actually performed (Ramnani 2006, Apps & Garwicz 2005). It has been suggested that the cortical regions send copies of their original commands to the cerebellum (called 'efference copies') (for a recent review, see Ramnani 2006).

An important feature of the cerebellum is the uniformity of its cellular organization and circuitry. Therefore, the cerebellum performs *the same computations for every function that it serves*; the reason for the cerebellum being involved in many different functions (including cognitive ones) lies in the different cytoarchitectonic organizations of the cortical areas from which it receives its inputs (Apps & Garwicz 2005, Ramnani 2006).

For a long time it had been supposed that the cerebellum and basal ganglia are involved solely in motor control and that they receive inputs from different areas of the cortex — including the prefrontal cortex which serves for cognitive functions — but send all of their outputs to the motor cortex. However, later findings showed that corticostriatal and corticocerebellar circuits also project to prefrontal cortex and hence may enter into cognitive functions as well (Leiner *et al.* 1993, Dezmond & Fiez 1998, Middleton & Strick 2001, 2002, Gebhart, Petersen & Thach 2002, Kelly & Strick 2003, Ramnani 2006).

In a study conducted by Schoenemann and colleagues, it was shown that while the amount of gray matter in the prefrontal cortex (the area of the cortex serving mainly for cognitive functions) does not differ much between human and nonhuman primates, prefrontal white matter differs greatly (Schoenemann *et al.* 2005). Gray matter is composed of the cell bodies of neurons, whereas white matter is composed of fibers. Moreover, a later study showed that there exists a relatively large prefrontal contribution to the corticocerebellar circuitry in humans when compared to Macaque monkeys. However, in Macaque monkeys the dominant contribution to corticocerebellar circuitry was from the cortical motor areas (Ramnani *et al.* 2006).

These findings suggest that there occurred a selective increase in inter-connectivity between the prefrontal cortex and the basal ganglia and cerebellum (i.e. the procedural memory system) in the human lineage, when compared to non-human primates. Hence it would be quite plausible to suggest that the circuits formerly mainly serving for motor functions (i.e. corticostriatal and corticocerebellar circuits, or stated otherwise 'procedural circuitry') were recruited for cognitive functions in the human lineage. This, in turn, may have played an important role in the great computational power of syntax in language and, as a consequence, to the qualitatively different computational, and ultimately conceptual, powers of the human mind.

6.   The Potential for Testing

People suffering from Broca's aphasia, generally understood to be a syntactic disorder, exhibit an interesting lack of abstract thought. In his study of Broca's aphasics, Kurt Goldstein distinguishes two attitudes, the abstract and the

concrete, observing that those with Broca's aphasia tend to be limited to the latter.

> In the concrete attitude we are given over passively and bound to the immediate experience of unique objects or situations. Our thinking and acting are determined by the immediate claims made by the particular aspect of the object or situation. For instance, we act concretely when we enter a room in darkness and push the button for light. If, however, we desist from pushing the button, reflecting that by pushing the button we might awaken someone asleep in the room, then we act abstractively. We transcend the immediately given specific aspect of sense impressions, we detach ourselves from the latter and consider the situation from a conceptual point of view and react accordingly.   (Goldstein 1948: 6)

Is this lack of abstractness a deficit in cognition by description? Technically no, since we defined cognition by description as the ability to think about entities or agents that one has never seen. We did not define it as the ability to think about states of affairs or situations which one has never perceived. Merge as such, by virtue of being productive, makes possible novel mental representations, so even external Merge, without internal Merge, could perhaps account for the ability to conceive of unperceived situations. The mere presence of a recursive procedure as such may be enough to explain abstractive thought, in Goldstein's sense, as a recursive deficit may also suffice to explain a lack thereof. The hypothesis of this paper, by contrast, predicts that defects in the procedural system will result in difficulties with grammatical transformations as well as difficulties in conceiving of *entities* (and agents) which have never been perceived. The distinction is important to bear in mind while looking for possible evidence.

Piattelli–Palmarini & Uriagereka (2005) conjecture that the uniquely hominid mutation of the *FOXP2* gene, which plausibly led to a boost in procedural memory, made possible transformational grammar (i.e. internal Merge) and hence a wide range of uniquely human cognitive abilities. Our hypothesis is compatible with theirs, although not identical to it. For one thing, we do not put so much weight on *FOXP2*.[18] Perhaps *FOXP2* alone accounts for uniquely hominid, or even uniquely human, procedural memory, but we are also open to roles for other genes as well (Özçelik *et al.* 2008, Tan *et al.* 2008, and references). Furthermore, Piattelli–Palmarini & Uriagereka do not discuss the relevance of Russell to these questions of human evolution.

Our hypothesis is testable. Aphasias have already been discussed. Clear evidence of an aphasia which disables internal Merge, or even just quantifier raising in particular, while leaving cognition by description unimpaired, would refute our hypothesis. 'Clear evidence', however, is an important qualification, because there might be a condition in which internal Merge remains intact but

---

[18]   As Piattelli–Palmarini & Uriagereka (2005: 60) write:

> As it turns out, it is not crucial that specifically *FOXP2* be involved in our hypothesis, but since this is the only gene we actually know for sure to be implicated in the language system, *largely for concreteness* we will articulate the proposal around it, and in particular the putative 'permissive' role of *FOXP2* in procedural memory.

cannot be applied in communication. A general inability to think recursively, or even just a general inability to exhibit the computational abilities required for transformational grammar, accompanied by unimpaired cognition by description would offer a clearer refutation.

An interesting potential field of research is to investigate the relation between basal ganglia impairment and deficits in cognition by description. There is some correlation between advanced stages of schizophrenia, basal ganglia dysfunction, and dementia. Since it is late developing, the demented condition is sometimes called 'tardive dementia' (Breggin 1990) or 'tardive dysmentia' (Wilson *et al.* 1983). The condition may be due to schizophrenia being partly a basal ganglia disorder (Graybiel 1997), or it may be a result of anti-schizophrenia medication damaging the basal ganglia (Breggin 1990, 1993, Dalgalarrondo & Gattaz 1994), or both. Either way, we submit that it is worthwhile to look for difficulties in the transformational aspects of grammar and impairment in cognition by description in individuals with subcortical dementia.

7.   Conclusion

Our conclusion is that Bickerton is mistaken in insisting that uniquely human conceptual structure, specifically cognition by description, must have evolved prior to the evolution of syntax. One can see how the emergence of recursion could have suddenly made possible cognition by description along with syntax. This is consistent with the hypothesis that Merge ushered in both syntax and uniquely human semantics, an hypothesis which Chomsky favors presumably because of its simplicity. However, we also leave open the possibility that external Merge appeared first, meaning that there was a semi-syntax prior to the evolution of full-blown syntax. Specifically, this would have been a phrase-structure grammar without transformations. It would also have been a 'protolanguage' in some sense, but not the a-grammatic sort of protolanguage which Bickerton posited in the earlier quotes. Full-blown syntax, because it utilizes internal Merge, could not have been utilized until a fully developed procedural memory system was in place. So it is possible that the evolution of the memory systems placed a constraint on the evolution of syntax, and hence uniquely human semantics as well, including cognition by description.

Our discussion has been extremely speculative and exploratory, as we noted at the outset. The evidence adduced could, no doubt, be interpreted in other ways. But our aim has been to arrive at a possible explanation of uniquely human semantics, an explanation which can be tested and will as a result of testing, almost certainly, be replaced by something better in time. Our rationale for proposing something so tentative is that one must speculate in order to have something to test. One cannot rule out hypotheses without having hypotheses in the first place. We agree wholeheartedly with Bickerton (2005: 2), when he writes that "Speculation is the horse that drags the chariot of theory".

References

Anglin Jeremy M. & George A. Miller. 1968. The role of phrase structure in the recall of meaningful verbal material. *Psychonomic Science* 10(10), 343-344.

Aoun, Joseph, Lina Choueiri & Norbert Hornstein. 2001. Resumption, movement, and derivational economy. *Linguistic Inquiry* 32, 371-403.

Apps, Richard & Martin Garwicz. 2005. Anatomical and physiological foundations of cerebellar information processing. *Nature Reviews Neuroscience* 6, 297-311.

Bergman, Thore J., Jacinta C. Beehner, Dorothy L. Cheney & Robert M. Seyfarth. 2003. Hierarchical classification by rank and kinship in baboons. *Science* 302, 1234-1236.

Berwick, Robert C. 1998. Language evolution and the minimalist program: The origins of syntax. In James R. Hurford, Michael Studdert–Kennedy & Chris Knight (eds.), *Approaches to the evolution of language: Social and cognitive bases*, 320-340. Cambridge: Cambridge University Press.

Bickerton, Derek. 2005. Chomsky: Between a Stony Brook and a hard place. Ms. [http://radicalanthropologygroup.org/pub_bickerton_on_chomsky.pdf (18 September 2008).]

Bickerton, Derek. 2007. Language evolution: A brief guide for linguists. *Lingua* 117, 510-526.

Bird, Helen, Matthew A. Lambdon Ralph, Mark S. Seidenberg, James L. McClelland & Karalyn Patterson. 2003. Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language* 48, 502-526.

Boolos, George S., John P. Burgess & Richard C. Jeffrey. 2002. *Computability and Logic*, 4th edn. Cambridge: Cambridge University Press.

Breggin, Peter R. 1990. Brain damage, dementia, and persistent cognitive dysfunction associated with neuroleptic drugs: Evidence, etiology, implications. *Journal of Mind and Behavior* 11, 425-464.

Breggin, Peter R. 1993. Parallels between neuroleptic effects and lethargic encephalitis: The production of dyskinesias and cognitive disorders. *Brain and Cognition* 23, 8-27.

Byrne, Richard W. & Anne E. Russon. 1998. Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences* 21, 667-721.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2, 113-124.

Chomsky, Noam. 1959. On certain formal properties of grammars. *Information and Control* 2, 137-167.

Chomsky, Noam. 1976. Conditions on rules of grammar. *Linguistic Analysis* 2(4), 303-351.

Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Blackwell.

Chomsky, Noam. 2002. *Syntactic Structures*. Berlin: Mouton de Gruyter. [First published by The Hague: Mouton, 1957.]

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1), 1-22.

Chomsky, Noam. 2006. Biolinguistic explorations: Design, development,

evolution. Paper presented at the *Department of Psychology's Distinguished Lecture Series and the Human Development Program Colloquium Series*, Boston University (January 2006). [http://www.buworldofideas.org/shows/2006/06/20060604.aspS (18 September 2008).]

Chomsky, Noam. 2007. Of minds and language. *Biolinguistics* 1, 9-27.

Dalgalarrondo, Paulo. & Wagner F. Gattaz. 1994. Basal ganglia abnormalities in tardive dyskinesia: Possible relationship with duration of neuroleptic treatment. *European Archives of Psychiatry and Clinical Neuroscience* 244, 272-277.

Devitt, Michael. 2004. The case for referential descriptions. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and Beyond,* 280-305. Oxford: Oxford University Press.

Donnellan, Keith. 1966. Reference and definite descriptions. *Philosophical Review* 75, 281-304.

Donnellan, Keith. 1968. Putting humpty dumpty together again. *Philosophical Review* 77, 203-215.

Epstein, William. 1961a. The influence of syntactical structure on learning. *American Journal of Psychology* 74, 80-85.

Epstein, William. 1961b. A further study of the influence of syntactical structure on learning. *American Journal of Psychology* 75, 121-126.

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in nonhuman primates. *Science* 303, 377-380.

Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the narrow language faculty: Clarifications and implications. *Cognition* 97, 179-210.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, Jerry A. & Thomas G. Bever. 1965. The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behaviour* 4, 414-420.

Gallistel, C. Randy. 1990. Representations in animal cognition: An introduction. *Cognition* 37, 1-22.

Gebhart, Andrea L., Steven E. Petersen & W. Thomas Thach. 2002. Role of the posterolateral cerebellum in language. *Annals of the New York Academy of Sciences* 978, 318-333.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440(7088), 1204-1207.

Gibson, Kathleen R. 1993. Tool use, language and social behavior in relationship to information processing capacities. In Kathleen R. Gibson & Tim Ingold (eds.), *Tools, Language and Cognition in Human Evolution*, 251-269. Cambridge: Cambridge University Press.

Goldstein, Kurt. 1948. *Language and Language Disturbances: Aphasic Symptom Complexes and Their Significance for Medicine and Theory of Language*. New York: Grune & Statton.

Goudarzi, Sara. 2006. Chomsky challenge: Bird said to grasp key aspect of language. *LiveScience.* [http://www.livescience.com/animals/060426_starling_linguist.html (18 September 2008).]

Graf, Richard & Jane W. Torrey. 1966. Perception of phrase structure in written

language. *Proceedings of the Annual Convention of the American Psychological Association*, 83-84.

Graybiel, Ann M. 1997. The basal ganglia and cognitive pattern generators. *Schizophrenia Bulletin* 23, 459-469.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569-1579.

Hornstein, Norbert. 2001. *Move! A Minimalist Theory of Construal*. Malden, MA & Oxford: Blackwell.

Hornstein, Norbert & Juan Uriagereka. 1999. Labels and projections: A note on the syntax of quantifiers. *University of Maryland Working Papers in Linguistics* 8, 249-270. [Reprinted in Uriagereka, Juan. 2002. *Derivations: Exploring the Dynamics of Syntax* (Routledge Leading Linguists), 115-135. London: Routledge.]

Joanisse, Marc F. & Mark S. Seidenberg. 1999. Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America* 96, 7592-7597.

Johnson, Neal F. 1965. The psychological reality of phrase structure rules. *Journal of Verbal Learning and Verbal Behaviour* 4, 469-475.

Karmiloff–Smith, Annette. 1992. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Keenan, Edward L. & Jonathan Stavi. 1986. A semantic characterization of natural language determiners. *Linguistics and Philosophy* 9, 253-326.

Kelly Roberta M. & Peter L. Strick. 2003. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *The Journal of Neuroscience* 23(23), 8432-8444.

Kim, Jaegwon. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Larson, Richard & Gabriel Segal. 1998. *Knowledge of Meaning: An Introduction to Semantic Theory*. New Delhi: Prentice-Hall of India. [First published by Cambridge, MA: MIT Press, 1995.]

Leiner, Henrietta C., Alan L. Leiner & Robert S. Dow. 1993. Cognitive and language functions of the human cerebellum. *Trends in Neurosciences* 16(11), 444-447.

Levelt, Willem J.M. 1970. Hierarchical chunking in sentence processing. *Perception and Psychophysics* 8, 99-103.

Longworth, C.E., S.E. Keenan, R.A. Barker, W.D. Marslen–Wilson & L.K. Tyler. 2005. The basal ganglia and rule-governed language use: Evidence from vascular and degenerative conditions. *Brain* 128, 584-596.

McClelland, James L. & Karalyn Patterson. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences* 6, 465-472.

McGonigle, Brendan, Margaret Chalmers & Anthony Dickinson. 2003. Concurrent disjoint and reciprocal classification by Cebus Apella in seriation tasks: Evidence for hierarchical organization. *Animal Cognition* 6, 185-189.

Mehler, Jacques, Thomas G. Bever & Peter Carey. 1967. What we look at when we read. *Perception and Psychophysics* 2(5), 213-218.

Middleton Frank A. & Peter L. Strick. 2000. Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Research Reviews* 31(2-3), 236-250.

Middleton Frank A. & Peter L. Strick. 2001. Cerebellar projections to the prefrontal cortex of the primate. *The Journal of Neuroscience* 21(2), 700-712.

Middleton Frank A. & Peter L. Strick. 2002. Basal-ganglia 'projections' to the prefrontal cortex of the primate. *Cerebral Cortex* 12(9), 926-935.

Miller, George A. 1962. Some psychological studies of grammar. *American Psychologist* 17, 748-762.

Miller, George A. S & Stephen Isard. 1963. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behaviour* 2, 217-228.

Morford, Jill P. & Susan Goldin–Meadow. 1997. From here and now to there and then: The development of displaced reference in homesign and English. *Child Development* 68(3), 420-435.

Neale, Stephen. 1990. *Descriptions.* Cambridge, MA: MIT Press.

Newman Aaron J., Michael T. Ullman, Roumyana Pancheva, Diane L. Waligura & Helen J. Neville. 2007. An ERP study of regular and irregular English past tense inflection. *NeuroImage* 34, 435-445.

Özçelik, Tayfun, Nurten Akarsu, Elif Uz, Şafak Çağlayan, Süleyman Gülsüner, Onur Emre Onat, Meliha Tan & Üner Tan. 2008. Mutations in the very low-density lipoprotein receptor *VLDLR* cause cerebellar hypoplasia and quadrupedal locomotion in humans. *Proceedings of the National Academy of Sciences of the United States of America* 105, 4232-4236.

Parker, Anna R. 2006. *Evolution as a Constraint on Theories of Syntax: The Case against Minimalism*. Edinburgh: University of Edinburgh dissertation.

Piattelli–Palmarini, Massimo & Juan Uriagereka. 2005. The evolution of the narrow faculty of language: The skeptical view and a reasonable conjecture. *Lingue e Linguaggio* IV, 27-79.

Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's so special about it? *Cognition* 95(2), 201-236.

Pinker, Steven & Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6(11), 456-463.

Ramnani, Narender. 2006. The primate cortico-cerebellar system: Anatomy and function. *Nature Reviews Neuroscience* 7(7), 511-522.

Ramnani, Narender, Timothy E.J. Behrens, Heidi Johansen–Berg, Marlene C. Richter, Mark A. Pinsk, Jesper L.R. Andersson, Peter Rudebeck, Olga Ciccarelli, Wolfgang Richter, Alan J. Thompson, Charles G. Gross, Matthew D. Robson, Sabine Kastner & Paul M. Matthews. 2006. The evolution of prefrontal inputs to the cortico-pontine system: Diffusion imaging evidence from Macaque monkeys and humans. *Cerebral Cortex* 16(6), 811-818.

Ramnani Narender & Chris Miall. 2001. Expanding cerebellar horizons. *Trends in Cognitive Sciences* 5(4), 135-136.

Reinhart, Tanya. 1976. *The Syntactic Domain of Anaphora.* Cambridge, MA: MIT dissertation.

Russell, Bertrand. 1905. On denoting. *Mind* 14, 479-493.

Russell, Bertrand. 1910. Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society* 11, 108-128.

Russell, Bertrand. 1919. *Introduction to Mathematical Philosophy*. London: Allen and

Unwin.

Russell, Bertrand. 1957. Mr. Strawson on referring. *Mind* 66, 385-389.

Russell, Bertrand. 1959. *The Problems of Philosophy*. Oxford: Oxford University Press.

Schoenemann P. Thomas, Michael J. Sheehan & L. Daniel Glotzer. 2005. Prefrontal white matter volume is disproportionately larger in humans than in other primates. *Nature Neuroscience* 8(2), 242-252.

Smith, Neil. 2004. *Chomsky: Ideas and Ideals*, 2nd edn. Cambridge: Cambridge University Press.

Spinozzi, Giovanna, Francesco Natale, Jonas Langer & Karen E. Brakke. 1999. Spontaneous class grouping behavior by bonobos (Pan Paniscus) and common chimpanzees (P. Troglodytes). *Animal Cognition* 2, 157-170.

Squire, Larry R. 2004. Memory systems of the brain: A brief history and current Perspective. *Neurobiology of Learning and Memory* 82, 171-177.

Stromswold, Karin. 1999. Cognitive and neural aspects of language acquisition. In Ernest Lepore & Zenon Pylyshyn (eds.) *What is cognitive science?*, 356-400). Oxford: Blackwell.

Suzuki, Ryuji, John R. Buck & Peter L. Tyack. 2006. Information entropy of humpback whale song. *The Journal of the Acoustical Society of America* 119(3), 1849-1866.

Tan, Üner, Sibel Karaca, Meliha Tan, Bekir Yılmaz, Namık Kemal Bağcı, Ayhan Özkur, & Sadrettin Pence. 2008. Unertan syndrome: A case series demonstrating human devolution. *International Journal of Neuroscience* 118, 1-25.

Thach, W. Thomas & Amy J. Bastian. 2004. Role of the cerebellum in the control and adaptation of gait in health and disease. *Progress in Brain Research* 143, 353-366.

Turing, Alan M. 1950. Computing machinery and intelligence. *Mind* 59, 433-460.

Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92, 231-270.

Ullman, Michael T. & Myrna Gopnik. 1994. Inflectional morphology in a family with inherited specific language impairment. *Applied Psycholinguistics* 20, 51-117.

Ullman, Michael T. & Elizabeth I. Pierpont. 2005. Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex* 41(3), 399-433.

Watkins, Kate E., Faraneh Vargha–Khadem, John Ashburner, Richard E. Passingham, Alan Connelly, Karl J. Friston, Richard S.J. Frackowiak, Mortimer Mishkin & David G. Gadian. 2002. MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain* 125, 465-478.

Wilson, Ian C., James C. Garbutt, Clayton F. Lanier, Joseph Moylan, William Nelson & Arthur J. Prange. 1983. Is there a tardive dysmentia? *Schizophrenia Bulletin* 9, 187-192.

John Bolender
  Middle East Technical University
Department of Philosophy
Eskişehir Yolu
06531 Ankara
Turkey
bolender@metu.edu.tr

Burak Erdeniz
University
Informatics Institute
İnönü Bulvarı
06531 Ankara
Turkey
burakerdeniz@yahoo.com

Cemil Kerimoğlu
Georg-August-Universität Göttingen
International Max Planck Research School
Neuroscience Program, Grisebachstr. 5
37077 Göttingen
Germany
gencemil@gmail.com

Optimal Growth in Phrase Structure

David P. Medeiros

This article claims that some familiar properties of phrase structure reflect laws of form. It is shown that optimal sequencing of recursive Merge operations so as to dynamically minimize c-command and containment relations in unlabeled branching forms leads to structural correlates of projection. Thus, a tendency for syntactic structures to pattern according to the X-bar schema (or other shapes exhibiting endocentricity and maximality of 'non-head daughters') is plausibly an emergent epiphenomenon of efficient computation. The specifier-head-complement configuration of X-bar theory is shown to be intimately connected to the Fibonacci sequence, suggesting connections with similar mathematical properties in optimal arboration and optimal packing elsewhere in nature.

Keywords: c-command; minimalism; phyllotaxis; projection; X-bar theory

1.   Introduction

This article addresses some theoretical issues in language design, adopting the biolinguistic concerns of the Minimalist Program (Chomsky 1995b). Within this framework, the line of inquiry pursued here is the attempt to explain linguistic properties in terms of 'laws of form' that may have nothing in particular to do with language, or even with biology, but rather seem to be at work at the deepest level in nature. Much has been written elsewhere clarifying and defending this sort of approach; see Chomsky (2005), Freidin & Vergnaud (2001), Uriagereka (1998), and Boeckx & Piattelli–Palmarini (2005), among others.

Within the Minimalist Program, much attention has been given to 'virtual conceptual necessity', and the intuition that 'that which is necessary is also sufficient'. As a result, one prominent trend in minimalist explanation is to reduce linguistic properties to requirements for 'legibility' with respect to the cognitive systems with which the linguistic system interacts (so-called 'bare output conditions'). Nevertheless, it deserves to be emphasized that various linguistic

properties can be both real and subject to minimalist explanation without being required in order for language to work at all, or in the simplest possible way. One of the most important lessons from applying 'Galilean' thinking to the natural world is that sometimes the best (i.e. natural) solution is not the simplest. Often more than one constraint must be satisfied by a system in some optimal way, and when the constraints conflict, interestingly complicated structure may emerge. In language as well, the biolinguistic viewpoint leads us to expect to find certain properties that are more complicated than would be strictly required for language to work *at all*, but are nevertheless 'natural' if language works *optimally*.

1.1.   Where We Are Headed

I propose in this article that certain properties of phrase structure have this kind of explanation, following not from bare output conditions but rather emerging 'for free' from concerns of efficient computation.[1] In particular, I propose here that the characteristic shape of phrases, as captured by the X-bar schema and similar forms, constitutes what we might think of as an 'optimal packing solution' or an 'optimal growth mode'. On the barest assumptions, Merge may apply freely to recursively build structure from terminal elements in any number of ways. However, if this implicitly free structure-building is subject to a constraint on efficient computation (related to minimizing computation involving c-command and containment relations), then some constructional choices will be preferred over others. Given basic concerns of locality of information flow in the derivation, it is plausible that this will induce certain consistent patterns in recursion (what amount to repeated structural 'templates'). Enumerating all possible recursive templates and comparing them with respect to this computational constraint, I show that the best templates have the shape of generalized X-bar projections. That is, the best way to 'pack' terminals into an iterated molecule of recursive structure (the best phrasal template) places a unique terminal at the bottom of the phrasal template, with 'slots' for several more objects of the same shape as the full 'phrase'. This kind of format is represented in (1).

(1) $[\,\alpha\,[\,\beta \ldots [\,\gamma\,[\,X^0\,\delta\,]\,]\ldots]\,]$

   As I will show, such a pattern of recursion produces fewer c-command and containment relations than any pattern of comparable complexity, a fact that I take to indicate computational optimality (e.g., minimizing the space searched by repeated probe-goal operations). In (1), X^0 is a terminal element, and α, β, γ, and so on are themselves constructed according to the pattern in (1). This is really shorthand for a class of optimal patterns, differing among themselves in how many self-similar 'slots' (α, β, γ, …) they permit. This includes (2), (3), and (4): In familiar terms, (2) corresponds to the geometry of the head-complement pattern, (3) to the specifier-head-complement pattern of the X-bar schema, and (4) to a

---

[1]   As will be familiar to connoisseurs of this enterprise, the idea is that explanation for linguistic properties can fruitfully be pursued in terms of the abstract derivation that generates expressions. The relevant sense of efficiency is to be understood as internal to this abstract computation, rather than directly reflecting online processes in language use.

pattern in which every 'phrase' may have two 'specifiers'.

(2) $[\ X^0\ \ \alpha\ ]$

(3) $[\ \alpha\ [\ X^0\ \ \beta\ ]]$

(4) $[\ \alpha\ [\ \beta\ [\ X^0\ \ \gamma\ ]]]$

We may describe the family of growth patterns fitting (1) as 'projective'. Some further factor(s) must act to select one particular choice (e.g., (3) instead of (4)) from the spectrum of projective solutions described by (1), a matter to which I return.[2] On the other hand, the format of (5) is not projective in the appropriate sense (because the terminal element X^0 is not at the 'bottom').

(5) $[\ X^0\ [\ \alpha\ \ \beta\ ]]$

I believe this result is surprising and significant. The options for structure-building allowed here are quite free; any finitely-defined scheme incorporating terminals into indefinitely recursive patterns is considered. Needless to say, only a small minority of these patterns 'look like' projections. Other possibilities have a repeating phrasal template which places terminals at (potentially many) designated locations other than the 'bottom', or recurse via units different than the 'top' of the template, and so on.[3] The considerations which enter into the investigation are of a purely configurational, geometric nature; no notion of 'head of a phrase', 'label', or other elements of the theory of projection are built into the assumptions. Yet something akin to projection (more precisely, a structural basis which could readily be mapped to a projection scheme) emerges 'for free' as an optimal solution. This suggests that the property of projection may be an epiphenomenon of 'blind' structural optimization.

A final point worth mentioning here is that what is explained is an optimal tendency, not an absolute law. As is the case with laws of form more generally, this kind of explanation is actually strengthened by finding occasional deviations from the predicted pattern (so long as they are rare). Consider, for example, the pervasive Fibonacci pattern in plant growth. A certain species may display this pattern as an overwhelming tendency, but individuals may show other patterns (or, as often happens, a deviation from the pattern is found on one portion of a single individual otherwise adhering to the pattern). In such cases, we are led to suspect even more strongly that the Fibonacci pattern is a result of a quite general law of form, rather than directly a result of some strict requirement. So too for the property of projection in language, I would like to suggest. That is, certain

---

[2]   To preview: There is arguably a cost associated with making the growth pattern too complicated, such that a growth pattern like $[\ \alpha\ [\ \beta\ [\ X^0\ \ \gamma\ ]]]$ places a heavier burden on resources than does a format like $[\ \alpha\ [\ X^0\ \ \beta\ ]]$. But the more complicated the pattern is allowed to be, the greater the reduction in c-command and containment totals. Thus, we expect language to settle on some 'minimax' compromise between the greater optimality of a more complicated growth rule, and the inherent costs of such further complication.

[3]   Of course, this invites the further question of whether those options are 'linguistically reasonable', or are ruled out for other reasons. I address this matter in section 5.

analyses propose that individual structures are not 'well-behaved' with respect to projection: small clauses, for example (see Moro 2000); see also the various proposals concerning exocentric or multi-headed structures (e.g., Williams 1994, Bouchard 1995, and Jackendoff 1977). If such analyses are on the right track, then it would seem misguided in principle to try to explain why projection is 'virtually conceptually necessary'.

1.2.   *Assumptions and Perspective*

The results presented in this article are primarily mathematical in nature. This departs from the usual practice in linguistics of close and careful attention to the intricacies of natural language data, where a proposal is judged by its success in covering new empirical paradigms, or in reinterpreting recalcitrant patterns in more illuminating ways. The perspective taken here is highly abstract, several steps removed from detailed empirical descriptions and from highly ramified empirical predictions. Instead, the goal is to attempt to explore one kind of explanation for some broad empirical generalizations that seem more or less well-established. It will not be my purpose here to defend these empirical descriptions, nor to refine them or extend their coverage to new kinds of data. The predictions of this study, insofar as they can be construed as empirical at all, would be definitively falsified by a discovery that linguistic structures overwhelmingly tended toward some characteristic recursive shape other than a projective one, or had no such characteristic shape at all.[4]

I will assume that syntax consists of a computational system utilizing recursive Merge, which may apply both to items drawn from the lexicon and to the output of other Merge operations. I keep to the simpler case of External Merge throughout the article, setting aside the complications that arise in treating Internal Merge. I furthermore assume that Merge is subject to the Extension Condition, and limited to strict binarity.

An anonymous reviewer points out that binary branching may be one of the facts of language most in need of explanation in terms of efficient computation.[5] Accounts of "why language is that way" with respect to binarity

---

[4]   The matter is muddied by the observation that *any* binary branching structure can be decomposed into *some* combination of different 'projective forms' in the present sense, if all that matters is bare geometry. Nevertheless, the claims advanced here are not the merest triviality: The idea is that some particular projective structure is applied more or less consistently.

[5]   The same reviewer wonders whether the approach pursued in this contribution may shed some explanatory light on the matter. As explained below, under strict binary branching, c-command and containment totals are exactly equal. As treated here, this is simply a convenient accident, allowing both measures to be lumped together in a single measurement. The reviewer suggests that some principle of grammar may favor this sort of balance, or that perhaps this fact tells us something about which of the two relations is more important in language design. The second point seems promising at first: Completely flat structure minimizes containment relations absolutely, while maximizing c-command relations (though doing no worse than worst-case binary branching). Does this suggest that binarity is favored for c-command? Closer examination is not encouraging. For example, [[a b c] [d e f]], with a mix of binary and ternary branching, actually results in lower totals of *both* c-command and containment relations (20 and 14, respectively) than any strictly-binary arrangement of the same elements (22 of each, at best).

exist — for example, Kayne's (1984) notion of unambiguous path, his theory of antisymmetry (Kayne 1994), or the general notion that "what is necessary is also sufficient". Although the matter is in no way trivial, I simply adopt the usual assumptions in this regard, trusting that readers will find it at least familiar.

I also do not attempt to deal with the possibility that adjuncts may lie 'on another plane', as has sometimes been suggested, thus ruling out some interesting possibilities. Thus, the present approach can be seen as aligning with Kayne (1994) and Cinque (1999) in assuming that adjuncts are in fact specifiers with unexceptional geometry. If that assumption should prove incorrect, and adjuncts have some special status in terms of their branching geometry, then this study is leaving out another important case over and above Internal Merge.

1.3.   *A Preview of Comparing Recursive Patterns*

As a first pass at the considerations to be explored here, suppose that a syntactic derivation has reached a stage where the following three objects remain to be combined:

(6) X^0, AP, BP

Let us take X^0 to be a bare lexical item, while AP and BP are internally complex objects constructed by Merge. For the purposes of this simplified example, let us ignore any distinction between AP and BP. The options for continuing the derivation are the following:

(7)   [AP [X^0 BP]]   (or [BP [X^0 AP]])
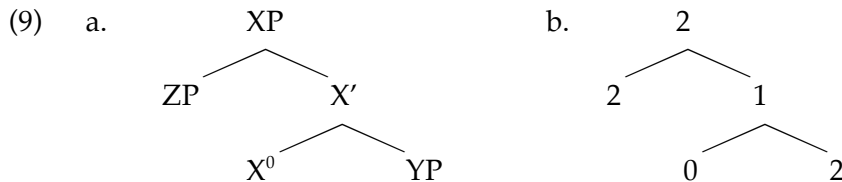
(8)   [X^0 [AP BP]]

Is there any basis for choosing between (7) and (8) in terms of their effects on c-command and containment relations? There is. Let a be the number of nodes in AP, and let b be the number of nodes in BP. Since AP and BP are internally complex, a, $b > 2$. When two objects Merge, the number of new c-command relations defined is simply the sum of the number of nodes in each; likewise, the operation also creates the same number of new containment relations (as the new mother node contains all of the nodes in each). Thus, creating (7) defines $(b + 1) + (a + b + 2) = a + 2b + 3$ new c-command and containment relations. Creating (8), on the other hand, allows $(a + b) + (a + b + 2) = 2a + 2b + 2$ new c-command and containment relations, which is strictly greater. Thus, fewer such relations are (potentially) computed at this stage if the derivation 'grows' according to (7) rather than (8). As argued in more detail below, this gives us good reason for preferring (7) over (8) in terms of efficient computation, all else equal.

Needless to say, this departs from the usual way of thinking about these matters. For one thing, it is usually assumed that given some real example, only one of (7) or (8) could apply; the other choice would 'crash', failing to meet the requirements of the items involved. Moreover, only some of the c-command and containment relations defined would actually be exploited to carry real linguistic

relations. I return to these issues in more detail later on. For now, the idea is that if we find as an empirical matter that the configuration in (7) tends to predominate as a structural pattern, while configurations matching (8) are relatively rare, we might be able to explain that fact in terms of this kind of comparison.
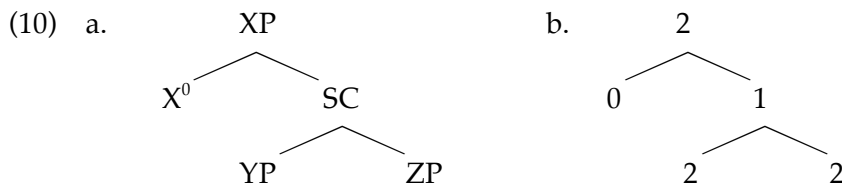
Note that (7) has the shape of an X-bar pattern of specifier, head, and complement, whereas (8) might correspond to a head taking a small clause complement, which seems to be a good deal less common (as an iterated pattern). What is at stake here has nothing to do with projection; questions such as whether X^0 is the 'head' of the construction do not enter into selecting one form over the other. Rather, the issue is one of branching form and its effects on c-command and containment relations.

In this light, consider the familiar X-bar schema in (9a). Setting aside the matter of projection (the fact that the complete syntactic object shares a lexical category label X with its head X^0), the relevant aspect for our purposes is that a complex syntactic object is formed by the particular recursive pattern in (9b).

(9)   a.

```
        XP                          b.        2
       /  \                                  / \
     ZP    X'                               2   1
          /  \                                 / \
        X⁰    YP                              0   2
```

At first, it looks like (9b) is just a matter of 'bar-level' notation: 0, 1, and 2 correspond to X^0, X', and XP respectively. But there is a way of thinking about (9b) which does not require reference to explicit 'bar-level' features (a grammatical device that has been discarded from minimalist theory for good reasons). The objects in (9b) are merely a convenient notation for describing the particular recursive pattern embodied by the X-bar schema. That is, a 0 in (9b) is a terminal (a lexical atom), while 1 and 2 are defined recursively: A 1 is an object resulting from Merging a 0 and a 2, and a 2 is the result of Merging a 1 and a 2. This is a template for recursion, implicitly expandable 'all the way down'.

On the other hand, the option followed in (8) manifests a phrasal format distinct from the X-bar shape, as in (10). (10a) gives a familiar linguistic interpretation of the shape (a head taking a small clause complement, as in the analysis of the copula by Moro 2000). What is of interest for present purposes is the abstract recursive characterization of the shape in (10b).

(10) a.

```
        XP                          b.        2
       /  \                                  / \
     X⁰    SC                               0   1
          /  \                                 / \
        YP    ZP                              2   2
```

Lest this be misunderstood, let me hasten to point out that I am not claiming by the representation in (10b) that small clauses are X' categories, or anything of the sort. Instead, the point is that this structure can be characterized in terms of three kinds of geometric object. One is a terminal, X^0, labeled 0 in

(10b). The other two objects (1 and 2) are distinguished by their recursive proper-
ties. The idea is that (10b) is an alternative to (9b) as a phrasal template. If this
pattern continued, the nodes labeled 2 at the lowest level of (10b) would them-
selves be head+small clause structures of the same shape as (10b), potentially 'all
the way down'. This would lead to different possible branching forms for linguis-
tic structure.

I illustrate in (11) and (12) the results of recursively expanding the X-bar
schema (9b) and the head+small clause pattern (10b). Expressions characterized
by these patterns would fill some finite portion of these full branching spaces.

(11)

(12)
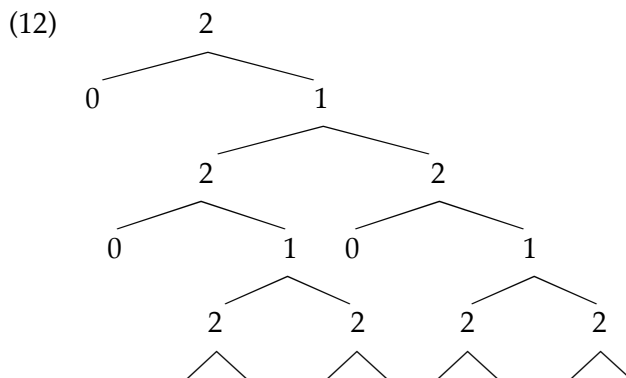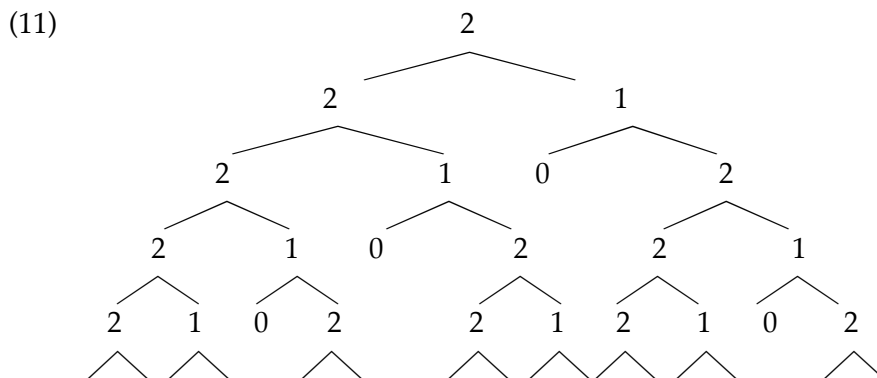
As is immediately clear, recursive expansion of the X-bar pattern creates a
space of branching forms which is intuitively 'denser' than the space associated
with the head+small clause pattern. This difference in 'branching density' turns
out to be simply another aspect of the difference between (9b) and (10b),
ultimately a part of the same fact underlying the local preference for (7) over (8).
Put simply, the more densely the space of forms generated by a phrasal template
branches, the better that phrasal template is for reducing the computational
burden of c-command and containment relations. The relationship between
recursive patterns (such as the X-bar format (9b) and the head+small clause
format (10b)) and c-command and containment relations is the matter that will
concern us in this article.

1.4.   On 'Explaining' Projection

What is 'projection', exactly? This question was obscured by the notational

conventions of earlier theories, wherein the notion was almost trivial. Taking trees as real objects, 'projection' has to do with how non-terminal nodes are labeled; whichever daughter shares its categorial 'label' with the mother node has projected.

Within a minimalist theory such as Chomsky's (1995a) *Bare Phrase Structure*, this familiar notion suddenly becomes problematic. Chomsky proposes a set-theoretic interpretation of linguistic structure building. On that conception, it is no longer so straightforward to 'label' non-terminal 'nodes'. A device is stipulated to capture labels, but it seems somewhat *ad hoc*; Merge of α and β is taken to yield not the simplest object $\{\alpha, \beta\}$, but rather $\{K, \{\alpha, \beta\}\}$, K the label; this requires further complication in introducing the notion of 'Term', essentially so that syntactic operations 'skip over' the label as a potential syntactic object in its own right. Collins (2002) objects to this complication, pointing out that it goes "way beyond" what a minimalist theory of phrase structure requires.

More recent work seems largely to agree with Collins; 'labels' are now taken to be implicitly defined, with Merge keeping to the simpler, 'bare' output of $\{\alpha, \beta\}$. Chomsky (2005) proposes that labels are identified by a search algorithm, and more recently has suggested that structures going beyond the head-complement format are 'unstable' in some sense (cf. Moro 2000 for a similar idea), and must be resolved by movement (Noam Chomsky, p.c.). Nevertheless, the idea of a 'label' perseveres, now motivated as a computational device carrying all information about a syntactic object relevant to further computation. Hornstein & Nunes (2008), following suggestions of Chametzky (2000) and Uriagereka (1998), challenge even this idea, arguing that for adjuncts at least, labels are unnecessary; the contribution of an adjunct to interpretation is understood via default 'conjunction' (here following Pietroski 2004).

Casting the matter in terms of interface interpretation in this way, we may well ask, with Wolfram Hinzen, whether forcing syntactic structure to reflect the relevant notions is really the right move:

> As for the notion 'head', why should phrase structure capture it, if the question of which of two lexical items that are merged becomes the head is decided by the *lexical* properties of these heads?   (Hinzen 2006: 182)

Where does this leave us? It is hard to deny that there is something substantive to the notion of projection; a verb phrase, say, is different from a noun phrase, and this difference can be traced to the differences between verbs and nouns. But it is precisely the 'therapeutic' value of minimalism that it leads us to demand more than empirical justification for the postulation of various devices; the goal is not merely to discover what language is like, but to explain "why it is that way". Regardless of the descriptive value of projection, or even its 'usefulness' for interpretation or syntax-internal computation, there remains the problem of mechanisms. That is, what structural device or process actually underlies the phenomenon that surfaces as projection, and where does that come from? If the mechanism can be explained 'naturalistically' rather than teleologically (i.e., as emerging 'for free' rather than being motivated by its eventual function), then we are closer to the goal of truly 'Galilean' explanation of language.

1.5.   *Organization of the Article*

This contribution attempts to cover some unfamiliar ground, exploring an unusual avenue of linguistic explanation at a highly abstract level. To avoid losing the way, it may be helpful to map out in advance where we are headed.

Section 2 examines one example of the kind of recursive pattern predicted by this account that is of particular interest: the specifier–head–complement configuration of X-bar theory. Here, I show that this pattern is fundamentally connected to the Fibonacci sequence. I include some speculation on the significance of this fact, and how it may relate to similar properties elsewhere in nature. Section 3 lays out the claim that c-command and containment relations are of central importance to certain aspects of linguistic computation, and that minimizing such relations results in more efficient computation. I briefly review several familiar empirical domains in which such concerns plausibly apply, and attempt to justify simply counting all such relations as an idealized measure of the relevant computational cost.
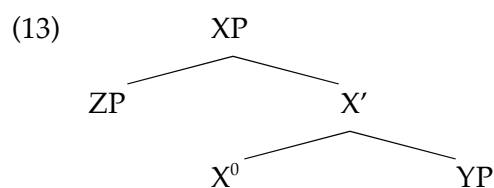
Section 4 tackles the problem of specifying what derivational patterns are available in principle to a Merge-based system. I develop a method to compare different patterns to each other in terms of c-command and containment relations, and map out how the various possibilities fare. The basic technique will be to compare different growth patterns to each other on the basis of the 'best trees' they can generate for a given number of terminal elements. Growth patterns will be partitioned into comparison sets on the basis of their complexity, and it will be shown that the best growth pattern from each comparison set is a member of the class of 'projective' patterns, with structural properties corresponding to endocentricity and 'non-head' maximality.

In section 5, I attempt to outline how the present study fits into the context of other current work, and where appropriate indicate why I have chosen to pursue an orthogonal line of inquiry. Section 6 concludes the article, drawing together the various threads and reviewing what has been established, and where it seems to point. Finally, I include an appendix presenting the formal results underpinning the claims made in section 4.

2.   **X-Bar Structure and the Fibonacci Sequence**

In this section, I show that X-bar configurations are related in a fundamental way to the Fibonacci sequence.[6] Following Uriagereka's (1998) identification of Fibonacci patterns in syllable shapes and theme-rheme structure, this is of some biolinguistic interest in itself. The mathematical structure at issue is the specifier-head-complement configuration of X-bar theory in (13):

---

[6]   The Fibonacci numbers form the sequence 1, 1, 2, 3, 5, 8, 13, 21, 34,… defined recursively by $a_0=1$, $a_1=1$, and $a_n=a_{n-1}+a_{n-2}$. Named for Leonardo da Pisa (ca. 1200, also known as Fibonacci), the numbers were known long before to Indian thinkers. These numbers, and the related golden section, seem to be favored in the natural world in myriad ways, very few of which will be discussed here.

(13)

```
              XP
          ╱       ╲
        ZP          X′
                 ╱      ╲
               X⁰         YP
```

(13)

$$
\begin{array}{c}
\text{XP} \\
\diagup \quad \diagdown \\
\text{ZP} \qquad \text{X}' \\
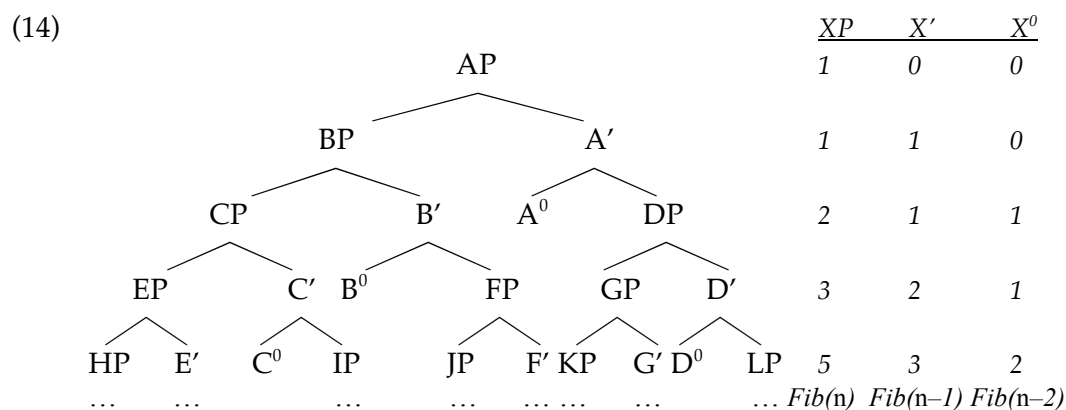\diagup \quad \diagdown \\
\text{X}^0 \qquad \text{YP}
\end{array}
$$

The object in (13) has played a central role in the empirical description of linguistic forms. The literature of X-bar theory is enormous; for some important developments, see Chomsky (1970), Jackendoff (1977), Stowell (1981), Kornai & Pullum (1990), Speas (1990), Kayne (1994), and Chametzky (1996). X-bar theory has been adopted widely even outside the Principles–and–Parameters tradition stemming from Chomsky (1981); see, for example, Bresnan (1982), Gazdar *et al.* (1985), and Pollard & Sag (1987, 1994). For now, suffice it to note that many researchers have taken (13) to be an important generalization about linguistic structure. Assuming so, we would like to know why phrases seem to pattern according to (13), if indeed they do; are there other possibilities? If so, why is (13) favored? As Hinzen puts it:

> What exactly does the X-bar scheme explain? And can its strictures be explained as following from more general and fundamental principles in the workings of the computational system? Or must we take it as an ultimate syntactic template that follows from nothing at all, accepting notions like headedness and projection as primitives? (Hinzen 2006: 180)

2.1. Iterated X-Bar: Fibonacci Numbers of Category Types

As noticed first by Carnie & Medeiros (2005), recursive expansion of the X-bar schema generates a Fibonacci sequence of bar-level categories at successive levels of embedding. Let us take the X-bar schema as recursively defining an X-bar space, and imagine 'filling' this space, such that all possible specifiers and complements are realized, each with their own specifiers and complements, 'all the way down'. If the X-bar schema is iteratively expanded in this way, the number of XPs, X's, and X^0s at successive levels of depth in the structure each form a Fibonacci sequence. This can be seen in the partially expanded structure in (14).

(14)

Structure	XP	X′	X^0
AP	1	0	0
BP — A′	1	1	0
CP — B′ A⁰ — DP	2	1	1
EP — C′ B⁰ — FP GP — D′	3	2	1
HP E′ C⁰ IP JP F′ KP G′ D⁰ LP	5	3	2
… … … … … … … …	*Fib(*n*)*	*Fib(*n–1*)*	*Fib(*n–2*)*

Recall that Fib(n) is defined recursively by $a_0=1$, $a_1=1$, and $a_n=a_{n-1}+a_{n-2}$.[7] In the X-bar schema, each XP at depth n introduces another XP at depth $n+1$ (its specifier), and another at depth $n+2$ (its complement). Thus, the number of XPs at depth n is the sum of the number of XPs at depth $n–1$ and $n–2$. There is a single XP at level 0 (the root node), and one at depth 1 (its specifier). Thus, letting XP(n) represent the number of XPs possible at depth n, XP(n) = Fib(n). Each X' at depth n is introduced by an XP at depth $n–1$, so the number of X's at depth n, or X'(n), is Fib($n–1$). Finally, each X^0 is introduced by an XP at depth $n–2$, so $X^0(n)$ (the number of X^0s at depth n) is Fib($n–2$). As a further consequence, the sum of number of objects of all types at each level of depth (i.e. XP(n) + X'(n) + $X^0(n)$) is a double of a Fibonacci number (2*Fib(n)) everywhere except at the root.

Figure 1 below provides a more perspicuous way to visualize how the Fibonacci sequence arises in the fractal space of forms generated by the X-bar pattern. Here, linear order is mapped to the counter-clockwise direction around the circle, starting at the top/'north' (assuming specifier–head–complement order). The binary Merge at the root of the tree corresponds to a division of the circle exactly in half; further binary branching deeper in the tree divides the relevant portion of the circle in half again. Where terminals occur in the expanded X-bar schema, the relevant portion of the circle is blacked out (no further subdivision will occur there). For example, in an X-bar tree the first terminal down from the root (the head of the root XP) occupies the left half of the right branch. Thus, the quarter circle between south and east is colored in. The next head down from the root is the head of the specifier phrase, corresponding to the shading of the eighth of the circle between the southwest and west directions. This process continues indefinitely; the result is a fractal diagram with Fibonacci numbers of successively smaller fractions blacked out, illustrating how the space of possible binary-branching forms may be 'populated' by terminals under perfect (infinite) iteration of the X-bar pattern of recursion.
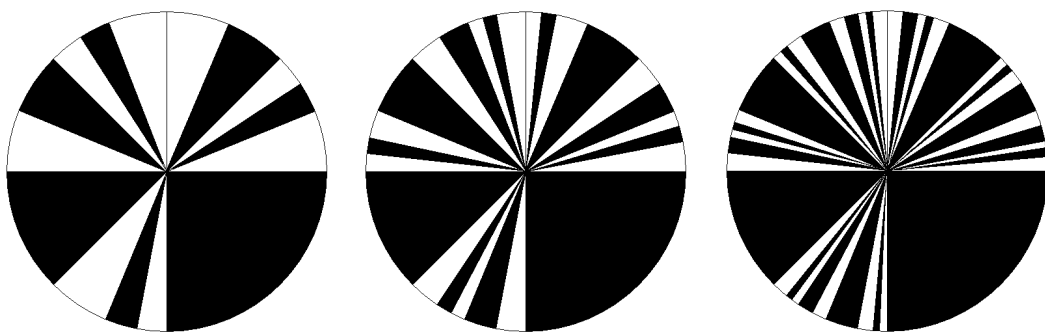

Figure 1.   Three steps in the recursive expansion of X-bar space

---

[7]   Frequently, the Fibonacci sequence is defined with $a_0=0$, $a_1=1$. It should be clear that the choice of index at issue is arbitrary, and irrelevant to the point being made here.

2.2.   *Fibonacci String Lengths and X-Bar Analyses*

There is another sense in which X-bar structure is related to the Fibonacci sequence. This fact is related to the question of what X-bar analyses can be assigned to a linguistic string of a given length. By an X-bar analysis, I mean an assignment of bracketing such every phrase contains a head, and up to two other phrases in the usual configuration of specifier and complement. That is, expanding the X-bar schema top-down, each phrase XP may have any of the following shapes, but no other possibilities are allowed:

(15)   a. $XP = X^0$
    b. $XP = [\ X^0\ YP\ ]$
    c. $XP = [\ ZP\ [\ X^0\ YP\ ]]$

Put another way, the X-bar scheme is taken to be a 'ceiling', but not a 'floor', on the internal complexity of a phrase. Assuming so, a number of different X-bar analyses are available for any string. Let us call the 'depth' of an analysis the maximum level of embedding of any element in the tree it assigns to the string. For example, a string of length 1 must have depth 0 (i.e., it is a trivial tree consisting of a single node), string length 2 requires depth 1, and string length 3 requires depth 2. For greater string lengths, some analyses will have different depths than others. As a function of the string length, we can identify the maximum depth of any possible analysis (clearly, 1 less than the string length), and also the minimum possible depth.
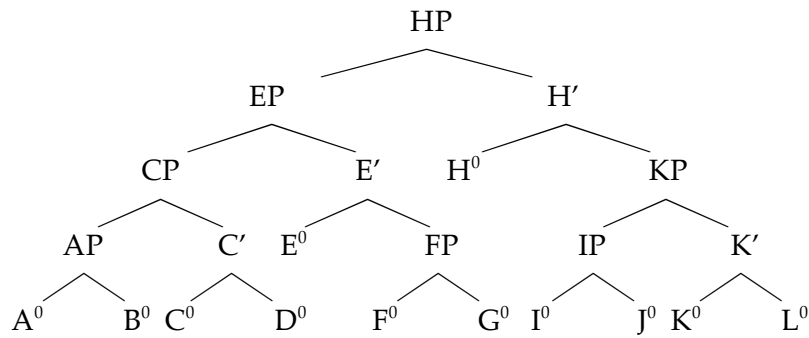
Fibonacci string lengths are *minimal depth milestones*, in the sense that a string of length Fib(n) is the first string length with a greater minimal depth than the previous string length. That is, a string of length 4 has a minimum depth of 2, the same as the minimal depth of string length 3; 5 is the first string length which forces an analysis of depth 3; likewise, string length 8 is the first with a minimal depth of 4, and so on.[8] Of course, real strings may have deeper analyses than the minimum. The point is simply that Fibonacci numbers have significance in terms of best-possible analyses, since minimal depth analyses are the 'best trees' within X-bar for a given number of elements, in terms of minimizing c-command and containment relations (see section 4 and the Appendix for discussion). As an illustration, consider (16) below:

---

[8]   This follows directly from the observation in the section 2.1 and this well-known identity:

(i) $\sum_{i=0}^{n} Fib(i) + 1 = Fib(n+2)$ (This identity is easily proven by induction.)
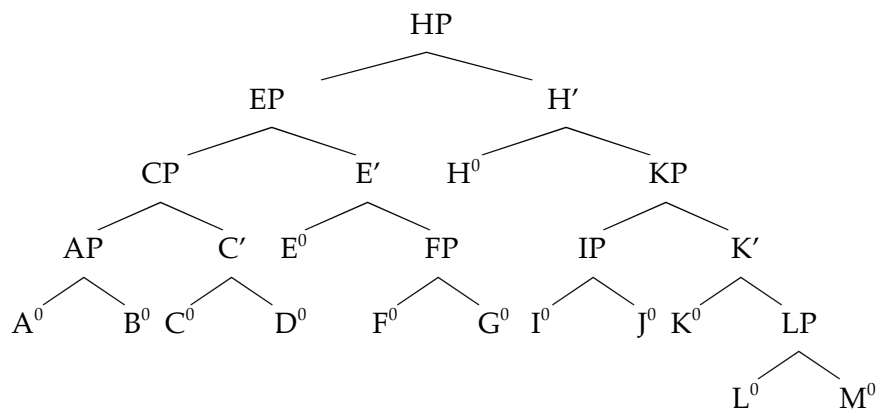
To see why, consider the X-bar analysis that packs the longest string possible into an X-bar analysis of a given depth. Given (23), in this analysis, all of the categories (including XP and X') at the greatest depth n are formatives in the surface string; thus, Fib(n) XPs + Fib(n–1) X's + Fib(n–2) X^0s, plus all of the X^0s introduced at lesser depths: Fib(n–3) + Fib(n–4)… + Fib(0). Adding one more terminal to the string forces the tree to depth n+1.

(16)

```
                                    HP
                        ┌───────────┴───────────┐
                       EP                        H′
                  ┌─────┴─────┐              ┌────┴────┐
                 CP           E′            H⁰         KP
             ┌────┴────┐    ┌──┴──┐              ┌──────┴──────┐
            AP        C′   E⁰    FP            IP             K′
          ┌──┴──┐   ┌──┴──┐   ┌──┴──┐       ┌──┴──┐        ┌──┴──┐
         A⁰    B⁰  C⁰    D⁰  F⁰    G⁰  I⁰    J⁰  K⁰        L⁰
```

The representation in (16) contains as many terminal nodes as possible for a depth 4 tree (*viz.* 12). The next string length, 13, is a Fibonacci number, and it is the first string length which forces the X-bar analysis to a minimal depth of 5. That is, to add another terminal element to (16) while adhering to the restrictions imposed by the X-bar format (understood as in (15)), one of the nodes at the bottom-most layer of the tree must be expanded, bringing the depth of the tree to 5 for the first time. (17) is an example of such a 'milestone' tree; no rearrangement of this number of elements into a structure consistent with the X-bar pattern has less depth.

(17)

```
                                    HP
                        ┌───────────┴───────────┐
                       EP                        H′
                  ┌─────┴─────┐              ┌────┴────┐
                 CP           E′            H⁰         KP
             ┌────┴────┐    ┌──┴──┐              ┌──────┴──────┐
            AP        C′   E⁰    FP            IP             K′
          ┌──┴──┐   ┌──┴──┐   ┌──┴──┐       ┌──┴──┐        ┌──┴──┐
         A⁰    B⁰  C⁰    D⁰  F⁰    G⁰  I⁰    J⁰  K⁰        LP
                                                        ┌──┴──┐
                                                       L⁰    M⁰
```

2.3. *Are the Fibonacci Properties of X-Bar Significant?*

It is tempting to see the appearance of the Fibonacci sequence in the X-bar pattern as being deeply significant in itself. But the X-bar schema is after all a very simple mathematical object, and there may be nothing particularly magical about the appearance of the Fibonacci sequence in the structures it generates. Their appearance in this domain could be no more of a surprise than their appearance in the family trees of bees, or in Fibonacci's idealized rabbit populations, or in the number of metrical possibilities for a line of Sanskrit poetry, or any of the myriad situations these numbers describe. To put it another way, it could be that these properties are an accident of no 'real' significance, or worse, merely a reflection of mathematical simplicity in linguists' description of language, rather than a property of language itself.

Yet it is undeniable that patterns related to the Fibonacci sequence play an important role in nature, especially in optimal packing and optimal arboration.

For example, botanical elements emerging from a central growth point tend to spontaneously organize into Fibonacci numbers of spirals, winding in opposite directions. In that case, it is known that the pattern is indeed the 'best possible' (dynamic) solution. Likewise, the pattern shows up in the branching patterns of many plants (e.g., sneezewort), and in a different sense in the proportions governing asymmetric branching in mammalian bronchial structure. The list goes on; see Uriagereka (1998) for discussion and further examples, including other Fibonacci patterns in linguistic structure. It seems that the pattern plays a 'spooky' role in nature (particularly in situations related to optimal self-similar growth). Thus, finding such a pattern in phrase structure suggests that this may be another manifestation of 'laws of form', reinforcing the biolinguistic suspicion that something deeper than just biology or linguistic principles are at work; the property may well "follow from principles of neural organization even more deeply rooted in physical law" (Chomsky 1965: 59).

All of this is intriguing, but of course it remains to specify exactly in what sense the X-bar pattern is optimal. This article attempts to go some distance towards exploring the details, but in the end falls short of motivating the X-bar pattern alone. Nevertheless, the weaker but more general conclusion reached below seems promising, namely that branching forms which look like a version of the X-bar pattern generalized to any number of specifiers are optimal. For now, I would like to point out the following intriguing analogy with plant growth.

2.4.   *An Analogy with Idealized Plant Growth*

Notice that in a binary-branching tree, each node is c-commanded and contained by a number of nodes equal to its depth in the tree. For reasons clarified in the next section, I will propose that the number of c-command and containment relations in a syntactic tree indicate a computational cost. This cost can be intuitively pictured as a 'force' pulling toward the root of the tree, in the sense that the deeper in the tree a given piece of structure is, the greater the number of c-command and containment relations it incurs.

Then the problem faced by the syntactic system is analogous to the following idealized problem of plant growth. Suppose that a plant is 'binary branching', and at each branching point, either new structure can become a terminal leaf, gathering sunlight but preventing further growth, or can grow a non-terminal stem which divides again in two. The plant 'desires' to grow as many leaves as possible (to gather sunlight energy more effectively) without making the resulting structure too tall/spindly. Vertical growth magnifies the structural strain involved in supporting the structure against gravity, wind, and so on, which is increasingly severe for each additional increment of growth away from the root (a longer stalk serving as a more effective lever for a given wind strength, and so on). Here, nature is searching for some compromise between growing as many leaves as possible, and not making the resulting form too tall.

If plant growth places two leaves at the very first branching, it is done growing. If it places one leaf at every branching point, only one branch will then remain available for further growth, resulting in a final form with a single stalk

(in linguistic terms, it is unidirectionally branching). Delaying leaf-generation for some number of branchings yields better results over the long term, as the final form will be bushier, shorter, and less likely to topple over from wind or its own weight. The very best final form would branch everywhere until spontaneously producing only leaves at the last generation. Of course, plants *grow*, making the notion of 'last generation' unavailable. What seems desirable is to strike some mini-max balance between growing as many leaves as possible immediately, and investing in optimality for future growth by growing more branches.

I will propose that syntax faces an equivalent problem (physical interpretation of the details aside, of course). That is, the 'cost' of branching structure grows with depth, such that each increment of deeper branching is costlier than the last (inducing more potential c-command and containment relations). The local 'force' on terminals is reduced by packing them as close to the root as possible, which is antagonistic to global optimality (each terminal which is too close to the root 'closes off' options for other structure, which must instead appear even deeper in the tree). In both botany and syntax, the Fibonacci pattern is a good compromise to this problem; perhaps even the best, depending on further details of the system.

3.   C-Command and Containment in Linguistic Computation

The primary tool of investigation in this article is the comparison of hierarchical structures on the basis of the number of c-command and containment relations they encode. Such relations are central to linguistic computations of various sorts (e.g., long-distance dependencies). Given the recent focus on principles of efficient computation, the hypothesis is that the derivation of structures with fewer such relations represents less of a computational burden. Insofar as different derivational patterns lead to structures with differing numbers of c-command and containment relations, there is then a basis in computational efficiency for preferring some derivational patterns over others. If we find that the recursive patterns which seem to characterize natural language are drawn from the patterns which are optimal in this sense, we may suspect that this aspect of phrase structure has a minimalist explanation.

I adopt the familiar definition of *c-command*, as follows:
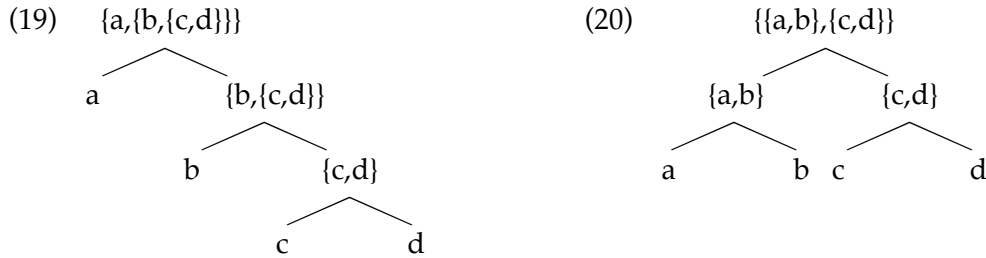
(18)   *C-Command* (Reinhart 1976: 32)

Node A c-commands node B if neither A nor B dominates the other and the first branching node which dominates A dominates B.

We will also be interested in *containment* (i.e. irreflexive domination), taken as the transitive closure of the 'immediately contains' relation. Note, first, that the totals of these relations are always equal in binary-branching trees. For each node α in a tree, the number of nodes which contain it is equal to its depth in the tree. Since the tree is binary-branching, the number of nodes that c-command α is also equal to its depth, because each node which contains α immediately contains a node β not containing α, which thus c-commands α; no other nodes c-command α.

3.1.   A Simple Observation

The point of departure for the present contribution is the simple observation that different patterns in Merge result in different totals of c-command and containment relations, even for the same input (number of terminals). For a simple example of this, consider sets (19) and (20).

(19)   {a,{b,{c,d}}}

```
        a        {b,{c,d}}

                 b        {c,d}

                       c        d
```

(20)   {{a,b},{c,d}}

```
        {a,b}        {c,d}

        a     b     c     d
```

These structures have equal numbers of terminals and of non-terminals, yet (19) has more c-command relations (12, compared to 10 in (20)). This is shown in (21) and (22), a listing of all the c-command relations present in (19) and (20), respectively (read "x: y, z, w" as "x c-commands y, z, and w").

(21)   {a,{b,{c, d}}}:   –
       a: {b,{c,d}},b,{c,d},c,d
       {b,{c,d}}:   a
       b: {c,d},c,d
       {c,d}:   b
       c: d
       d: c

       Σ = 12

(22)   {{a,b}{c,d}}:   --
       {a,b}:   {c,d},c,d
       {c,d}:   {a,b},a,b
       a: b
       b: a
       c: d
       d: c

       Σ = 10

For completeness, I list all containment relations for (19) and (20) in (23) and (24), respectively. Here, "x: y, z, w" means "x contains y, z, and w".

(23)   {a,{b,{c,d}}}:   a,{b,{c,d}},b,{c,d},c,d
       {b,{c,d}}:   b, {c,d}, c, d
       {c,d}:   c, d

       Σ = 12

(24)   {{a,b}{c,d}}: {a,b},a,b,{c,d},c,d
       {a,b}:   a,b
       {c,d}:   c,d

       Σ = 10

3.2.   C-Command in Linguistic Relations

The notion of c-command is central to numerous linguistic relations. Reinhart's (1976) original concern was describing the distribution of anaphora. While still relevant to binding theory, c-command is also implicated in linearization (Kayne 1994), the determination of relative scope (May 1985), and the probe-goal mechanism of Chomsky (2000, 2001), the latter taken to underlie long-distance agreement and to be a pre-condition for displacement.

Epstein *et al*. (1998) provide a natural reason for the ubiquity of the c-

command relation in terms of a derivational view of syntax. As they point out, c-command amounts to the condition that syntactic objects can enter into linguistic relations with elements of the sub-tree they are merged with. This suggests a view of c-command as following from a search operation (potentially) accompanying each Merge operation. The property of Minimality, as encoded by principles such as the Minimal Link Condition, Shortest Move, Attract Closest, and Relativized Minimality (the relevant literature is vast; see Chomsky 1995b, Rizzi 1990, among many others), reinforces this interpretation of c-command. The basic observation is that in configurations like (25), where X could enter into a dependency with either Y or Z but Y is 'closer' to X in some appropriate sense than Z is, a dependency may hold between X and Y but not between X and Z.

(25)   X … Y … Z

This closeness is usually measured by c-command relations: If Y asymmetrically c-commands Z, then Y is closer to a c-commanding X than Z is. To a first order of approximation, we might reasonably say that syntax seems to 'minimize links', presumably for reasons related to efficient computation. The idea is that long-distance dependencies reflect a search operation 'probing' for a 'goal' in the searched category in a top-down fashion (Chomsky 2001). Once an appropriate goal is found, the search terminates, thereby blocking a dependency with a more deeply embedded but otherwise legitimate goal (so-called 'intervention effects', possibly unifiable with the A-over-A Condition).

As one aspect of 'least effort' conditions on efficient computation, Chomsky (2000: 99) explicitly includes principles aimed to "reduce 'search space' for computation: 'Shortest Movement/Attract', successive-cyclic movement (Relativized Minimality, Subjacency), restriction of search to c-command or minimal domains, and so on". The last point is especially significant for our purposes: In terms of individual instances of search, the burden is less if a smaller domain is searched. But note that the total number of c-command relations in a syntactic object is simply the sum over the size of domains that have (potentially) been searched during its derivation. Thus, it is a natural extension of the drive to restrict the domains for individual searches to prefer structural patterns leading to lower c-command totals, since that amounts to restricting the aggregate domain for iterated searches.

C-command totals may be taken to indicate computational cost in other ways as well. Beyond the search interpretation of the probe-goal mechanism, Kayne's (1994) Linear Correspondence Axiom can be understood as a process 'reading' c-command relations and deriving linear order. Moro's (2000) theory of dynamic antisymmetry reinforces this view of linearization as a computational process at the interface, for him crucially applying after syntactic displacement has resolved points of symmetry. Likewise, scope is affected by displacement, again suggesting that some process 'reads' c-command relations at the interpretive interface. It seems natural to suppose that processes of linearization and the determination of scope are less burdensome if applied to objects with fewer c-command relations.

3.3. *Containment Computations*

There are reasons to believe that certain linguistic computations are 'measured out' by containment relations, such that minimizing the number of such relations improves computational efficiency. For example, Chomsky & Halle (1968) note a relationship between stress levels and hierarchical set structure in complex expressions. They propose a cyclic rule of stress assignment (the Nuclear Stress Rule) that re-computes stress in successive applications from most-to-least embedded levels. See Halle & Vergnaud (1987) for a broadly similar system, as well as Hayes (1995). In all of these theories, the stress on individual items may potentially be readjusted at each level of embedding. Importantly, an arrangement like (20) involves fewer total (potential) adjustments of the stress levels on individual elements than (19). That is, in (19) the most deeply embedded elements (*c* and *d*) will be subjected to three cycles of stress computation; the element *b* will undergo two cycles, and *a* just one: The total is 9 (potential) readjustments of individual stress levels. On the other hand, in (20) each element is twice-embedded, hence subject to 2 cycles, for a total of 8 potential readjustments. This suggests that assigning stress to (20) is a simpler computation than doing the same for (19).

Along the same lines, consider theories that relate displaced elements to their position of canonical interpretation in the way that Head-Driven Phrase Structure Grammar (Pollard & Sag 1987, 1994) does.[9] To encode discontinuous dependencies (e.g., in *wh*-questions), HPSG utilizes a feature on a verb (a SLASH feature in the HPSG parlance) marking its semantic deficiency, which propagates up the tree along the path of dominating nodes until it encounters a category that can satisfy it. If some such mechanism underlies displacement phenomena in general, then one natural condition of efficient computation is that the feature propagation path should be as short as possible. Maximally balanced trees like (20) provide a scaffolding with the minimal propagation path-length sum possible; in general terms, the 'average' containment path is shorter in such a tree, and the worst-case paths are shorter than in any other structure.

3.4. *Is Counting Enough?*

I will resort to simply counting all of the c-command (equivalently, containment) relations in a structure, adopting the working hypothesis that this is a reasonable proxy for the 'real' computational cost incurred in actual expressions. It might be objected that counting all c-command relations may overestimate the relevant cost in important ways. For one thing, it is often assumed that in a given configuration, the relations for which c-command matters are one-sided. Thus, when α and β merge, only one (say, α) can search the other; dependencies cannot be established from β into the interior of α. Moreover, the very fact of intervention means that not all probe-goal searches are computed; the search

---

9   Of course, HPSG is a model-theoretic approach to syntax. It is not clear that concerns of 'efficient computation' are as relevant to such an approach as to proof-theoretic derivational accounts pursued within the Principles–and–Parameters tradition.

stops once the first legitimate target is found. Thus, it seems we are crucially over-counting the c-command relations that should matter for such a comparison. Similar concerns apply to containment; for the case of stress assignment, it seems only the containment relations involving terminals matter for optimality.

This is a necessary casualty of the idealizations here. To restrict dependencies such that only one of the operands of Merge may search the other requires some basis for the asymmetry. Given the range of constructional options considered here, there simply is no way to reconstruct such an asymmetry on configurational grounds in full generality. Since any other grounds for the asymmetry (say, properties of the individual lexical items involved) are ignored as well, we shall have to live with this. Similarly with the intervention effect: Without knowing what dependencies might actually be established or not, we are left with a bare scaffolding of possible dependencies and no way of choosing how it might be filled out. The only basis for comparison is the scaffolding itself.

Even so, I think the approach here is not unreasonable. Recall that the goal is to find a basis for selecting certain structural patterns over others. At this level of idealization, it may make sense to abstract away from the details and consider the total space of possible relations latent in branching forms themselves. However the possibilities are eventually exploited in particular expressions by some defined relations entering into linguistic computations while others do not, it is a fact that some structures put a tighter cap than others on the computational cost that could be incurred in principle.

Furthermore, it is a crucial point that the measure of computational cost need not be strictly accurate for our purposes; all that is important is that it reflects the *relative* optimality of the structural options being compared. In this regard, it is encouraging to note a general property of scale-invariance in the comparison between different recursive possibilities. As we will see, the self-similarity of the patterns to be explored implies that if one pattern produces fewer c-command and containment relations than another in small domains, their relative optimality will not be reversed in larger domains. The same property of self-similarity suggests that the comparison will tend to go the same way if domains are restricted in principle in equivalent ways.

The hypothesis — and it is only that — is that at this level of abstraction, this simple expedient of counting will suffice to illuminate at least the outlines of where such an approach will lead. But suppose it turns out that simply counting total numbers of c-command and containment relations is wrong in some fundamental way, and a more detailed look at the properties involved leads to different measurements, making different predictions. Even so, I would like to suggest that such predictions should be taken seriously, and their explanatory potential explored. In other words, the methodology employed here may prove to be too simplistic, but I think the underlying concerns deserve attention, in that (to my knowledge) computational efficiency in this form has not been examined before, and the potential for 'deep' linguistic explanation in these terms appears promising.

## 4.   Optimal Syntactic Growth Modes

What I propose to investigate and compare below are phrase structural patterns, in the sense of characteristic aspects in the branching geometry formed by Merge applying recursively to lexical items and its own output. The hypothesis being entertained is that the forces which govern the process, in the sense of selecting some binary-branching structures over others, will give rise to identifiable and repeated tendencies (what might be thought of as 'optimal growth modes'). To determine what tendencies we might expect, I generate all possible patterns that could be used as consistent 'phrasal templates' to build infinitely recursive structures from lexical atoms, and develop a technique to compare them to each other.

4.1.   A Domain for Terminals

One condition that will need to be imposed is that the recursive templates include a characteristic place for terminal elements. This makes a good deal of sense on several levels. First, the objects are recursively defined, which requires some 'base step'; it is hard to see what aspects of branching structure could provide this other than terminals. From another point of view, these are ultimately discrete, finite patterns, built bottom up from lexical items; they are 'about' structuring terminals into larger structure. Without terminals to 'ground' the patterns, there can be no distinctive shape, hence no 'pattern' at all; the only rule would then be 'anything goes'.

      The concern in this regard is structures like (26) below, which are 'maximally balanced', with all terminals at the same depth (or at two adjacent levels of depth). These structures provide absolute minimization of c-command and containment relations.

(26)

$$A^0 \quad B^0 \; C^0 \quad D^0 \; E^0 \quad F^0 \; G^0 \quad H^0 \; I^0 \quad J^0 \; K^0 \quad L^0 \; M^0 \quad N^0 \; O^0 \; P^0$$
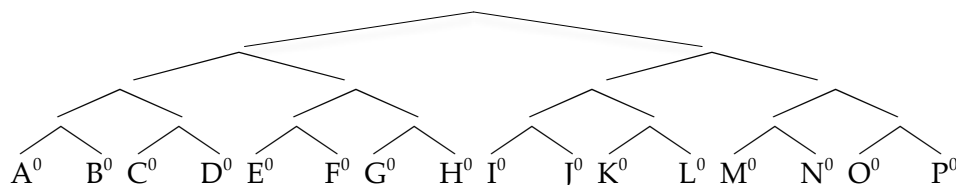
      If the concerns in this article really do 'matter' in the determination of structure, why do we not see such forms in natural language? If the only problem were optimizing *at once* the positioning of a full set of elements, we would indeed expect to see something like this.

      But one guiding theme in minimalist work is the idea that syntactic forms are to be explained dynamically, by local (informationally limited) optimization at each step of a syntactic derivation. In these terms, the structure above looks decidedly unnatural. To actually derive such a form, Merge must apply as symmetrically as possible. This involves unbounded 'vertical' information flow at each step; the internal structure of syntactic objects must be accessible 'all the way down' so as to match objects (terminals, pairs of terminals, pairs of pairs of terminals, etc.) appropriately. But even this 'local' (i.e. one Merge operation at a

time) matching of object structures is not enough. The derivation must be kept in appropriate synchrony across the entire set of parallel sub-derivations; if one process of merging terminals into ever-larger sets proceeds too many steps beyond other combinations occurring in parallel, we may be left with a final stage where only unmatched objects remain. Information must thus be shared 'horizontally' as well, in effect amounting to global pre-planning of the derivation.

We can identify a parallel situation in botanical growth. Recall the idealized problem of leaf-placement in section 2.4: The ideal representation for solving the problem produces no leaves until the last generation, when only leaves are produced. There is something distinctly unnatural about this; organic growth proceeds by a local logic, where notions such as 'final form' have no power to shape the dynamics of growth. Similar concerns apply to the pattern of Fibonacci spirals in phyllotaxis: If the only problem were to pack at once a certain number of elements into a limited space, a hexagonal lattice structure would be best. But the observed patterns *grow*, with the result that what we in fact observe is not the best form, but the best growth pattern, a crucial distinction.

Given the dynamic view of syntax adopted here, similar constraints are expected to apply: The best configuration is 'ungrowable'. Parallel to the phylotactic case, we expect to observe at best an optimal derivation, not an optimal final representation, because the dynamic system is limited by a fundamental locality. This is why (26) is not predicted here; no local pattern of growth can produce it.

4.2. Possible Growth Modes

Such concerns lead us to expect that the considerations which enter into derivational choices will be limited by an informational horizon. Recall that one of the problems with (26) was that it required syntactic objects to be matched 'all the way down'. Limiting this informational flow means that only some of the recursive structure of the operands of Merge is 'visible' to optimization concerns. For example, if one level of internal structure can be examined, then terminals can be distinguished from more complex objects. Allowing two layers of structure to be visible allows further distinctions, which allows more internal complexity in recursive patterns, and so on.

As an idealization to aid the investigation of these matters, I will suppose that whatever pattern might be found will be consistent (i.e. deterministic). A consistent recursive scheme carried out within a finite derivational window can be described by a finite number of distinct 'types' of syntactic object (terminals, or objects recursively defined as the result of Merging other terminals or recursively defined objects), which 'loop' into each other in a finite cycle.
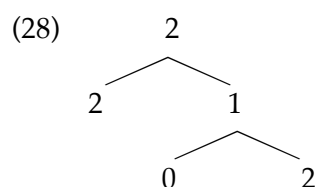
4.2.1. Notational Conventions

To allow the full range of recursive possibilities, let us simply use the natural numbers to represent the relevant distinctions among outputs of different Merge operations, reserving 0 for terminal elements. Let us furthermore use the largest

number in a pattern to designate the root symbol (held constant, under the 'top-down' formulation discussed below). Here, we will take the appearance of the same number on two different nodes to mean that the structures so labeled have isomorphic recursive structure. In these terms, the simplest recursive pattern (both including terminals and allowing indefinite recursion) will be represented as below:

(27)

```
        1
       / \
      1   0
```
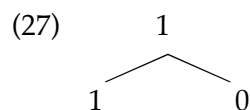
Likewise, in this formulation the X-bar specifier-head-complement pattern will have 0-level terminals marked as 0s, while 'single-bar-level' intermediate categories are 1s, and 'phrases' are 2s.

(28)

```
         2
        / \
       2   1
          / \
         0   2
```

Thus, the numerical designations might be thought of as something like a generalization of conventional 'bar-level' notation. To be clear, this is *not* a proposal about reviving bar-level notation as an explicit grammatical device, thus violating Inclusiveness. Instead, the notation is a device for reasoning about possible derivational sequences; the relevant information is not to be understood as somehow reified in any way 'on' the node, but is a matter of information that is in the way the derivation itself proceeds. If these patterns do characterize natural language, that fact presumably emerges from dynamic considerations, rather than being explicitly enforced by some mechanism like 'bar-level features'.

Insofar as a pattern is consistent, its elements (other than 0) can be characterized by what amount to 'rewrite rules' (again, this is a matter of investigational convenience, not a proposal for a 'real' grammatical device):

(29)   i → j k   i in $\{1, 2,\dots n\}$; j, k in $\{0, 1, 2,\dots n\}$

The simplest structure (27) can thus be expressed as in (30), and the X-bar schema as in (31):

(30)   1 → 1 0

(31)   2 → 2 1
       1 → 2 0

4.2.2. Generating All Possibilities

Let us now set to exploring the options systematically. If the 'derivational

window' is as small as possible (i.e. the growth pattern is as simple as possible), then there is only one option for how to build recursive structure from terminals. I call this the 'spine', for obvious visual reasons (intuitively, it generates a uni-directionally branching tree); I will likewise use descriptive names for the other patterns for mnemonic convenience.

(32) 1 → 1 0 ('spine'):

```
        1
       / \
      1   0
```

We obviously need at least this much structure to have recursion at all. Ignoring linear order (as I do throughout), and requiring the pattern to be built recursively from terminal elements and the output of Merge, for distinct objects 0, 1 the other combinations can be ruled out (1 → 1 1 is not built from terminals, while 1 → 0 0 does not recurse).

Moving on to the next level of complexity in sequencing Merge, we consider patterns involving two types of non-terminals (equivalently, two-stage sequencing of Merge operations). Given the remarks above, we have at first pass $6^2 = 36$ distinct options for recursive patterns involving two order-irrelevant Merge rules (i.e. non-terminal characterizations) defined over three object types (0, 1, 2); for arbitrary n, there are $(n(n+1)/2)^{n-1}$ options. Being a little more careful, we can restrict this further by ruling out the following types of characterizations:

(33) i → i i does not terminate (DNT)
    n → 0 0 does not recurse (DNR)
   n → n 0 isomorphic to the Spine

That is, any object which immediately contains two isomorphic copies of itself cannot be recursively constructed from terminals. If the root node (designated as the largest number n) consists of two terminals, recursion is impossible. Finally, if the root node consists of a terminal and an object isomorphic to the root, it is isomorphic to the spine (1 → 0 1), hence is not really a member of the higher-order comparison set. The table below lists all the options for the comparison set built from {0, 1, 2}; non-viable options are grayed out.

|	2 → 2 1	2 → 1 1	2 → 1 0
1 → 2 2	DNT	DNT	high-headed D-bar
1 → 2 1	DNT	DNT	high-headed X-bar
1 → 2 0	X-bar	D-bar	(spine)
1 → 1 0	spine of spines	pair of spines	(spine)
1 → 0 0	double-headed spine	DNR	DNR

Table 1: Options for the comparison set built from {0, 1, 2}

      I have also grayed out the option described as a 'pair of spines', which, as the name is intended to suggest, consists of two spines merged at the root. It should be clear that this is not a repeating structure; the configuration at the root is unique, and thus it is not a growth pattern in the desired (basically, self-similar) sense. I illustrate the remaining options below, including their repeating 'molecular' structure as a partial tree diagram.

(34) a.   2 → 2 1   ('X-bar')
             1 → 0 2

      b.

```
            2
           / \
          2   1
             / \
            0   2
```
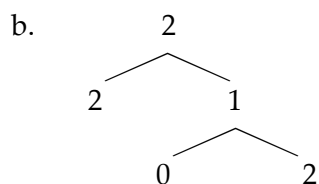
(35) a.   2 → 1 0   ('high-headed X-bar')
             1 → 2 1

      b.

```
            2
           / \
          0   1
             / \
            2   1
```
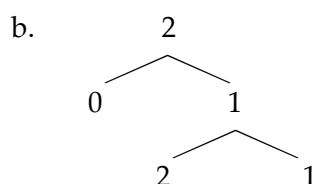
      Options (34) and (35) form a natural pair, as do (36) and (37) below, in that the members of the pairs are really the *same* recursive cycle caught at different times, with a different selection of which non-terminal serves as the root. I call the member of each pair of patterns in which the terminal occurs nearer to the root 'high-headed'. See the discussion in 4.3.3.2 below.
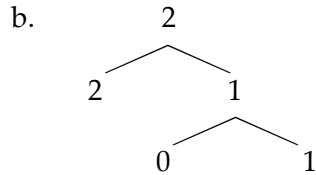
(36) a.   2 → 1 1   ('D-bar')
             1 → 2 0

      b.

```
            2
           / \
          1   1
             / \
            2   0
```
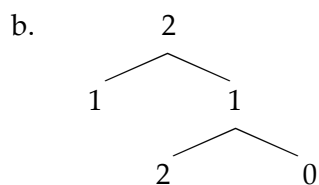
(37) a.   2 → 1 0   ('high-headed D-bar')
             1 → 2 2

      b.

```
            2
           / \
          0   1
             / \
            2   2
```
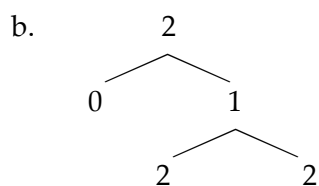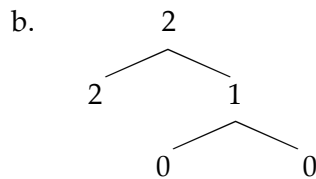
This pair (again, really different 'snapshots' of the same pattern) has a fundamental symmetry; the D in D-bar is meant to stand for 'double' for this reason.

(38) a. $2 \rightarrow 2\ 1$ ('spine of spines')
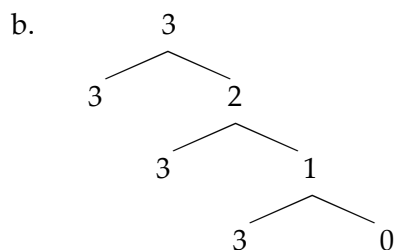$1 \rightarrow 1\ 0$

   b.
```
            2
          /   \
         2     1
             /   \
            0     1
```

(39) a. $2 \rightarrow 2\ 1$ ('double-headed spine')
$1 \rightarrow 0\ 0$

   b.
```
            2
          /   \
         2     1
             /   \
            0     0
```

Enumerating all of the options for further comparison sets (allowing three stage Merge sequences/three non-terminal types) would be a good deal more tedious. For illustrative purposes, I include just one of the options. This represents the 'projective' geometrical format, and thus is the optimal member of its class (for reasons discussed below, and proven in full generality in the Appendix). Intuitively, it corresponds to the structures described by Jackendoff's (1977) 'uniform three-level hypothesis', an X-bar-like structure with two specifiers. In other words, it is a version of the X-bar schema utilizing three non-terminal types; hence, '3-bar'.

(40) a. $3 \rightarrow 3\ 2$ ('3-bar')
$2 \rightarrow 3\ 1$
$1 \rightarrow 3\ 0$

   b.
```
            3
          /   \
         3     2
             /   \
            3     1
                /   \
               3     0
```

4.3. Comparing Growth Modes

Now that we have developed a way of enumerating the possibilities for recursive growth modes, we turn to the task of comparing them to each other. Recall the fundamental observation underlying this investigation, that building structure in some ways results in fewer c-command and containment relations than other options. I have argued that having fewer such relations lessens the computational

burden for the derivation. The hypothesis is that this results in a preference for patterns in the application of Merge that will tend to reduce c-command and containment relations. Our goal in this section will be to develop a technique to compare the recursive options we have enumerated on the basis of their consequences for c-command and containment totals.

4.3.1. Comparison Sets Based on Cycle Complexity

Each of the recursive patterns we are considering is defined within the bounds of some fixed amount of sequential complexity. Some patterns have more or less internal structure than others: The spine is 'simpler' than the X-bar schema. The X-bar schema requires more in the way of (relatively local) information flow to structure the derivation appropriately. Different choices of the size of the derivational window (i.e. the number of different types of object, or equivalently, the number of derivational steps in a characteristic cycle) will partition the possibilities into natural comparison sets. That is, we will compare recursive patterns of comparable complexity to each other. In present terms, we will be comparing patterns that can be specified with the same number of symbols, so that a comparison set will consist of all the recursive possibilities that can be described with numbers from 0 to some fixed n.
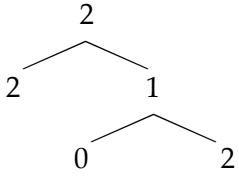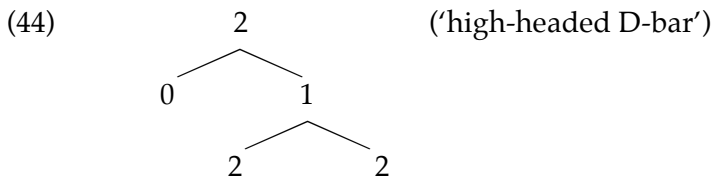
4.3.2. Direct Comparison

How can one growth mode (recursive pattern) be compared to another? Sometimes the comparison can be made quite directly. Consider again the following example from the introduction. We are given the problem of combining the syntactic objects AP, BP, and X^0 via binary Merge. AP and BP are internally complex, while X^0 is a terminal. The options are these:

(41)   [AP [X^0 BP]]   (or [BP [X^0 AP]])

(42)   [X^0 [AP BP]]

Again, given just the information that AP and BP are internally complex, the first option produces fewer c-command and containment relations than the second. Noticing the monotonic way in which c-command and containment relations accumulate in a derivation (i.e. additively), this local superiority gives us very good reason for preferring to apply the pattern manifested in the first option over the second more generally, if we are forced to choose one or the other as a repeated format. Put another way, it motivates the choice of the growth mode (43) over (44):

(43)   2   ('X-bar')
            2   1
                  0   2

(44)   2   ('high-headed D-bar')

```
          2
         / \
        0   1
           / \
          2   2
```

      However, this sort of direct comparison will not work for the full comparison set they belong to. Consider another member of that set:

(45)   2   ('double-headed spine')

```
          2
         / \
        2   1
           / \
          0   0
```
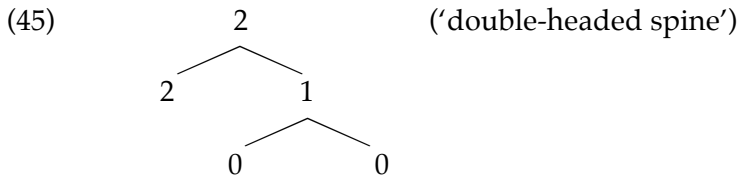
No local, direct comparison with the previous two patterns is possible, since they take different inputs (23 calls for two terminals); in general, where (43) and (44) can be applied, (45) cannot.

4.3.3. Indirect Comparison

To get around this problem, I will proceed as follows. First, it is an inescapable fact that these are discrete patterns, ultimately built from some finite number of terminal atoms. This suggests an alternative, slightly indirect way to compare different growth patterns: Compare the set of tree-forms they can generate for some constant number of terminals.

      These patterns implicitly define a class of trees. For example, The Spine can be applied to generate (46); that unidirectionally branching structure belongs to the set of trees associated with the growth mode (such a tree can be 'grown' by the pattern). On the other hand, (47) does not belong to the class of trees associated with the Spine.

(46) $[W^0 [X^0 [Y^0 \ Z^0]]]$

(47) $[[W^0 \ X^0] [Y^0 \ Z^0]]$

      For a fixed number of terminals, there are many different binary-branching arrangements of that number of elements. Some of those branching structures will belong to the class of trees associated with a particular phrase-structure pattern, and some will not. These will typically differ in their number of c-command relations. However, for a fixed number of terminal elements and a particular recursive pattern, we can identify the *best* tree(s), which contain the fewest number of c-command relations of any of the trees associated with a particular pattern. These best trees for a number of terminals then serve as a basis for comparison among the patterns themselves (since, as it turns out, this comparison is monotonic: If a pattern allows a better tree for n terminals than any competing pattern, it also has a better tree for $n+m$ terminals).

4.3.3.1. The 'Bottom of the Tree' Problem

However, this requires some further clarification. The idea is to find some way to compare templates for infinite growth, by isolating them and seeing what happens when they are followed as faithfully as possible. The problem is that none of these rules can be followed completely faithfully. This is an inevitable consequence of insisting that they allow for indefinite recursion: Any such growth pattern must contain 'slots' for other objects of indefinitely large size. Yet the objects which manifest these patterns must ultimately be finite, with nothing but terminal nodes at the bottom of the tree. As a result, some 'slot' that calls for a larger object must be filled with a terminal instead.

To illustrate, consider the simplest possible growth rule for combining terminals into an indefinitely large recursive structure:

(48)

```
        1
       / \
      0   1
```
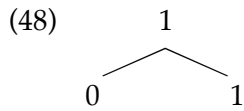
Even in this, the simplest pattern, the very first step in a derivation presents a problem, as it does not follow the rule. Any derivation whatsoever must begin by creating a structure of the form [X^0 Y^0]; there simply is no other option. So for a pattern like (48), we will accept a structure like (49) as manifesting it as faithfully as possible:

(49)

```
        1
       / \
      0   1
         / \
        0   1
           / \
          0   1
             / \
            0   1/0
```
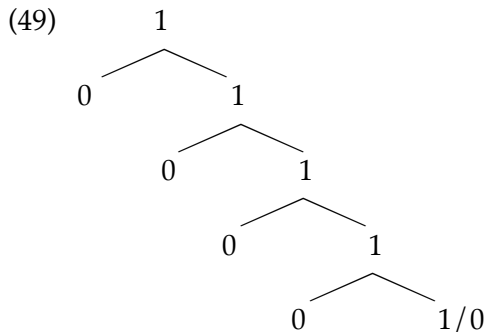
The notation 1/0 indicates where we have deviated from following the growth rule (necessarily, since the tree is finite), here including a terminal where the rule calls for a complex object.

However, if we must allow some 'fudging' at the bottom of the tree, we can at least be faithful everywhere else. Keeping in mind that our ultimate goal is to find some basis for comparing one growth mode to another, we reason that we do not want to 'truncate' the pattern encoded in the growth rule anywhere not required by the brute fact of discreteness. In particular, we will insist that the growth pattern be followed faithfully 'in the middle' of the derivation, so to speak. This amounts to the formal specification that the only deviation from the recursive pattern allowed will be replacing a called-for non-terminal with a terminal. We rule out non-terminal to non-terminal sequencing that violates the pattern, as in (50) below. Here, the notation *0/1 marks the illegitimate portion:

A called-for terminal has been filled with a non-terminal instead.

(50)

```
                      1
                    ⌒
          * 0 / 1          1
          △    △
```
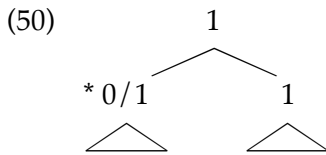
4.3.3.2. Top-Down Generation

Note that we have imported a further complication by the convention of assuming that one of the non-terminal types (*n*, the highest of the numbers designating the non-terminal types) will be uniformly associated with the root. Formally, this amounts to generating the trees to be compared from the root down, allowing any branch to terminate. It is an important (if subtle) point that this is not a matter of committing to a top-down view of syntactic derivation, though it should be recognized that a Merge-based system need not be quite so literally bottom-up as often assumed:

> Thus if X and Y are merged, each has to be available, possibly constructed by (sometimes) iterated Merge. […] But a generative system involves no temporal dimension. In this respect, generation of expressions is similar to other recursive processes such as construction of formal proofs. Intuitively, the proof 'begins' with axioms and each line is added to earlier lines by rules of inference or additional axioms. But this implies no temporal ordering. It is simply a description of the structural properties of the geometrical object 'proof'. The actual construction of a proof may well begin with its last line, involve independently generated lemmas, etc. The choice of axioms might come last.   (Chomsky 2007a: 6)

Regardless, in the present investigation top-down generation is an artifact of notational choices, rather than a substantive claim.[10] Recall that the objects of interest are recursive cycles. Understood as time-neutral geometric patterns of recursion, these patterns do not properly have a 'beginning' or an 'end' (other than terminal elements, which can in principle appear anywhere in the looping structure as inputs to Merge, but not outputs). Their structure is a matter of how outputs from one step loop into the input to other steps. But we have kept to the familiar tree-diagram notation, assigning numerical designations to non-terminal types. The result is that certain patterns are multiply represented. For example, 'X-bar' and 'high-headed X-bar' are really the same recursive pattern, with a different choice for which non-terminal occurs at the root.

However, it turns out that a certain orientation of the pattern (fixing one or

---

10   In light of this point, the claim made in this article about 'projective structures' needs to be clarified somewhat. Represented in the format [α [β … [γ [X^0 δ]] …]], the claim is a little too strong. What is motivated here is rather the recursive cycle underlying this format. Put another way, even universal strict adherence to such a growth mode in reality would not necessitate that the root node be maximal; the recursive cycle could be oriented differently at the root, thus showing up as one of the 'high-headed' alternatives (such a situation would look like a 'small' projection at the root embedding an otherwise well-behaved projective structure).

another of the non-terminal types at the root) will consistently provide better results than others. Thus for each looping object we can generate a set of alternate versions fixing one or another of the stages as the 'top', corresponding to the 'root' of a tree, and see which are best. Since the basis for comparison is best performance, this should not present a problem in any way.

4.3.3.3. Some Results from Indirect Comparison

Figure 2 graphs the growth in c-command and containment relations for several recursive patterns. Recall that for each growth mode, there is an associated set of trees generated by adhering to the structural pattern consistently from the root down, allowing terminals to appear in 'slots' calling for non-terminals (required for finite trees). For a given number of terminals, a number of trees can be generated by a given pattern. These will differ in the number of c-command and containment relations they encode, but for each choice of growth mode and number of terminals, there will be a best tree (or set of such trees). A 'best tree' has the fewest possible c-command and containment relations that could be produced by that growth mode for that number of terminals. It is these totals which appear in Figure 2 (as a function of the number of terminals).
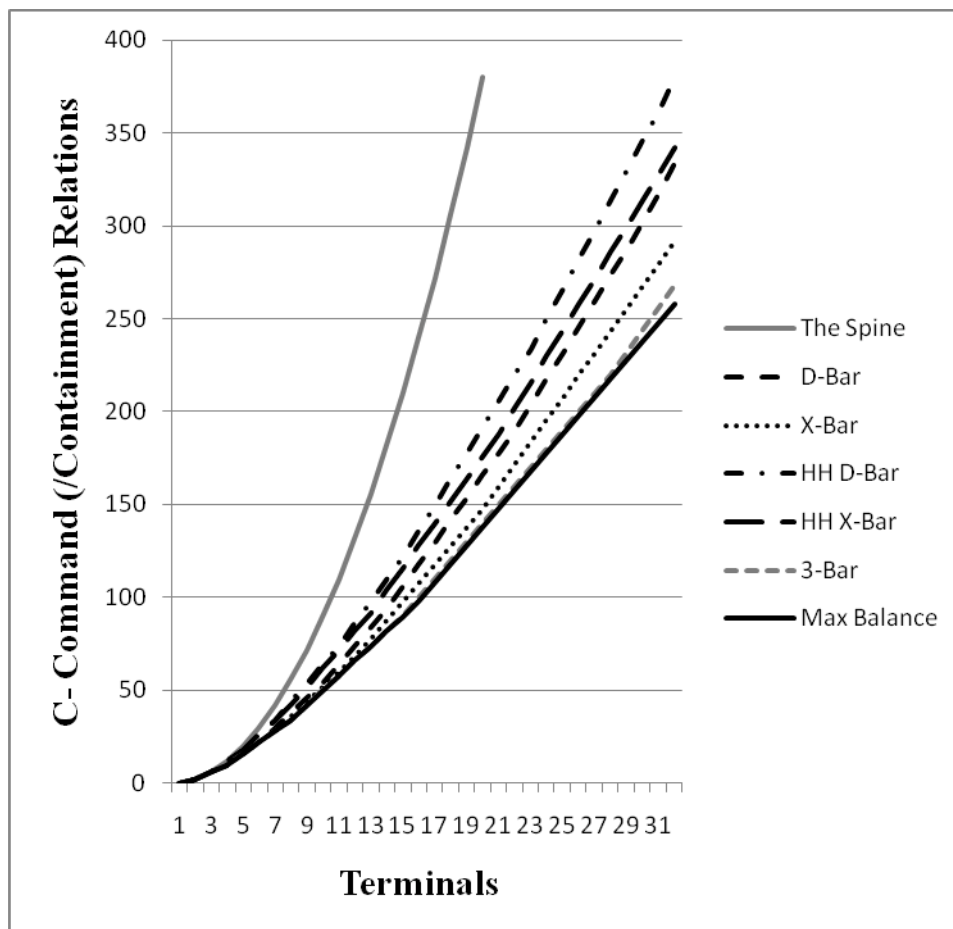

Figure 2: C-command and containment totals as a function of terminals in best trees

I include in the figure 'best trees' in X-bar (34), as well as three other two-layered constructional schemes (35-37). I also include the best system utilizing a 4-way combinatorial distinction (40), which I call '3-bar' (intuitively, an X-bar-like system with two types of intermediate category). The spine (32) forms the upper boundary curve; no growth pattern results in worse performance (in the sense of creating more c-command and containment relations for a given number of terminals). There is also a lower boundary curve, here labeled 'Max Balance'. This is the number of c-command and containment relations in a maximally balanced tree like (26) from section 4.1; the pattern is not the result of any finite growth pattern, but forms the boundary on best-case performance.

Among the growth modes in its comparison set, X-bar has the best performance: Its curve is closer to the best case lower boundary ('Max Balance'). The optimal pattern from the next comparison class, '3-bar', has slightly better performance (the best trees that can be 'grown' by that pattern have fewer c-command and containment relations for the same number of terminals).

To be clear, the figure is meant as an illustration, not a proof. The general result that projective growth modes are best is established formally in the Appendix.

4.4.   *Deriving Projection*

As suggested by Figure 2, X-bar is the best growth mode that can be achieved by any two-stage scheme for constructing recursive structure from terminals via binary Merge. What I call '3-bar' is better still, though it requires more distinctions (more recursive complexity, more information flow) to construct. Generalizing, these are examples of the 'projective' format in (51), where X^0 is a terminal at the 'bottom' of the repeating structure, and α, β, and so on are objects themselves constructed according to (51).

(51) $[\, \alpha \, [\, \beta \, \ldots \, [\, \gamma \, [\, X^0 \, \delta \,]\,] \, \ldots \,]\,]$

The structural properties of (51) can be captured in our alternate notation as in (52), where 0 is a terminal, and *n* the non-terminal associated with the root.

(52)   n → i n
       i → j n
       …
       k → 0 n

The specifier-head-complement format of X-bar theory is one example of such a 'projective structure': Specifically, it is (52) with *n*=2. The more optimal '3-bar' system of (40) is another example, this time with *n*=3. As I prove in the Appendix, this is the optimal format for *n*+1 (i.e. 0, 1, … *n*) types of category (many other less optimal possibilities exist). Intuitively, the idea is as follows. The phrase structural possibilities are understood to be (partially) realized by finite expressions, built bottom-up by Merge. As such, every recursive pattern must include terminals (0s) as one of its structural types. Moreover, no categories are

built solely from non-terminals 'all the way down'.

Given these restrictions, and the determinacy of the structural characterizations assumed, any non-terminal node must dominate a terminal node within depth n, for $n+1$ types. The best kind of structure, following the format in (52), introduces terminals no closer to the root than forced by this. In essence, introducing terminals too close to the root 'closes off' branches, forcing complex structure to appear deeper in the tree, where it will induce more c-command and containment relations than if it were shallower. The format in (52) allows arbitrarily large structures to be as balanced as possible given the limitations resulting from finitely many structural distinctions.

Note two very interesting properties of (52):

(53) a. Every non-terminal immediately dominates a root-type node.
      b. Terminal nodes and root-type nodes are associated one-to-one; a single terminal occurs at the lowest level of the chain of non-root-type nodes dominated by a root-type node.

Replace 'root-type node' with 'maximal projection', 'terminal' with 'head', and 'chain of non-root types dominated by a root type' with 'projection chain', and we have:

(54) a. Every non-terminal immediately dominates a maximal projection.
      b. Heads and maximal projections are associated one-to-one; a single head occurs at the lowest level of the projection chain.

That is, the recursive scheme that best minimizes c-command and containment relations has geometric properties corresponding to (54a) the maximality of non-head daughters, and (54b) endocentricity. Such properties are the essence of the theory of projection. But the notions entering into (53) are purely structural ones. Does this 'derive' projection? Not in the sense of literally providing labels on non-terminal nodes. But it suggests a reason for syntactic objects to tend to take the form of structures which are 'ready-made' to be 'read' as projections, in that there is a natural one-to-one association in the optimal format between larger molecules of structure and unique terminals at their 'bottom'.

5.   Lexical Features and Projection

This article has been concerned with matters of pure hierarchical geometry, paying no attention to lexical details at all. To say the least, this is orthogonal to the sort of approach pursued in recent work. Following the seminal work of Speas (1990), phrase structure is generally understood to be determined by the specific featural requirements of lexical items, hence 'projected from the lexicon'. Taking this view in conjunction with the principle of Last Resort, which holds that syntactic operations are driven by strict necessity (Chomsky 1995b), there is little room for other principles to play a role in phrase structure. In the strongest version of such a conception, each step in the structure-building process is

required; if some other step were taken, some lexical feature would not be checked appropriately, and the derivation would crash.

5.1. Beyond Features and Last Resort

To be clear, I see no reason to deny that Last Resort accurately describes the mechanisms in play, at an appropriately detailed level of description. But focusing too narrowly on the mechanisms involved may limit the depth of explanation that might be achieved. Consider, for example, the mysterious EPP (Extended Projection Principle) property (Chomsky 1981), which requires that T(ense) must have a filled specifier. This can be enforced by supposing that the relevant head has an EPP feature (or some equivalent device), and that the derivation will crash if the specifier position is not filled. But does this actually explain the EPP, or merely describe it? On this view, it is an accident that T has such a feature; it could just as well have lacked that feature, and then there would be no EPP. What is left unanswered is *why* T should have such a requirement in the first place; can that be explained in some naturalistic way?[11]

   The usual approach is to take lexical properties as given *a priori*, with the task of syntax being to accommodate them as best it can. The present approach could be understood as exploring causation in the other direction (i.e. the extent to which syntactic effects might explain lexical properties). Without presupposing that lexical requirements *have* to be what they are, all options are on the table, so to speak. The ultimate goal is to eventually use the insights of the present investigation to achieve a deeper understanding of lexical facts (for example, why an EPP feature for T might be preferred), though this further step is left for future work. That is, independent of the mechanisms which effect structuralization, we may ask about the optimality of the patterns they induce. Insofar as those patterns turn out to be optimal in the sense explored here, they are as expected — 'perfect', Galilean, and explainable in the minimalist mode.

   Thus, the point of view here is compatible with even the strictest understanding of how a derivation might be driven by lexical features, if it is allowed that principles of optimality might play some role in determining lexical features.[12] If so, the concerns explored in this article are rather far in the background, indirectly realized through patterns 'frozen into' the lexicon. On the

---

11   Earlier drafts of this work included material showing how concerns of minimizing c-command and containment relations plausibly play a role in displacement, including EPP-movement. The crucial point is that with respect to some of the computations involving such relations at the interfaces, displaced elements are effectively in their displaced position and *not* in their 'base' position (linearization and scope being clear cases), thus opening up the possibility that displacement might serve to derive a more economical form, in the relevant sense. Although the predictions here seem extremely promising, the issues that arise go too far beyond the scope of this article.

12   This may seem odd at first, if 'features' are understood as properties related to interpretation at the interface. Two comments are in order. First, the sorts of features that could most readily be explained by computational considerations are so-called uninterpretable features, which have the dual properties of not being interpreted directly, and seemingly playing a crucial role in structuralization. Second, it may after all be sensible to rethink some properties formerly considered to be properties of interpretation in terms of syntax-internal concerns, pursuing the general program of Hinzen (2006) along these lines.

other hand, in his most recent work Chomsky has suggested that Merge may be driven by a non-specific generalized 'Edge Feature' EF (Chomsky 2007a), which is undeletable in syntax (hence allowing unbounded Merge). If that point of view is adopted, the options for structure are not nearly so rigidly and predictably forced by the choice of lexical items, and the considerations here may play a rather more direct role in determining structure.

5.2.   *Lexical Requirements: Projection vs. Structuralization*

An essential step in the argument is the idea that many possibilities are logically possible for phrase structure. In particular, I argue that we should be willing to be surprised that syntax makes use of 'projective' geometries, wherein terminals occur at the bottom of phrases. But at first glance, this would seem to present a problem: How could it be any other way? That is, if local structures are indeed enforced by lexical requirements, how could a terminal affect structure anywhere except in the sort of domain defined by a projective structure?

What is at stake is the power to enforce structural features, and how far that extends. It is uncontroversial that certain lexical items can force certain structural choices in subsequent Merge operations beyond the first they occur in (i.e. higher up the tree). For example, the item T (Tense), even after Merging with its complement, is able to enforce further details of the derivation, in the form of its EPP property requiring a phrase to occur in its specifier. It seems that T has the reach to place non-local requirements on the structure it occurs in.

But what is required for (lexically-driven) non-projective geometries is that structural enforcement can reach down the tree as well as up. That would require lexical requirements to be discharged 'before' the enforcing head has been Merged; is that not a paradox? Crucially, this sort of situation is not just possible, but empirically attested. The relevant case is long-distance selection: Selectors have the power to enforce properties not just in their complements, but in the interior of their complements as well (thus 'down' the tree).

Boeckx (2008) gives the following example from Hebrew. In Hebrew, as in English, the verb meaning *ask* selects for a [+*wh*] CP. What is important, for our purposes, is the presence of the topicalized phrase *ha sefer* to the left of the [+*wh*] element *le mi*:

(55) Sa'alta   oti et   ha sefer le mi   le haxzir.
      asked.2.SG   me ACC   the book to whom to return
      'You asked me to whom to return the book.'   (Boeckx 2008: 16)

This is a clear case of long-distance selection going beyond strict head–complement. That is, assuming the articulated left periphery of Rizzi (1997), the selected [+*wh*] material occurs embedded inside the phrase hosting the topic. Boeckx notes that Grimshaw's (1991) notion of Extended Projection, or other feature-passing devices, doesn't solve the problem:

> More specifically, it is unclear what it would mean to allow for [+*wh*] information to be 'passed onto' Topic^0, given that [*wh*] marks new

information, and [Topic] old information. Such a semantic clash of feature composition would be expected to bring the percolation of the relevant feature to an end.   (Boeckx 2008: 16)

For another relevant case, consider the analysis of the copula as a head taking a small clause complement (Stowell 1978, Moro 2000), illustrated in (56) below. Here, the lexical copula selects a small clause, an atypical structure resulting from Merge of two full phrases. Small clause structure is a matter of geometry internal to the complement of the copula, hence plausibly another case of long-distance selection in the relevant sense.

(56)   copP
            /   \
      copula   SC
                    /   \
                  XP   YP
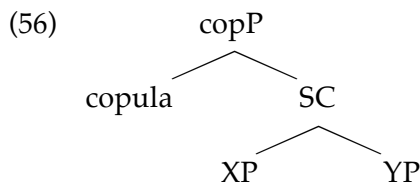
The conclusion seems to be that lexical requirements can enforce structural details down the tree as well as up. This is one reason for carefully separating a notion of 'phrase' tied to projection from a notion involving structure. Surely we do not want to say that portions of the lower clause in the Hebrew example above are 'projections' of the selecting verb, nor that the small clause is a 'projection' of the copula. The point is simply that such an element must occur in a characteristic structure with 'deeper roots', so to speak: Its lexical requirements have structuralization effects that reach down the tree. It should be clear that if such items were the rule, syntactic structures would be drawn from a different set of tree-forms than those described by a projective structure like the X-bar pattern. This only deepens the mystery of why real linguistic structures should tend to adhere to something like the X-bar scheme.

5.3. *Antisymmetry and Teleological Reasoning*

One important theory which has received little mention throughout is the anti-symmetry approach to linear order following from Kayne (1994). Kayne proposes that linear order is not a primitive relation of the syntactic component, but rather a consequence of certain structural properties. Specifically, he proposes that linear order follows from asymmetric c-command, and derives from this the result that only phrase structures obeying a particularly rigid X-bar format are linearizable.

In a sense, then, Kayne's work may be seen as deriving X-bar shape in terms of PF interface requirements. If one adopts such a view, the concerns explored in this article may seem superfluous, in that X-bar-like structure is 'over-determined' both by interface requirements (antisymmetry) and by minimization of c-command and containment computations (in the present account). But Chomsky has repeatedly emphasized that such a redundancy of explanation should be taken to indicate that some further theoretical reduction is required. Put another way, why would we want to explain the relevant facts in

another way, if antisymmetry already does the job?

Let us examine the antisymmetry account in more detail. One crucial point that must be faced by any version of antisymmetry is that pure geometry as such is insufficient to linearize the result of Merging two complex objects. In Kayne's original work, this problem was resolved using the 'segment/category' distinction proposed by May (1985). That is, when objects XP and YP merge, the result will be, say, YP; in this case, the lower YP is then a 'segment' of the full YP, and does not 'count' for the linearization rule, which is restated in terms of full categories. Notice that this requires both a device of projection (to distinguish whether Merge of XP and YP is an XP or a YP), and some explicit notion of bar-level. Without the latter to distinguish Y^0 and YP, the head Y^0 of YP would be a further segment (with Y') of the larger category (YP), with the undesired result that the X-bar configuration would be an unlinearizable multiple-adjunction structure.

Chomsky's (1995a) Bare Phrase Structure Theory incorporates a similar linearization scheme that improves on this somewhat, in effect recreating the segment/category distinction by insisting that intermediate categories, as neither maximal nor minimal projections, are 'invisible to the computation', a claim which seems empirically supported at least. Nevertheless, projection is still integral to the system, and must be explicitly represented by some device that is visible to the PF component.

In either Kayne's or Chomsky's theory, projection is required for PF demands at least. Without such a device, language would fail to be 'usable' by PF; thus one might argue that antisymmetry could explain projection (and the projective X-bar structure) in terms of requirements imposed by the interface. Why then should any alternative explanation for such properties be counte-nanced?

I would like to argue that, if we wish to 'explain' syntax naturalistically, we should be suspicious of teleological reasoning of this sort (see especially Hinzen 2006 on this point). That is, supposing that interface conditions — what syntax is 'good for' — explain the mechanisms that syntax has at its disposal is problematic; in a sense, it amounts to a denial of the autonomy of the syntax. Rather, the preferred mode of minimalist explanation is (or ought to be) to explain syntactic facts in terms of concerns internal to the syntax itself. This is the sort of intuition expressed by Uriagereka's notion that it is "as if syntax carved the path that interpretation must blindly follow" (Uriagereka 2002: 64). Thus, whatever the functional necessity of projection, it is something we would like to derive rather than stipulate.

It may be 'good' for language to have a mechanism for projection. But so what? Language, without such a mechanism, would be whatever it was (perhaps unusable, or at least unpronounceable). As Darwin was careful to point out, the 'desire' to fulfill a certain function does not induce internal complexity. This is not the sort of explanation we should be satisfied within the biolinguistic enterprise. It may be 'good' for tigers to have stripes, so that they may be more effectively camouflaged in tall grass, but that does not *cause* them to have stripes. Likewise, the usefulness of flight does not explain how certain creatures come to have the necessary anatomy in the first place. In the case of projection, labels may

be good things to have, but where do they come from?

These concerns are all the more pressing given the biolinguistic perspective, and particularly the rejection of adaptationist accounts. The preferred mode of explanation, if it can be achieved, is to show that properties of language are not just examples of good design, but of *minimal* design as well. Insofar as language seems to have properties that are 'custom-made' for its eventual function, we may feel we have explained more, and in a more satisfying way, if we find that those properties 'emerge' from the optimal functioning of more primitive components. Suppose we have two kinds of accounts for a property like projection. One account appeals to the use to which projection is eventually put to explain why it exists in the first place, in terms of the interpretation (e.g., linearization) of syntactic objects. An alternative account purports to explain projection in an internalist, autonomous way which makes no reference to the eventual use to which language may be put, instead reflecting optimality in 'bare' combinatorics. Then the second account is more of what we are looking for, so to speak. We would prefer to find, not that projection is a principled complication whose mechanism we must stipulate as a primitive, but rather that the mechanism itself is an example of 'order for free', expected to emerge from the optimal operation of more basic components of the system.

5.4.   *'Emergent' Projection?*

I would like to suggest that this should lead us to rethink the nature of projection in the grammar. Indeed, that concept has a problematic status under current understanding. Simply put, the 'technology' of projection seems to require something non-minimalist, such as assuming that Merge is fundamentally asymmetric, or that Merge necessarily includes a labeling function, both *prima facie* departures from the virtual conceptual necessity of truly 'bare' sets of lexical items. Recognizing this problem, Chomsky suggests that projection is not a primitive notion of syntactic theory, but is to be explained in some way:

> It seems now that much of the architecture that has been postulated can be eliminated without loss, often with empirical gain. That includes the last residues of phrase structure grammar, including the notion of projection or later 'labeling', the latter perhaps eliminable in terms of minimal search.
>
> (Chomsky 2007b: 24)

If the ubiquity of c-command relations in linguistic phenomena reflects a search process (as discussed in section 2), then explaining projection in terms of minimal search is, in fact, exactly what I have proposed. More precisely, what is explained here is why structural correlates of projection are expected in the products of a dynamically optimal derivation. But what is left curiously hanging is the idea of labeling itself: No specific mechanism for enforcing the association between a phrase and its head is motivated. This may well be a good result, given the problems surrounding the technical implementation of labels pointed out by Collins (2002) and others.

In the quote above, it seems that Chomsky has in mind 'eliminating' labeling as a matter of notation explicitly reified by some device in phrase

structure (for example, by complicating the set structure produced by Merge to directly encode the label, as in Chomsky 1995a). But the essential notion remains as a derived fact: Some designated element is required to be readily accessible to determine appropriate interpretation. In the present proposal, a more radical reduction is on offer: The pattern which gives a special place to some designated terminal is independently derived, an accident of optimal branching form. I would like to suggest that this might amount to a deeper explanation; the fact 'emerges' for naturalistic reasons internal to the workings of the computation itself. This is one way of cashing out Uriagereka's notion that it is "as if syntax carved the path that interpretation must blindly follow" (Uriagereka 2002: 64).

6.   Conclusions

The Minimalist Program is concerned with the degree to which the abstract mental system that generates syntactic expressions is 'perfect'; that is, as simple and optimal as it could be. This is often cast as the search for explanation in terms of 'virtual conceptual necessity'. But another important facet of minimalist theorizing is that (internal) optimality is also important. This is precisely the nature of 'emergence': Sometimes, very simple systems behaving optimally give rise to complicated structure. In other words, while superficial complexity may seem problematic from the point of view of the minimalist expectation of perfect simplicity, sometimes complex structure is the most perfect solution.

I have argued that the property of 'projection' might be explained in this way. That is, rather than supposing that projection is strictly required for the linguistic system to function at all (a teleological concern which says nothing about where the instantiating mechanism might come from), I argue that a structural basis for projection might emerge from the optimization of unlabeled branching forms. If we suppose that Merge may apply freely, the full spectrum of binary-branching forms is available in principle. But I have argued that there is a computational burden associated with establishing relations based on c-command and containment, such that some derivational choices are better than others. Taking this claim together with the idea that the information which can influence derivational choices is rather local in character, it stands to reason that a syntactic derivation will face the same 'problem' repeatedly, and thus that it might consistently apply the same solution, in the form of a self-similar pattern of recursion. It turns out that the best such patterns correspond to exactly the sort of structures described as 'projections'.

Moreover, there may be reasons for singling out the X-bar pattern of specifier-head-complement from among these projective patterns. The X-bar form has played an important role in linguistic theory for several decades. Here, this pattern has been shown to have properties related to the Fibonacci sequence, a mathematical pattern which pervades nature. It is not much of an exaggeration to claim that patterns related to the Fibonacci sequence are nature's 'favorite solution' to problems of self-similar growth. Of great relevance to the biolinguistic enterprise is the robust, unselected nature of the pattern: Although it is an optimal solution to certain problems, it is apparently not produced by

successive approximations under evolutionary 'tinkering', but emerges robustly and spontaneously from quite general laws of form that shape the inorganic world as well (see Douady & Couder 1992, Thompson 1917/1992, Ball 1999).

I have tried to demonstrate the potential value of considerations orthogonal to another trend in minimalism, the general program of reducing syntactic properties to lexical requirements. One way of understanding the ideas here is as potentially underlying some otherwise mysterious lexical properties, while still maintaining that featural requirements are the mechanism which drives derivational choices. In its strongest form, this proposal could also be taken to indicate a more direct role for hierarchical optimization in determining syntactic forms. In that case, linguistic computation takes on the appearance of a dynamically self-organizing system, and the explanatory burden placed on features and interface requirements is reduced.

This article is one attempt at explaining substantive properties of language in terms of efficient computation and 'laws of form'. This has only been achieved by way of considerable idealization and abstraction; surely the present approach has a long way to go before approaching anything like the rigorous empirical standard to which linguistic research is usually held. It is not clear that detailed predictions are even possible at the level of abstraction here, and it may turn out that nothing more than intriguing analogies will follow from taking it seriously. Even if the specific ideas here prove to be misguided, I hope that the article may at least suggest some new avenues toward deeper explanation of the sort invited by the biolinguistic perspective.

Appendix: Proof of the Optimality of Projective Structure

Take a recursive pattern P to be defined as above over terminal type 0, non-terminal types 1, … n, with properties of *determinacy* (every non-terminal i branches according to a unique rule i $\rightarrow$ j k, with j, k in 0, 1, … n), and *termination* (no non-terminal dominates only non-terminals 'all the way down').

The reasoning here will involve the infinite tree-space T generated by maximal iteration of a recursive pattern P. In such trees, every non-terminal node in the recursive pattern will be recursively expanded, and the non-terminals thus introduced will be expanded, and so on 'all the way down'.

Now, we may consider mapping nodes in the tree-space T_1 generated by one pattern P_1 to nodes in the tree-space T_2 generated by another pattern P_2. The idea is to find immediate-containment-preserving maps of sets of nodes in T_1 to sets of nodes in T_2 such that:

(A1) the image of the root node of T_1 is the root node of T_2, and
(A2) if node α immediately contains node β in T_1, the image of α immediately contains the image of β in T_2.

Let us say that T_2 *contains* T_1 if there is some mapping of the set of all nodes

in T_1 into nodes of T_2 meeting this condition, and that T_2 *properly contains* T_1 if T_2 *contains* T_1 but T_1 does not *contain* T_2. (If T_1 *contains* T_2 and T_2 *contains* T_1, then T_1 and T_2 are isomorphic, and so are P_1 and P_2.)

We will also consider finite trees within these infinite trees, i.e. *contained* by them in the sense above. For notational clarity, we reserve T_i for infinite tree-spaces generated by maximal expansion of P_i. Clearly, if T_1 *properly contains* T_2, every finite tree generable by P_2 can be generated by P_1.

We are interested in comparing the optimality, with respect to number of c-command and containment relations, of best finite trees (with equal numbers of nodes) generated by distinct recursive patterns P_1, P_2. At the very least, if every arrangement possible under P_2 is also possible under P_1, but there are arrangements generated by P_1 more optimal than any arrangement of the same number of nodes under P_2, we will judge P_1 to be more optimal than P_2.

(A3) *Lemma 1*
     If T_1 *properly contains* T_2, P_1 is more optimal than P_2.

Clearly, every finite tree generable by P_2 can be generated by P_1. For proper containment to hold, T_1 cannot be mapped to T_2. The mapping from T_1 to T_2 fails first at some finite depth d (succeeding at all depths less than d); the maximal finite trees in T_1 and T_2 can be mapped to the other up to depth d–1.

For the mapping to fail, T_1 must have one or more non-terminals at depth d–1 that map to one or more terminals at the same depth in T_2. Then consider the maximal finite tree in T_1 of depth d (all recursive options expanded to depth d, all non-terminals in T_1 at depth d replaced with terminals). This tree has fewer c-command and containment relations than any tree in T_2 with the same number of terminals. One or more of the non-terminals at depth d–1 that were expanded in T_1 must terminate at that depth in T_2. Then some number of nodes in T_1 at depth d cannot be mapped to corresponding nodes in T_2 at the same level, and the same number of nodes must appear at depth d+1 or greater in T_2; all other nodes correspond. Since the number of c-command and containment relations induced by a node is equal to its depth in the tree, it follows that any tree in T_2 containing the same number of nodes as the maximal finite tree of depth d in T_1 must have strictly more c-command and containment relations.

Thus, if T_1 properly contains T_2, P_1 is more optimal than P_2: Every arrangement possible under P_2 is also possible under P_1, but there are arrangements generated by P_1 superior to any arrangement of the same number of nodes under P_2.

(A4) *Lemma 2*
     The infinite tree space T_p generated by the projective recursive pattern P_p defined over some number *n* of non-terminal types *properly contains* all tree-spaces T_i generated by distinct recursive patterns P_i defined over the same number of non-terminal types.

To see this, we will need one more concept, that of 'least path-to-terminal'. A 'path' leading from node α to node β is the set of nodes containing α, β, and all

nodes dominating β which are also dominated by α. For any non-terminal node in a tree, we can identify the paths of nodes leading to terminals it dominates, and measure the depth of those paths. Among these paths, there will be one or more least paths-to-terminals (clearly, of depth at most n, for n non-terminal types). Let us consider these paths under the sort of mapping described above.

First, in T_p, the least path-to-terminal from the root node has length n. Let us call an 'off-branch' from this path a sub-tree whose root node is immediately dominated by a node on the path, but is not on the path itself. In T_p, the least path-to-terminal from the root of any off-branch is itself of length n (since any off-branch is isomorphic to the root node).

Now suppose T_i is a tree-space distinct from T_p defined over the same number of n non-terminal types. First, T_p contains T_i. For this to be false, there must be some finite depth d at which the mapping first fails. Find the shortest path-to-terminal from the root in T_i (or select one of them, if there are several of the same shortest length). Let us map the nodes in this path to nodes in the least path-to-terminal in T_p. This mapping succeeds, because this path is of depth at most n, and the path-to-terminal in T_p is of depth n. Now, for each off-branch from the path in T_i, we can map a least path-to-terminal successfully to the least path-to-terminal on the corresponding off-branch in T_p, which again is of the greatest possible depth n. And so on, for off-branches of off-branches; this exhausts the set of nodes in T_i, since (due to the termination requirement) every non-terminal lies on some least path-to-terminal. Thus, T_p contains T_i.

It cannot be the case that T_i contains T_p, because we have supposed that T_p and T_i are distinct. Thus, T_p properly contains T_i.

Then from Lemma 1 and Lemma 2, P_p is more optimal than P_i; since P_i was an arbitrary recursive pattern distinct from P_p defined over the same number of non-terminal types, we conclude that the projective pattern is the most optimal.

References

Ball, Phillip. 1999. *The Self-Made Tapestry: Pattern Formation in Nature*. Oxford: Oxford University Press.

Boeckx, Cedric. 2008. *Bare Syntax*. Oxford: Oxford University Press.

Boeckx, Cedric & Massimo Piattelli–Palmarini. 2005. Language as a natural object; linguistics as a natural science. *The Linguistic Review* 22, 447-466.

Bouchard, Denis. 1995. *The Semantics of Syntax: A Minimalist Approach to Grammar*. Chicago: University of Chicago Press.

Bresnan, Joan (ed.). 1982. *The Mental Representation of Grammatical Relations* (Cognitive Theory and Mental Representation). Cambridge, MA: MIT Press.

Carnie, Andrew & David Medeiros. 2005. Tree maximization and the Extended Projection Principle. *Coyote Working Papers in Linguistics* 14, 51-55.

Chametzky, Robert. 1996. *A Theory of Phrase Markers and the Extended Base*. Albany, NY: SUNY Press.

Chametzky, Robert. 2000. *Phrase Structure: From GB to Minimalism* (Generative Syntax 4). Malden, MA: Blackwell.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*, 184-221. Waltham, MA: Ginn and Co.

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris.

Chomsky, Noam. 1995a. Bare phrase structure. In Gert Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program* (Generative Syntax 1), 385-439. Malden, MA: Blackwell.

Chomsky, Noam. 1995b. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89-155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language* (Current Studies in Linguistics 36), 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1-22.

Chomsky, Noam. 2007a. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax–Semantics* (Studies in Generative Grammar 89), 1-29. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2007b. Of minds and language. *Biolinguistics* 1, 9-27.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic Perspective* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Collins, Chris. 2002. Eliminating labels. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program* (Generative Syntax 6), 42-64. Malden, MA: Blackwell.

Douady, Stephane & Yves Couder. 1992. Phyllotaxis as a physical self-organized process. *Physical Review Letters* 68, 2098-2101.

Epstein, Samuel David, Erich Groat, Ruriko Kawashima & Hisatsugu Kitahara. 1998. *A Derivational Approach to Syntactic Relations*. Oxford: Oxford University Press.

Freidin, Robert & Jean-Roger Vergnaud. 2001. Exquisite connections: Some remarks on the evolution of linguistic theory. *Lingua* 111, 637-666.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Grimshaw, Jane. 1991. Extended projection. Ms., Brandeis University.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An Essay on Stress* (Current Studies in Linguistics 15). Cambridge, MA: MIT Press.

Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press.

Hinzen, Wolfram. 2006. *Mind Design and Minimal Syntax*. Oxford: Oxford University Press.

Hornstein, Norbert & Jairo Nunes. 2008. Adjunction, Labeling, and Bare Phrase Structure. *Biolinguistics* 2, 57-86.

Jackendoff, Ray. 1977. *X'-Syntax: A Study of Phrase Structure* (Linguistic Inquiry Monographs 2). Cambridge, MA: MIT Press.

Kayne, Richard S. 1984. Unambiguous paths. In Richard S. Kayne, *Connectedness and Binary Branching*, 129-163. Dordrecht: Foris.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax* (Linguistic Inquiry Monographs 25). Cambridge, MA: MIT Press.

Kornai, Andras & Geoffrey Pullum. 1990. The X-bar theory of phrase structure. *Language* 66, 24-50.

May, Robert. 1985. *Logical Form* (Linguistic Inquiry Monographs 12). Cambridge, MA: MIT Press.

Moro, Andrea. 2000. *Dynamic Antisymmetry* (Linguistic Inquiry Monographs 38). Cambridge, MA: MIT Press.

Pietroski, Paul. 2004. *Events and Semantic Architecture*. Oxford: Oxford University Press.

Pollard, Carl & Ivan Sag. 1987. *Information-Based Syntax and Semantics*, vol. 1 (CSLI Lecture Notes 13). Stanford: CSLI Publications.

Pollard, Carl & Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Reinhart, Tanya. 1976. *The Syntactic Domain for Anaphora*. Cambridge, MA: MIT dissertation.

Rizzi, Luigi. 1990. *Relativized Minimality* (Linguistic Inquiry Monographs 16). Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Lilliane Haegeman (ed.), *Elements of Grammar: Handbook of Generative Syntax* (Kluwer International Handbooks of Linguistics 1), 281-337. Dordrecht: Kluwer.

Speas, Margaret. 1990. *Phrase Structure in Natural Language* (Studies in Natural Language & Linguistic Theory 21). Dordrecht: Kluwer.

Stowell, Tim. 1978. What was there before there was there? In Donka Farkas, Wesley Jacobson & Karol W. Todrys (eds.), *Papers from the Fourteenth Regional Meeting of the Chicago Linguistic Society*, 458-471.

Stowell, Tim. 1981. *Origins of Phrase Structure*. Cambridge, MA: MIT dissertation.

Thompson, D'Arcy Wentworth. 1917/1992. *On Growth and Form* [abridged edn., prepared by John Tyler Bonner]. Cambridge: Cambridge University Press.

Uriagereka, Juan. 1998. *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.

Uriagereka, Juan. 2002. *Derivations: Exploring the Dynamics of Syntax* (Routledge Leading Linguists). London: Routledge.

Williams, Edwin. 1994. *Thematic Structure in Syntax* (Linguistic Inquiry Monographs 23). Cambridge, MA: MIT Press.

David P. Medeiros
University of Arizona
Department of Linguistics
P.O. Box 210028
Tucson, AZ 85721
USA
medeiros@email.arizona.edu

On the Nature of Syntax

Alona Soschen

There is a tendency in science to proceed from descriptive methods towards an adequate explanatory theory and then move beyond its conclusions. Our purpose is to discover the concepts of computational efficiency in natural language that exclude redundancy, and to investigate how these relate to more general principles. By developing the idea that linguistic structures possess the features of other biological systems this article focuses on the third factor that enters into the growth of language in the individual. It is suggested that the core principles of grammar can be observed in nature itself. The Faculty of Language is an efficient mechanism designed for the continuation of movement in compliance with optimization requirements. To illustrate that, a functional explanation of syntactic Merge is offered in this work, and an attempt is made to identify some criteria that single out this particular computational system as species-specific.

1. Introduction: Natural Law and Syntactic Trees

Alongside the other two important factors — genetic endowment and experience — a third factor is particularly important to our discussion. According to Chomsky (2008), it includes the objective principles of architecture that restrict outcomes determining attainable languages. We will follow the minimalist research program in seeking to identify aspects of language that are determined by the properties of natural phenomena. At this point, to advance our understanding of the common properties of human language, we need to present further proof of the advantages which would arise from the application of

physical laws to the analysis of syntactic structures.

Syntax is viewed in this article as a unique subtype of recursive systems designed for the continuation of movement. The Faculty of Language (FL) in the broad sense (FLB) includes a sensorimotor system, a conceptual-intentional system, and the computational mechanisms for recursion. If we accept the hypothesis that FL in the narrow sense (FLN) includes only recursion, the ideas offered in this article may help to explain what basic operations underlie FLN.

Natural Law (N-Law), a physical phenomenon exemplified as the Fibonacci patterns where each new term is the sum of the two that precede it, can be observed in language, just as it is in other mental representations (Uriagereka 1998, Carnie *et al*. 2005, Soschen 2006).[1] These structures share certain remarkable properties with the linguistic system, according to minimalism: Both of them are characterized by discreteness and economy.[2] Based on that, it will be shown that the same condition accounts for the essential properties of syntactic trees: binarity of branching, the mechanism of labeling, and the properties of Merge. First, the article provides a functional explanation of thematic domains and phase formation, on the example of applicative constructions. Second, it offers a principled account of label-free parallelism of phases across languages by presenting a short discussion of the Exceptional Case Marking structures. The analysis derives the types of cross-linguistically available argument representations, and explains the attested relative frequencies of various basic word order patterns.

In the present system, syntactic composition is in effect reduced to conjunction, or Merge, of two elements without asymmetry, thus eliminating the X'-level of representation. *Conjunctivism* achieves a remarkable degree of simplicity for Occam's Razor-like methodological reasons. As it is further developed to handle an increasingly broad range of constructions and theoretical considerations, it will inevitably become more complex.

The *Fibonacci sequence* (henceforth, *FS*) is one of the most interesting mathematical curiosities that pervade the natural world. These numbers are evident in every living organism. For example, they appear in the spiral shapes of seashells and in the arrangement of leaves, petals, and branches of trees:
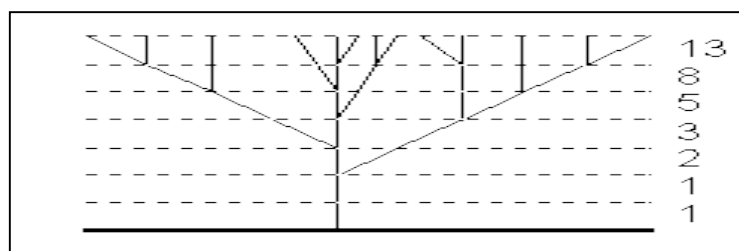

Figure 1: Fibonacci Numbers in a Tree

---

[1]   The number of 'growing points' in plants corresponds to the Fibonacci Sequence: $X(n) = X(n–1) + X(n–2)$: {0, 1, 1, 2, 3, 5, 8, 13, …}. The limit ratio between the terms is .618034…, the Golden Ratio.
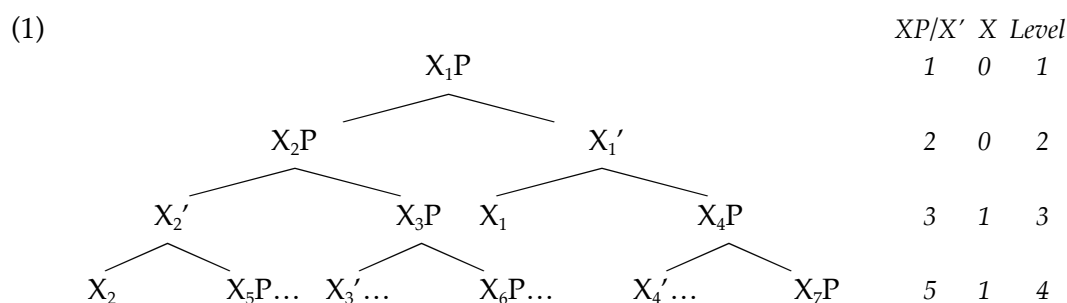
[2]   For a good overview of the roots and core ideas of minimalism, see Boeckx (2006).

Early approaches to FS in nature were purely descriptive with a focus on the geometry of patterns. Later, Douady & Couder (1992) developed a theory of plant growth (phyllotaxis), which explained the observed phenomenon as following from efficient space filling. A particular pattern related to maximizing space is important in the case of closely-packed leaves and branches, because it ensures maximum exposure to the sun. This system is based on simple dynamics that impose constraints on the number and order of constituents to satisfy optimal conditions. Successive elements of a certain kind form at equally-spaced intervals of time on the edge of a small circle, representing the apex. These elements repel each other (similar to electric charges) and migrate in a radial manner at some specified initial velocity. As a result, motion continues and each new element appears as far as possible from its immediate successors. In humans, the Golden Ratio appears in the geometry of DNA and physiology of the head and body. On a cellular level, the '13' (5+8) Fib-number present in the structure of microtubules (cytoskeletons and conveyer belts inside the cells) may be useful in signal transmission and processing. The brain and nervous systems have the same type of cellular building units, so the response curve of the central nervous system may also have FS at its base. This suggests a strong possibility that N-Law or general physical laws that ensure efficient growth apply to the universal principles that govern linguistic representations as well.

As has already been mentioned, it was confirmed recently that syntactic structures exhibit certain mathematical properties. Like other systems that comply with N-Law, tree structures are maximized in such a way that they result in a sequence of categories that corresponds to FS. The syntactic tree is generated by merging two elements; the next operation adds a newly introduced element to the already formed pair. Each item is merged only once; every subject/specifier and every object/complement position is filled. In the traditional sense of Chomskyan X-bar theory, a label immediately dominated by the projection of another category is an XP (phrase).[3] Other non-terminal nodes are annotated as X', and Xs are 'heads'. If $XP(n)$ is the number of XPs in the nth level, then $XP(n) =$ $Fib(n)$. This property is true of all trees that are maximized by having specifiers and complements filled.

In (1) below, one can see that N-Law provides an external motivation for Merge to distinguish between syntactic labels in a particular way. Determining whether a node is XP or X follows directly from the functional pressure of cyclic derivation. The Fib-based system distinguishes between *sums of terms* (XP and X') and *single terms* (X), rather than between either XP/X' or X'/X: Level 2 has one XP and one X', Level 3 has one X' and one X. The assumption that syntactic structures have an intermediate X' projection does not hold in the present system: Basic representations appear to be monadic — cf. the dyadic model of X-bar theory, for example; see also Collins (2002) on the elimination of labels.

---

[3]   The Fibonacci sequence in a tree is related to the fact that each node dominates exactly one maximal projection. Thanks to Hans Broekhuis (p.c.) for pointing this out. Possibly, hierarchical structures created by adjunction (pair-Merge, in the Chomskyan system) comply with NL as well. Rubin (2003) proposes the (obligatory) existence of a functional category, Mod, in the structure of adjuncts ([Mod [[YP] Adjunct]]) that is parallel in nature to functional categories in clauses.

(1)

						XP/X'	*X*	*Level*
			X$_1$P			*1*	*0*	*1*
	X$_2$P			X$_1$'		*2*	*0*	*2*
X$_2$'		X$_3$P X$_1$			X$_4$P	*3*	*1*	*3*
X$_2$ X$_5$P…		X$_3$'…	X$_6$P…	X$_4$'…	X$_7$P	*5*	*1*	*4*

What is the reason behind compositionality that motivates combining exactly two terms in a set? The requirement to achieve tree maximization explains why the trees are constructed out of binary units. If Merge were allowed to optionally select three terms and combine them into a ternary structure, then FS of maximal categories would disappear. The sequence where each term A_n combines with the two that precede it is {1, 1, 1, 3, 5, 9, 17, 31, 57, …}. The ternary branching system shows a Fib-like sequence; however, the arrangement of elements displays a ratio different from the Golden Ratio, which fails to meet the condition of optimization. As a result, ternary branching or any operation that merges more than two syntactic elements is disallowed.[4]

The requirement to fill specifier and complement positions faces a problem: It creates a 'bottomless' tree by eliminating a line with only terminal Xs (Carnie 2002). However, real sentences always have an ending point. In the present work, the solution to this problem lies in redefining syntactic binarity to include zero-branching — in other words, to start FS with 0 instead of 1. This follows directly from the requirement of N-Law: Each successive element is combined with a sum of already merged elements, not with one. For example, merging 2 with 1, where 1 is a sum of 1 and 0, yields a new element 3, while merging two elements one of which is not a sum (2+0) does not. Consequently, (2a) and (2b) are instances of Merge, while (2c) is not.

(2)   a.   XP   b.   X/Y   c.   X
                  /\   /\   /
              X   Ø   X   Y   X
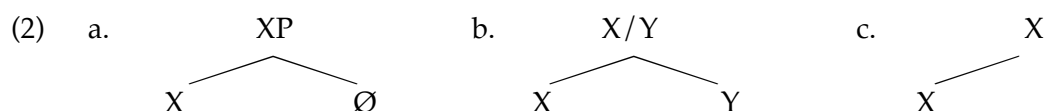
When the sum of terms is present at each step, it provides the 'bottom line' in the syntactic tree. The newly introduced *zero-Merge* (Ø-Merge) distinguishes between terms {1}/X and singleton sets {1, 0}/XP. This way the process of merging *terms* with *sets* is initiated, to ensure continuation of motion. Following from that, *singleton sets are indispensable for recursion*.

The suggestion to regard an empty element as functional in Merge has serious consequences for the theory of binary branching. The minimal building block that enters into linguistic computation is re-evaluated to include Ø-Merge, and is identified as the product of Ø-Merge.[5] As a result, binarity is preserved,

---

[4]   Chomsky (2007a: 8) asserts that "Merge cannot create objects in which some object W is shared by the merged elements X, Y. It has been argued that such objects exist. If so, that is a departure from SMT, hence a complication of UG".

[5]   For the discussion of zero-branching constructions, see, e.g., Roodenburg (2004).

while there is no problem caused by the requirement to fill specifier and complement positions. XPs and Xs are disambiguated, which eliminates the necessity to proceed with further branching below the bottom level.

Furthermore, the proposed analysis along the lines of N-Law clarifies the notion of labeling, and answers the question why labels can be disposed of in syntax. If the same element can be represented as either a singleton set or a term, it follows that X and XP are not syntactic primitives.[6] The idea that constituent structures are labeled appears to be a stipulation; this part of Merge should be abandoned in favor of a rule that offers a more adequate explanation.[7] As grammar evolves toward a generalized syntactic representation, the only necessary mechanism is not the one that determines which node is XP and which is X or X', but the one that determines whether a node is a result of Merge or not, thus eliminating labels altogether.

In sum, in the present system,

- a bottom node is identified as either XP or X, depending on whether or not it undergoes Ø-Merge;
- a node is identified as either XP or X, depending on whether or not it is the result of Merge.

2.   Merge and Displacement

2.1.   *Constraints on External Merge*

Syntactic Merge builds elementary trees and combines them into larger structures. Under External Merge (henceforth, EM), α and β are separate objects; under Internal Merge, one is part of the other, and Merge yields the property of displacement (Chomsky 2001).[8] The argument structure is the product of EM. The application of Fib-like logic to the analysis of thematic domains makes some

---

[6]   Heads can behave like phrases and *vice versa*, according to Carnie (2000), Collins (2002), and Chomsky (2004, 2008). There exist numerous instance of label-switching between X and XP: *that* may behave as X and XP in the same sentence (i).

    (i) $_{XP}$That $_X$that is, is; $_{XP}$that $_X$that is not, is not — we all know $_{XP}$that.

In addition, a group of Russian nouns (*toska* 'boredom', *grex* 'sin', *vremja*, *pora* 'time', etc.) can be either predicate heads (ii) or arguments (iii).

    (ii)   Vam $_X$grex   žalovat'sja.
                you.DAT *sin*   *complain.INF*
                lit. 'For you a sin to complain.'

    (iii) $_{XP}$Grex budet iskupljon.
                sin   *will.be redeemed*
   'The sin will be redeemed.'

[7]   "It seems now that much of the architecture that has been postulated can be eliminated without loss, often with empirical gain" (Chomsky 2007b: 24).

[8]   The pressure for the tree to be maximized justifies the basic principle of organization in both types of Merge. Move is just one of the forms of Merge: EM induces IM by virtue of the fact that already conjoined elements have to be *linearized* at the level relevant for pronunciation.
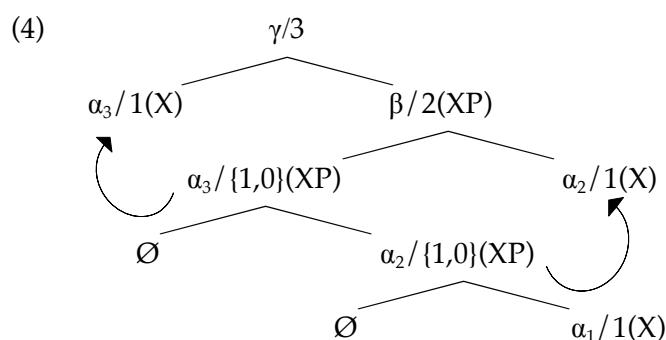
interesting predictions about the constraints on EM, such as a fixed number of nodes (1, 2, and 3) in these domains.

Assume that Ø-Merge is the operation responsible for constructing elementary *argument-centered* representations, the process that takes place prior to lexical selection.[9] As already pointed out, this kind of Merge is relevant at the point where a distinction between *terms/entities* — represented as *{1}/X* — and *sets* — or *{1, 0}/XP* in the present system — is made.

The functional pressure of cyclic derivation to merge terms of *different types only* accounts for the *type-shift*, or type-lowering, from *sets* to *entities* at each level in the tree. As a result, at some level, a node is XP (set); at the next level, it is X (entity). The Impenetrability Condition ensures the *continuity* (vs. discreteness) of constituents: Once X is formed, it cannot be broken up into parts. To clarify this point, (3) shows an example of a type-shifting operation in an FS-based numeric system. At the point where 3 is merged with 2, element 3 is the sum of 1 and 2 (set {1, 2}, XP), but 2 is a single entity ([2], X).

(3)

$$
\begin{array}{c}
5 \\
3/\{1, 2\} \qquad [2] \\
1 \qquad 2/\{1, 1\} \quad \textit{type-shift}
\end{array}
$$
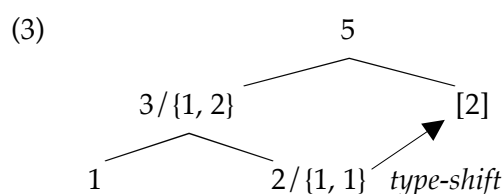
Assume that in a tree built by EM in compliance with N-law, the recursively applied rule adjoins (in a bottom-up manner) each element to the one that has a higher ranking, starting with the term that is 'Ø-merged first'. Recall that in the present system, FS starts with 0: {0, 1, 1, 2, 3, 5, …}. In (4), α_1 is entity, α_2 and α_3 are singleton sets, and β and γ are non-empty (non-singleton) sets. At each level, the Impenetrability Condition induces a type-shifting operation from *sets* to *entities*. The type of α_2 is shifted from *singleton set* (XP) to *entity* (X), to be merged with α_3 (XP); the type of α_3 is shifted from *singleton set* (XP) to *entity* (X) and merged with β (XP).

(4)

$$
\begin{array}{c}
\gamma/3 \\
\alpha_3/1(X) \qquad \beta/2(XP) \\
\alpha_3/\{1,0\}(XP) \qquad \alpha_2/1(X) \\
\emptyset \qquad \alpha_2/\{1,0\}(XP) \\
\emptyset \qquad \alpha_1/1(X)
\end{array}
$$

There is a limited array of possibilities for EM, depending on the number of

---

[9]   Chomsky (2007a) specifies other argument constructs, such as Pritchett's (1992) theta-driven model of perception. In such and similar models, a verb is theta-role assigner. In a proposed *primitive* model of EM, the only function that matters is the one that identifies *arguments*.

positions available to a term adjoining to the tree. This operation either returns the same value as its input (Ø-Merge) or the cycle results in a new element (N-Merge).

- Term α_1 can be Ø-merged *ad infinitum* (5a): The function returns the same term as its input and the result are zero-branching structures.
- Ø-merged α_1 is type-shifted to α_2 and N-merged with α_3 (5b): The process creates a single argument position made explicit by intransitive (*unergative* and *unaccusative*) verbs, e.g., in sentences such as *Eve$_1$ laughs* or *The cup$_1$ broke*.[10]

(5) a.

b.
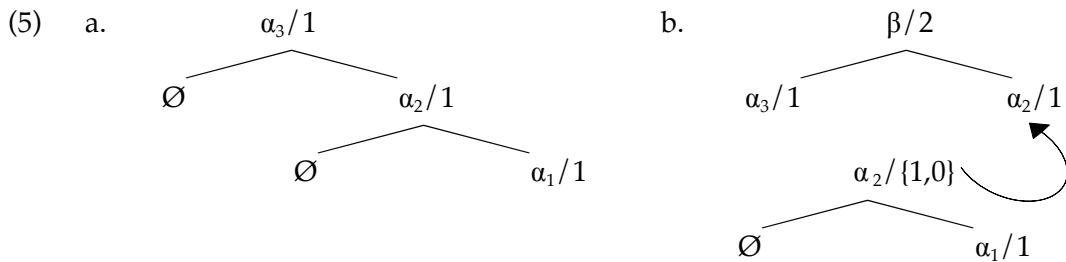
- Terms α_2 and α_3 assume positions where each can be merged with a non-empty entity: The result are two argument positions, e.g., *Eve$_1$ saw Adam$_2$* (6a).[11]
- There are three positions to accommodate term 1 (i–iii): This may explain why the number of arguments permitted is limited to three in maximal thematic domains, represented, by the sentence *Eve$_1$ gave Adam$_2$ an apple$_3$*, for example (6b).

(6) a.

b.

---
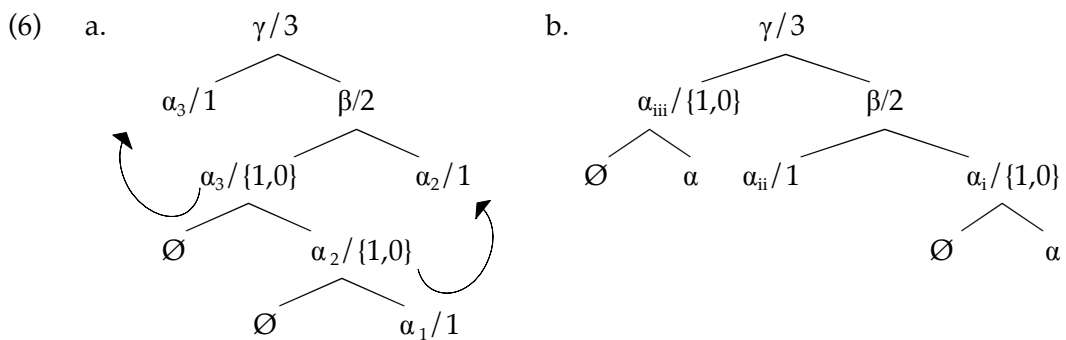
[10] Certain verbs of spatial configuration, such as *lean*, are unergative with an agentive subject but unaccusative when they take a non-agentive subject (Levin & Rappaport Hovav 1995). A term may undergo Ø-Merge either first or second, which explains why the same verb appears with either *agent* (Ø-merged first) or *theme* (Ø-merged second).

[11] The supporting evidence that a term may undergo Ø-Merge either first or second comes from Japanese. In (i), the argument position of *girl* is 'Ø-merged second' in the matrix clause and 'Ø-merged first' in the subordinate clause.

(i) Yoko-ga kodomo-o koosaten-de mikaketa onnanoko-ni koe-o kaketa.
Yoko.NOM child.ACC intersection.LOC saw girl.DAT called
'Yoko called the girl who saw the child at the intersection.' (Pritchett 1992)

We have shown so far that the N-Law logic can be applied to the analysis of EM to account for the limited number of argument positions in thematic domains. The argument structure is built upon hierarchical relations; the term that is Ø-merged first has the highest ranking.[12]
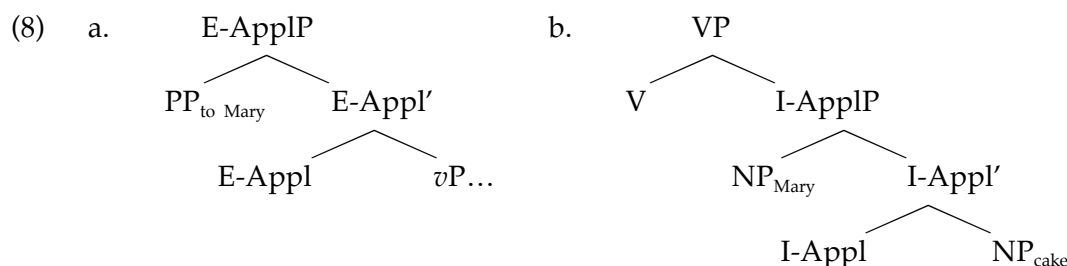
2.2. *Maximal Thematic Domains*

The applicative and double object constructions of the kind *John baked Mary a cake* and *John gave Mary a cake* vs. *to-* and *for*-constructions *John baked a cake for Mary* and *John gave a cake to Mary* have a maximal number of arguments, which is essential for the explanation of limitations imposed on thematic domains.

Recent research on argument structure has resulted in a complex representation that consists of two levels: One involves two individuals, and another expresses an individual-event relation (Marantz 2003, McGinnis 2001, Pylkkänen 2001, 2003). Sentences like *John baked/gave [Mary]$_{individual}$ [a cake]$_{individual}$* are of the first type; other structures, such as *[John baked a cake]$_{event}$ [for Mary]$_{individual}$* or *[John gave a cake]$_{event}$ [to Mary]$_{individual}$*, belong to the second.

It was suggested that a relation between individuals is established by means of Event Applicative, heading an E-ApplP ((7a,b) for (8a)), and by means of Individual Applicative, heading an I-ApplP ((7c) for (8b)).[13]

(7) a.   John gave a cake$_{event}$ to Mary$_{individual}$.
     b.   John baked a cake$_{event}$ for Mary$_{individual}$.
     c.   John baked/gave Mary$_{individual}$ a cake$_{individual}$.

(8) a.   E-ApplP   b.   VP

         PP$_{to\ Mary}$ E-Appl'   V   I-ApplP

                   E-Appl vP…   NP$_{Mary}$ I-Appl'

                                                              I-Appl   NP$_{cake}$

The generalized thematic structure that incorporates both ApplPs is shown in (9), where Y_E is E-Appl and Y_I is I-Appl.

(9) a.   [$_{vP}$ v [$_{E\text{-}ApplP}$ E-Appl [$_{VP}$ V [$_{I\text{-}ApplP}$ I-Appl NP]]]]
     b.   [$_{vP}$ v [$_{Y\text{-}EP}$ Y_E [$_{VP}$ V [$_{Y\text{-}IP}$ Y_I XP]]]]

---

[12]   Hierarchy is assumed to be automatic for recursive operations (Chomsky 2008).
[13]   This classification is viewed as necessary to account for the difference in semantic interpretation. See Erteschik–Shir (1979) and Snyder (2003) on the semantics of the English *to*-dative and double object constructions with *give*. Studies of texts show that there is a preference for the double object construction since recipients are typically human and, therefore, likely to be given, while themes are typically inanimates and, therefore, less likely to be given.

   (i)   a. Nixon's behavior gave Mailer an idea for a book.
          b. *Nixon's behavior gave an idea for a book to Mailer.   (Snyder 2003)

When the trees are maximized and all positions are filled, as in (10), the sum of heads, specifiers, and complements yields a maximal space of 13 — the Fib-number.

(10) a. $[_{XP}\ v\text{P}\ [_{v'}\ v\ [_{XP}\ \text{VP}\ [_{V'}\ \text{V}\ [_{XP}\ Y_I\text{P}\ [_{YI'}\ Y_I\ \text{XP}\ ]]]]]]$
      b. $[_{XP}\ v\text{P}\ [_{v'}\ v\ [_{XP}\ Y_E\text{P}\ [_{Y\text{-}E'}\ Y_E\ [_{XP}\ \text{VP}\ [_{V'}\ \text{V}\ \text{XP}\ ]]]]]]$

In theory, maximal thematic domains may be constructed in a certain way to accommodate all possible *argument* configurations:[14]

(11)
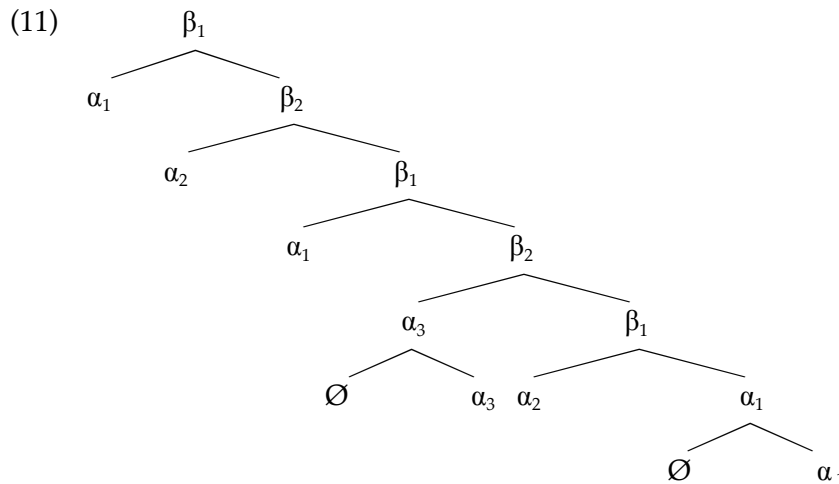
There does not seem to be any intrinsic reason why thematic domains should be spaces with a particular number of nodes — 13. However, from a broader perspective, there is a sense in which the domains under discussion are maximal. As was already pointed out, the Fib-number '13' is present in the structure of microtubules; the brain and nervous systems have the same type of cellular building units. This may account for the limitations imposed on thematic domains — the core units built by syntactic Merge.[15]

3.   Internal Merge

3.1.   (Non-)Propositionality of Phases

The application of Fib-like logic not only makes interesting predictions about the constraints on EM but also explains the properties of Internal Merge (IM), an operation relevant at the point of pronunciation that assigns the order to lexical items. As was already shown, EM creates a hierarchical structure with a restricted number of arguments. It is possible that optimization requirements also justify the principle of organization in IM, a highly efficient mechanism

---

[14]   See section 4 for further elaboration as to why this possibility is strong, and the relevant one.
[15]   Interestingly, in (11) the number of binary chunks is 7, which roughly corresponds to the human short-term memory capacity with an average of 7±2 limit.

designed for the continuation of movement in derivations. In this sense, restructuring is not an imperfection but a necessity to satisfy conditions on the ordering of syntactic elements at the point of pronunciation. The explanation of IM is very straightforward if we assume that derivations proceed by phases, and that movement depends on the qualification of phrases as phases.[16]

Are phases propositional? According to Chomsky (who suggests that vP and CP are phases, while VP and TP are not), the answer is most probably yes. Only a fully-fledged phrase can qualify as a phase. *Bill likes Mary* is possible because there is an additional position x in [Spec,vP] to accommodate the NP *Bill*. This position is projected by the phasal head v in [$_{vP}$ x$_{Bill}$ v [$_{VP}$ likes Mary]]. In contrast, *likes Mary* is not a phase as there is no available position x to accommodate the NP *Bill*; representations of the kind [$_{VP}$ x V NP] are not feasible. As was already discussed, ternary branching or any operation that merges more than two syntactic elements is disallowed in syntax.

The analysis developed in this paper leads one to the conclusion that any XP can in principle head a phase. This idea is based primarily on regarding phases in a particular way. Phases are characterized by their *ability to induce a new cycle* (to ensure continuation of movement) by projecting extra Spec positions, thus providing a 'landing site' for a moved constituent.[17] In this sense VP and TP, for example, may constitute *internal* phases, however incomplete.[18] In the next section, phases are redefined as *maximal* (propositional) and internal, i.e. *minimal* (non-propositional), constructs. Then it is shown that the formation of a *minimal* phase should be regarded as language-specific.

3.2.   *Minimal and Maximal Phases*

A 'derivation-by-phase' approach to applicative and double object constructions constitutes a crucial step toward an explanatory account of phase formation. As previously described, I-Appl establishes a relation between two individuals, while E-Appl is instrumental in expressing a relation between an individual and an event. It was maintained in the above-cited literature that only the relation between individuals and events constitutes a phase, in order to provide an account of passive formation in these constructions. It was also concluded that the absence of an extra Spec-position in I-ApplP, the Individual-Applicative Phrase, disqualifies it from phasehood, by blocking DO-movement, i.e. movement by the direct object. As a result, sentences like *A cake was baked Mary t$_{cake}$* and *A cake was given Mary t$_{cake}$* are unacceptable. At the same time, sentences of the

---

[16]   See Chomsky (1995, 2004, 2007a, 2007b) for the discussion of phase formation. See also Bošković (2002), Epstein & Seely (2002), Legate (2003), Müller (2004), Suranyi (2004), and Wexler (2004).

[17]   Thinking positively, we are interested in what prompts movement, the steps by which it proceeds — and only then considering non-phasal configurations to account for the *barriers* to movement.

[18]   This distinction between *internal* and *complete* phases is analogous to what is found in other natural systems of efficient growth. In Figure 1, the tree constitutes a *complete* stage/phase, while its constituent parts (i.e. a branch, a flower) are *internal* stages in the development of the tree. Meanwhile, both types of phases comply with the optimization requirement imposed by N-Law.

kind *A cake was baked t_{cake} for Mary* and *A cake was given t_{cake} to Mary* are grammatical due to DO-movement of NP_{cake} to [Spec,E-ApplP], which is a phase.[19]

The distinction between the two structures (12a) and (12b) below is in the movement of object to subject position in E-ApplP. This movement is possible because E-Appl projects an extra Spec-position, while I-Appl does not, rendering (12b) ungrammatical. The DO *cake* can raise to the subject position in (12a), but not in (12b).[20] Based on this analysis, the conclusion has been reached that only a propositional (eventive) E-ApplP, but not I-ApplP, constitutes a phase.

(12) a.   A cake was given to Mary / baked for Mary.
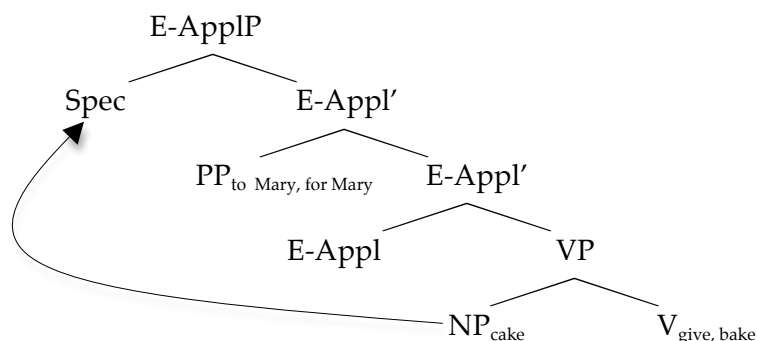
b.   *A cake was given / baked Mary
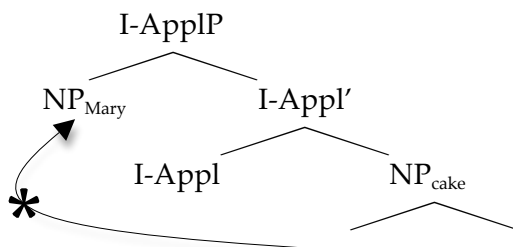
Recently, however, it has been shown that Individual-Applicative Phrases behave like phases in certain languages, by allowing DO-movement and blocking IO-movement in passives, as (13) sketches (Soschen 2005). In synthetic (inflectional) languages, such as Russian, Italian, and Hebrew, I-ApplPs exhibit the properties of minimal (min)-phases. [21]

---

[19]   Note that indirect object or IO-movement is ok: *Mary was given/baked a cake.*

[20]   Move is driven by a need to check a feature (Chomsky 1995, Richards 2001). In passives, direct object moves to [Spec,E-ApplP] to check *uninterpretable* features on a phase head. When the head — in this case, I-Appl — does not have these features, no Spec-position is projected, and movement is blocked.

[21]   As one example, applicative constructions in Kinyarwanda (Bantu) exhibit either indirect (i) or direct (ii) object movement in passives. There is no morphological evidence (i.e. PP) that (ii) involves E-ApplP; the conclusion that E-ApplP is a phase but I-ApplP is not relies solely on object movement.

(13) *Italian, Russian, Hebrew, Kinyarwanda*

[$_{VP}$ V [$_{I\text{-}ApplP}$ DO [$_{I\text{-}ApplP}$ IO [$_{I\text{-}Appl'}$ I-Appl t_{DO}]]]]

I-ApplP: minimal phase

In contrast, in analytical languages I-ApplP is not a phase but *v*P is:[22]

(14) *English, Icelandic*

[$_{vP}$ IO *v* [$_{VP}$ V [$_{I\text{-}ApplP}$ t_{IO} [$_{I\text{-}Appl'}$ I-Appl DO]]]]

v*P: maximal phase*

Both synthetic and analytical groups have maximal phases such as E-ApplP:

(15) *Italian, Russian, Hebrew, Kinyarwanda, English, Icelandic*

[$_{E\text{-}ApplP}$ DO [$_{E\text{-}ApplP}$ PP [$_{E\text{-}Appl'}$ E-Appl [$_{VP}$ V t_{DO}]]]]

E-ApplP: maximal phase

The absence of min-phases is characteristic of languages with fixed word order. When subject and object have to be ordered with respect to the verb, *v*P is the phase. The process is different when relations between words are established by means of inflections and the requirement of ordering is not so strict.[23,24]

(i) Umukoôbva$_i$ a-ra-andik-ir-w-a t_i íbárúwa   n'ûmuhuûngu.
     girl   *SP-PRES-write-APPL-PASS-ASP*   *letter*   *by.boy*
     'The girl is having a letter written for her by the boy.'

(ii) Íbárúwa$_i$ i-ra-andik-ir-w-a   umukoôbva t_i n'ûmuhuûngu.
     letter   *SP-PRES-write-APPL-PASS-ASP* *girl*   *by.boy*
     'The letter is written for the girl by the boy.' (McGinnis 2001)

[22] There is a restriction on movement of the direct object of ApplP *Haraldur* in Icelandic (i) but not in Italian (ii). A unifying explanation of these constructions can be provided if I-ApplP is a phase in Italian but not in Icelandic.

(i)   *I-ApplP is not a phase*
      a. Jón   telur [mér$_i$ virðast t_i [Haraldur   hafa   gert þetta vel]].
         John.NOM *believes* *me.DAT* *seem.INF*   *Harald.NOM* *have.INF* *done this   well*
         'John believed Harald to seem to me to have done this well'.
      b. *Jón   telur [Haraldur$_i$ virðast   mér [t_i hafa   gert þetta vel]].
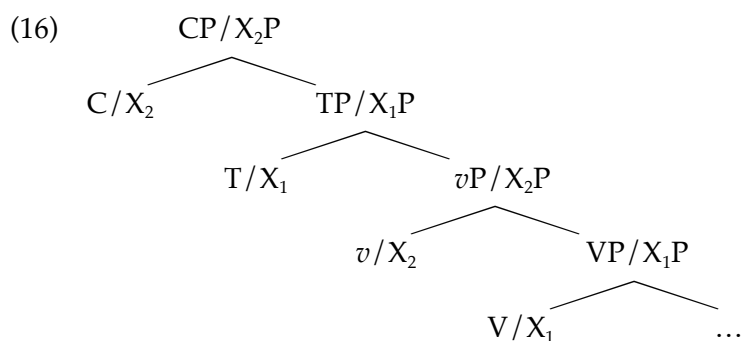
(ii) *I-ApplP is a phase*
      Gianni   non   [gli   sembra [t   fare   il   suo   dovere]].
      Gianni.NOM *not* *him.DAT* *seems*   *do.INF* *the his   duty*
      'Gianni does not seem to him to do his duty.'

[23] When English is compared to languages with overtly marked dative case in sentences with *give*, the recipient NP in the *to*-construction is sometimes equated with dative NP. In this sense, those languages lack constructions with I-ApplP (they have only E-ApplP). This does not explain the cross-linguistic distribution of object movement and consistency of passive formation in both applicative and double object constructions with *give, send*, and the like.
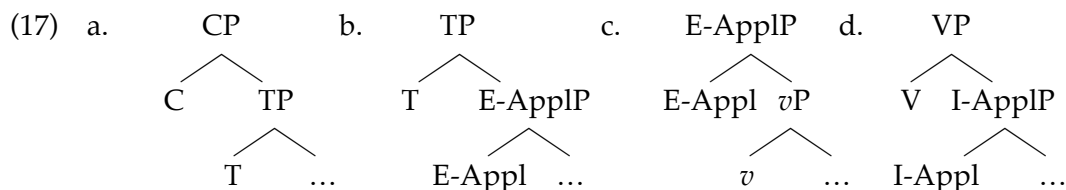
[24] There is additional evidence that syntactic structures that express a relation between *individuals* should be considered more basic than those expressing a relation involving events. In languages with phasal I-ApplPs, sentences such as *A boy tore a girl a skirt, My*

3.3. Phase Parallelism

As was already proposed, phase selection is language specific, while any syntactic phrase may in principle constitute a phase. These label-free phases — compared along the lines of their configurations only — exhibit parallelism. For example, $[_{CP}$ C $[_{TP}$ T$]]$ and $[_{vP}$ v $[_{VP}$ V$]]$ are parallel because both have a no-label dyadic representation $[_{X2P}$ X$_2$ $[_{X1P}$ X$_1]]$ at their base (16). The difference between two types of phases is in whether a phase is *minimal/incomplete* (X$_1$P) or *maximal* (X$_2$P).

(16)

```
            CP/X₂P
          /        \
      C/X₂          TP/X₁P
                   /      \
               T/X₁        vP/X₂P
                          /      \
                      v/X₂        VP/X₁P
                                 /      \
                             V/X₁        …
```

At some level, $[_{CP}$ C $[_{TP}$ T$]]$ and $[_{VP}$ V $[_{I\text{-}ApplP}$ I-Appl$]$ are parallel (17). If I-ApplP can in principle constitute a minimal phase, then one may expect to identify other minimal phases (such as TP) in a language where I-ApplP is phasal.[25]

(17) a.   CP   b.   TP   c.   E-ApplP   d.   VP
            /   \   /   \   /   \   /   \
          C   TP   T   E-ApplP   E-Appl   vP   V   I-ApplP
                /   \   /   \   /   \   /   \
              T   …   E-Appl   …   v   …   I-Appl   …

What happens when TP behaves as a minimal phase? A certain class of verbs assigns structural case to an embedded subject in Exceptional Case Marking (ECM) constructions in sentences such as *Eve wanted Adam$_{ACC}$ to taste an apple*, where the NP *Adam* is assigned accusative Case by the matrix verb *want*. This fact was accounted for in terms of CP-reduction. If this is a universally accessible rule, it is not clear why many languages — with Hebrew, Spanish, and Russian among them — lack ECM:

(18) *Hebrew*

a. * Hava   racta   Adam   lakahat et   ha-tapuax.
       Hava.NOM   wanted   Adam.ACC   take.INF ACC   the-apple
       'Eve wanted Adam to take the apple.'

---

friend broke me glasses, She fixed her neighbor a car, and *A daughter washed her mother the dishes* are regular grammatical structures.

[25] Recall that in the present system, phases are primarily characterized by their capacity to project extra Spec-positions, to ensure continuation of movement. It is possible that *minimal* phases are *incomplete*. TP is not a *maximal* phase; it is internal to CP and dependent on CP.

b. Hava   racta   še   Adam   ekah   et   ha-tapuax.
    Hava.NOM wanted that Adam.NOM would-take ACC   the apple
    (lit.) 'Eve wanted that Adam took the apple.'

(19) *Spanish*

a. * Eva   quisiera Adam   tomar   la manzana.
      Eva.NOM wanted   Adam.ACC take-INF the apple
      'Eve wanted Adam to take the apple.'

b. Eva   quisiera que Adám   tomara   la manzana.
    Eva.NOM wanted   that Adam.NOM would-take   the apple
    (lit.) 'Eve wanted that Adam would take the apple.'

(20) *Russian*

a. * Jev-a   xotela   Adam-a   vzjat'   jabloko.
      Jev.NOM wanted   Adam.ACC take-INF apple
      'Eve wanted Adam to take the apple.'

b. Jev-a   xotela   čtoby Adam   vzjal jabloko.
    Jev.NOM wanted   that   Adam.NOM took   apple
    (lit.) 'Eve wanted that Adam took the apple.'

The explanation of this contrast lies in the distribution of the language-specific types of phases. The absence of ECM can be accounted for if a language has minimal (internal) phases, and one such phases is TP. The explanation is as follows. Once the lower T_{INF}P-phase is complete, subject NP in [Spec,T_{INF}P] requires Nominative Case that cannot be assigned in this position due to the properties of T_{INF}. The conflict between Case requirements and phasal status of TP cannot be resolved, and derivation crashes. In English, TP is not a phase, and subject moves to object position of matrix verb to receive Accusative Case.[26]

For the same reason, these languages lack Optional Infinitival Stage. English-speaking children at some stage between 1;10-2;7 on occasion omit TPs by producing sentences such as *Mary like John* (Wexler, 1998). Cross-linguistic data shows that this stage is absent in Polish, Russian, Italian, and Spanish. Evidently, minimal phases such as TP cannot be omitted even at an early stage of language development. The cross-linguistic distribution of Optional Infinitives in child language is consistent with the proposed universal phase parallelism and the existence of two types of phases.

3.4. *Strict Cycle Condition*

Chomsky (1973: 243) states the Strict Cycle Condition as follows: "No rule can apply to a domain dominated by a cyclic node A in such a way as to affect solely a proper sub-domain of A dominated by a node B which is also a cyclic node." The Strict Cycle Condition is borne out in Russian, a language characterized by

---

[26] When nominative Case assignment is unnecessary (e.g., in *Eve wanted PRO to taste an apple*), the derivation survives in a language characterized by min-phases (e.g., the corresponding Russian *Jeva xotela vzjat' jabloko*).

min-phases that allow DO-movement (21). This blocks IO-movement (22).[27]

(21) a.   Rubaška   byla   vyšita / dana   Petru.
            shirt.NOM   was   embroidered / given   Peter.DAT
   (lit.) 'A shirt was embroidered Peter.'
      b.   [$_{vP}$ DO v [$_{I-ApplP}$ t_{DO} [$_{I-ApplP}$ IO [I-Appl t_{DO}]]]]
            I-ApplP: minimal phase

(22) a.   *Petr   byl   vyšit / dan   rubašku.
            Peter.NOM   was   embroidered / given   shirt.ACC
            'Peter was embroidered / given a shirt.'
      b.   *[$_{vP}$ IO v [$_{E-ApplP}$ E-Appl [$_{VP}$ V [$_{I-ApplP}$ [$_{I-ApplP}$ t_{IO} [I-Appl DO]]]]]]

The above restriction complies with the Phase Impenetrability Condition (PIC) —
it requires that the domain of a phase be opaque; only the edge and the head are
visible to later syntactic operations. From a more general perspective, in a system
where X(n) = X(n–1) +X(n–2), GR — the Golden Ratio — between the terms is
preserved only when each term is combined with the one that *immediately* pre-
cedes it. Once a phase is complete, it is impossible to extract yet another element
from its domain. For example, 5 is a sum of 3 and 2. If the sum were formed by
adding 1 (instead of 2) to 3 etc., a sequence would yield (1, *1*, 2, *3*, 4, 6, …),
violating GR.

   Among other things, this explains why DO-movement bleeds IO-
movement in Russian applicative constructions, and presents yet another proof
that I-ApplP is a phase in this language:

(23)
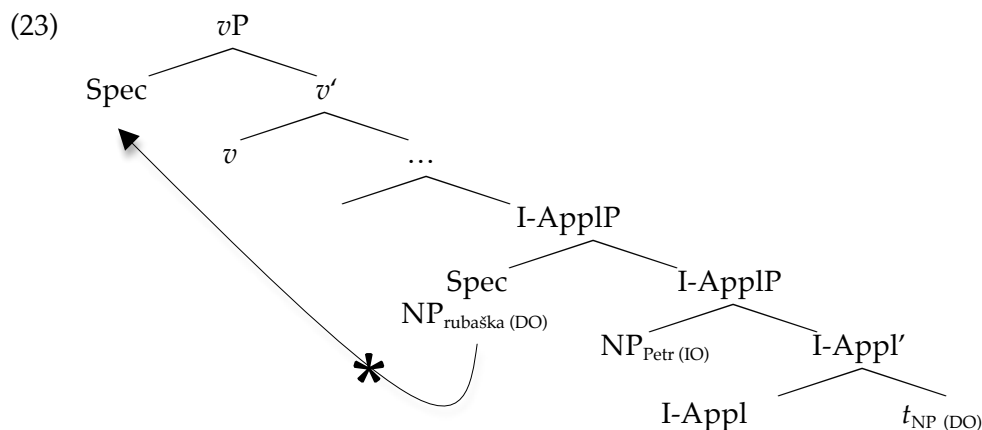
3.5. *Spell-Out and Interpretation of Phases*

Chomsky (2001) identifies vP and CP as fully-fledged phases, relatively
independent at the interface and spelled out cyclically. Epstein & Seely (2002)
find this specification problematic: How do we know they are independent at the

---

27   In (21b) and (22b), X'-nodes are subsumed. In (23) below, the nodes E-Appl' and V' are
      subsumed. This is not a contradiction to the claim that X'-levels should be eliminated. At
      present, the existing X'-model is indispensable for syntactic representations.

interface if Spell-Out applies before the interface is reached? The explanation is as follows: These are the phases within which all theta-roles are discharged, evidence that the underlying label-free argument-centered component is preserved throughout derivations.

Consider, for example, the sentences *John left his girlfriend with a baby* and *John left his girlfriend with a smile*. The interpretation of these and similar sentences (inspired by Chomsky's examples) varies, which can be made explicit by the extended semantics of V (meaning 'John impregnated his girlfriend' vs. 'John walked away from his girlfriend with a smile on his face'). The argument-centered representations below make the distinction transparent. There is no requirement to extend V_{left}: *John left his girlfriend (with a smile)* has two obligatory participants ((6a), repeated here), while *John left his girlfriend *(with a baby)* has three ((6b), also repeated here). Clearly, a rule that determines the number of arguments and their hierarchy is preserved through derivations until PF is accessed.[28]

(6)   a.   b.
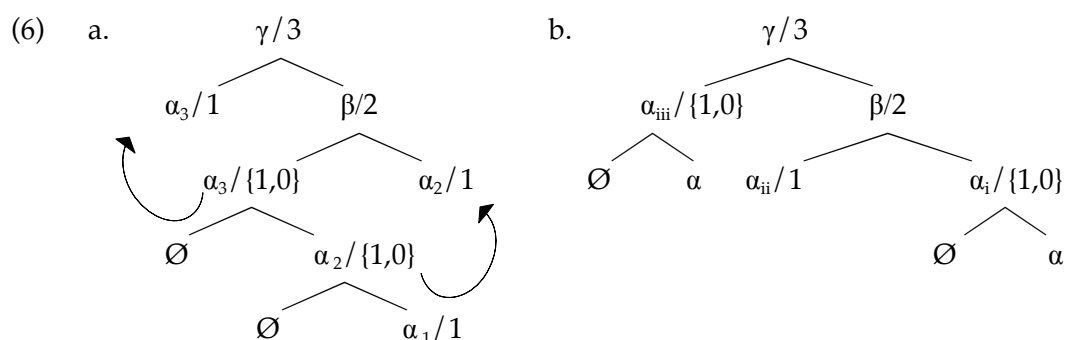
At the end of each phase, derivations are sent off to PF (Phonological Form, Spell-Out) and LF (Logical Form, Interpretation). The next important question is how PF and LF are derived in a language system with two types of phases — maximal and minimal. According to Epstein & Seely (2002), some features of lexical items are illegitimate at one or the other interface. For instance, the pronoun *him* seems synonymous with *he*, even though their PF-interpretations are distinct. It was assumed that unvalued lexical features are illegible at both LF and PF; valuation, however, is a necessary, but not sufficient, condition for LF convergence. The Case feature of a DP/NP may be valued by the operation Agree, but a valued Case feature is by hypothesis still not interpretable at LF, and can be interpreted only at PF.

Let us assume that this interface disassociation is crucial to the distinction between min- and max-phases. If (all) languages have max-phases (CP, E-ApplP), and certain languages in addition have min-phases (TP, I-ApplP), heads of min-phases ensure realization of Agree for the continuation of movement, but it is max-phases that are sent to PF. One example are 'garden-path' sentences (Gibson

---

[28]   Argument-based representations in (23) above may also provide valuable insight into the existing similarity between applicative and double object constructions. Even though the former has an optional argument (*John baked (Mary) a cake*) and the latter has three obligatory arguments (*John gave Mary a cake*), object movement in passives shows consistency in both: These constructions have a common core.

2000). The sentence [$_{CP1}$ *The horse raced past the barn*] is interpreted as complete; the resultant derivation is sent to PF and LF. In *The horse raced past the barn fell*, this constituent CP$_1$ is interpreted as NP — [$_{CP2}$ [$_{NP}$ *The horse raced past the barn*] *fell*]. At the end of derivation, a completed max-phase (CP$_2$) is sent to PF. To conclude, in a label-free system underlying syntactic representations,

- phase heads are characterized by their ability to project Spec-positions;
- any phrase may in principle constitute a phase;
- phases can be compared along the lines of their configurations;
- all languages have maximal (propositional) phases, certain languages also have minimal/internal phases;
- at the end of derivation, maximal phases are sent to PF. [29]

4.   Argument-Centered Representations

4.1.   'Verbless' Languages

A relation between individuals may constitute a phase and induce movement (recursion). This means that the core syntactic representations do not necessarily require a verb. The argument-centered logic of minimal syntactic units relies to a large extent on the data from language acquisition: Nouns are acquired first by children who have 'perfect grammar', equipped with the innate principles of universal syntax that allow them to master any language. Child language abounds in 'verbless' and 'copulaless' constructions. These structures are preserved in English in small clauses, such as *We consider [$_{SC}$ Mary a friend]*, for example. Furthermore, many languages construct sentences of the kind *Mary is smart* without a copula.

Across language systems, nouns have a special status that ranks them higher than verbs.[30] Certain languages have a very restricted number of verbs. For example, the aboriginal language Jingulu spoken in Australia has only three verbs: *do*, *go*, and *come*. Igbo (Ibo), a language of approximately 18 million speakers in Nigeria, does not have verbs at all. Instead, Igbo uses clusters termed 'inherent complement verbs' (ICV) that have the structure –*gbá* plus a noun. The root *gbá* is the only root in Igbo "devoid of meaning", and the most productive one (Chinedu Uchechukwu, p.c.; see also Uchechukwu 2004). Here are some

---

[29]   For reasons already given, languages with min-phases always have max-phases, while the max-phase group may in principle (but not necessarily) have min-phases. An example appears to be Icelandic that has ECM-constructions found in languages with max-phases. It also has dative experiencer constructions characteristic of languages with min-phases (e.g., the equivalent of *John.NOM me.DAT likes* meaning 'I like John'). Such a min-phase is I-ApplP [*John.NOM*$_{NP}$ *me.DAT*$_{NP}$]. Moreover, certain dialects of English appear to have an I-ApplP phase by allowing constructions such as *A cake was baked Mary*.

[30]   Additional evidence comes from iconic languages of children of deaf parents. Deprived of formal linguistic input, these children simultaneously invent languages in which the gesture for *give* is associated with three noun phrases, the gesture for *kick* with two, and the gesture for *sleep* with one (Lidz & Gleitman 2004).

examples of Igbo clusters: *–gbá egwú* 'dance a dance' — *egwú* 'dance', *–gbá igwè* 'ride a bicycle' — *igwè* 'bicycle', *–gbá ákú̯/egbè* 'shoot' — *àkú̯* 'arrow'/*égbè* 'gun', *gbá ú̯kwú̯* 'kick' — *ú̯kwu̯* 'foot', *–gbá oso̯* 'run a race' — *o̯so̯* 'race', *–gbá motò* 'travel with a vehicle' — *motò* 'vehicle', etc.

The structure with ICV in Igbo linguistics has always been problematic for analysis. The first characteristic that differentiates its use from light verbs in other languages is that it is a regular linguistic means. The second is that these structures do not have any simple verb equivalent. As a matter of fact, *gbá* cannot be considered equal to a light verb: In expressions *take a leap*, *take a leak*, etc., there is no sharp divide between word and phrasal special meanings (Marantz 1997). In contrast, in Igbo, the semantic meaning of *–gbá*-clusters encodes the intrinsic connection between two key arguments, agent and theme, based on the primary function of the theme with respect to the agent. For example, the basic function of a car is to carry passengers. Accordingly, *–gbá motò* means 'travel with a vehicle'; it does not mean 'repair a vehicle' or 'sell a vehicle'.[31]

For Igbo, we postulate a Relational Phrase (RelP) whose head Rel is expressed overtly as a semantically vacuous element *–gbá* that establishes a relation between individuals (similar to I-Appl):

(24)   [$_{RelP}$ Spec [Rel–*gbá* [α, β]]]

Igbo clusters make explicit a hierarchical distribution of arguments in the absence of a verb. In a label-free representation of argument structure, agent is the first element to be Ø-merged to form a singleton set {α, 0}, type-shifted to α, and then moved to [Spec,RelP]:

(25)

---
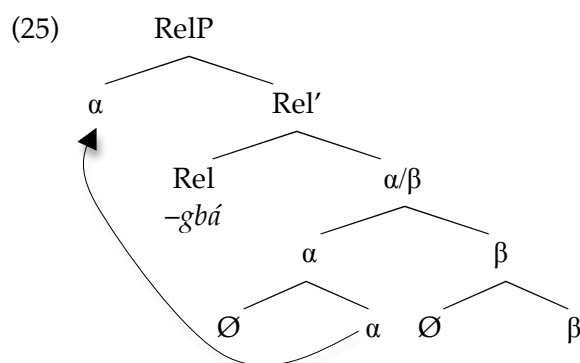
[31]   Note that inflected *–gbá* roots are not semantically empty: For example, *–do* is a suffix that expresses 'fixation of the activity' in *–gbá–do*. Other roots (e.g., *–tu*, *–kpa*,*–ma*) check semantic features of the nouns they are combined with, such as 'animacy' and 'shape'. This feature-checking is similar to what is reflected as the SER/ESTAR alternation in Spanish and Portuguese. The choice of a particular copula is consistent with a ±permanency feature of the predicate: SER is chosen over ESTAR when 'sourness' is a permanent property of the subject.

(i)   a.   Os   limões   são [SER]   ácidos.   / *Os   limões   estão [ESTAR]   ácidos.
             the   *lemons*   *be.3PL*   *sour*   /   *the*   *lemons*   *be.3PL*   *sou*
             'The lemons are sour.'
        b.   As   maçãs   estão [ESTAR]   ácidas. / *As   maçãs   são [SER]   ácidas.
             the   *apples*   *be.3PL*   *sour*   *the*   *apples*   *be.3PL*   *sour*
             'The apples are sour.'   (Costa 1998)

In the propositional setting, verbs cannot be eliminated. In contrast, from the point of view of Fib-like logic, the operation Merge is unconstrained, and any two successive elements may be merged to form a part of recursive system. If a certain type of phase can be defined as non-propositional, then EM can be represented as a mechanism that establishes the hierarchy of α and β, depending on whether α or β is Ø–merged first. The representation {{α, Ø}, {β, Ø}} in (26a) below is chosen over {α, β} in compliance with the NL-requirement: The sum of terms needs to be represented at each level.

Thus, the core EM operates on two symmetrically conjoined elements — α, β. The 'argument-oriented' mechanism establishes a hierarchical relation between α and β in some *relational* configuration RP — this means that its R carries a certain feature [R] that triggers the selection. In principle, either α or β can check the R-feature. The choice of the element depends on which sum undergoes EM first: If α is Ø-merged first, then α is ranked higher (26b).

(26) a. $[_{RP}$ **RH** $[[α, Ø], [β, Ø]]]$
      b. $[_{RP}$ α $[$ **RH** $[β, Ø]]]$

The requirement of EM to disregard order in favor of hierarchy is evident in the following.[32] When asked to complete a sentence, readers preferred conjuncts with a shared subject over object conjuncts, and both over clause conjuncts (Hoeks & Hendriks 2005, from which the following examples are taken). S-coordinated sentences such as (27) were used, the first of which was temporarily ambiguous, whereas the latter served as a control sentence, made unambiguous by inserting a comma after the first object NP.

(27) a.   The model embraced the designer and …
      b.   The model embraced the designer (,) and the photographer opened a
              bottle of expensive champagne.

(28) a.   The model *embraced the designer* and *laughed*.   *VP-conjunct*
      b.   The model embraced *the designer* and *the photographer*.   *NP-conjunct*
      c.   *The model embraced the designer*, and *the photographer*
              opened a bottle of expensive champagne.   *TP-conjunct*

Language users strongly prefer to continue a fragment such as (27a) for VP-coordination (28a). The second NP was interpreted by the readers as the object of the first clause (28b) rather than the subject of the second clause (27c). Both sentences *The model embraced the designer and laughed* and *The model embraced the designer and the photographer* were ranked higher than the one that had conjoined clauses, such as *The model embraced the designer, and the photographer opened a bottle of expensive champagne.* An account for the above-mentioned differences can be provided if VP-conjuncts are selected because both VPs share the same agent for

---

[32]   Kayne's (1994) Linear Correspondence Axiom derives linear order from strict asymmetric c-command. Linearization applies only at the level relevant for pronunciation — Spell-Out (Chomsky 2000).
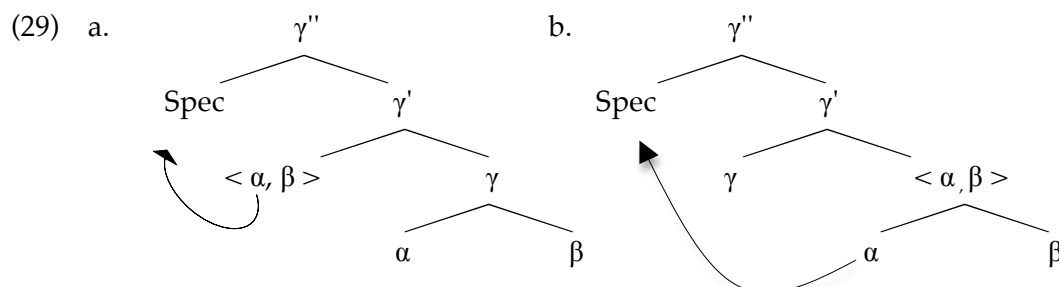
both verbs, ranking (28a) higher than (28b). Theme is next in the hierarchy of arguments; hence, (27b) is selected rather than (28c). The conclusion is that not all conjuncts are equal, and preference is given to the structure that identifies agents first, before a verb is introduced.[33]

4.2.   *Word Order: Subject-First*

Grammatical linguistic expression is the optimal solution, the reason why a particular word order 'Subject-first' is preferred across languages. In this section, it will be shown that cross-linguistic differences regarding the order of major constituents (Subject–Object) reflects the ways the system implements the notion 'preference', which attests to the intrinsic hierarchy of arguments. The SO order remains constant in the majority of languages (96%, Dryer 2005); SOV (rather than SVO) is the predominant pattern. The highest preference is given to languages that are either SO-first, or S-first. The canonical word ordering in optimal terms is SOV, SVO, VSO, VOS, OVS, and OSV.[34]

The introduction of R-function as a means of hierarchical prioritization is offered as an account for the ranking of word order across languages. The structure $\{\alpha, \beta\}$ is *symmetrical*; α and β share an equal chance for movement. The Relational head R takes a pair $\{\alpha, \beta\}$ and establishes a hierarchy of elements in RP. The choice of which element is ranked higher depends on which sum is merged first. If α is $\varnothing$-merged first, α is displaced first. The output of the function R is an *hierarchically* ordered pair — either $<\alpha, \beta>$ or $<\beta, \alpha>$. The order $<\alpha, \beta>$ is preferred to $<\beta, \alpha>$. In our system, α corresponds to Subject and β to Object. Once S and O are ordered in RP, SO undergoes (Verb)-linearization. It has two options, with the first ranked higher than the second:

- Constituent SO is displaced: The resulting order is either $<<\alpha, \beta>, \gamma>$ or $<\gamma, <\alpha, \beta>>$, where γ is V; $<<\alpha, \beta>, \gamma>$ (SOV) is preferred to $<\gamma, <\alpha, \beta>>$ (VSO) (29a).
- S is displaced: The resulting word order is $<\alpha, \gamma, \beta,>$ (SVO) (29b).

(29)   a.

b.

In Object-first languages, R takes a pair {α, β} with an output of the ordered pair <β, α> (OS), then the verb merges with OS. These are the options:

- The whole constituent OS is merged with V: The order <γ, <β, α> > (VOS) is preferred to < <β, α>, γ> (OSV).
- The first constituent O is merged with V: < β, γ, α > (OVS).

It may be argued, however, that even though S+O (in SO languages) and O+S (in OS languages) in some cases display syntactic independence such as moving as a constituent, it is far from being typical or unmarked. This can be explained if movement is re-evaluated as the 'internal' version of Merge, thus not an 'imperfection' of language. The *internally* merged elements A, B have to be independent to occupy positions in a tree that are justified by the principles of efficient growth. However, in a symmetrical representation of *externally* merged arguments, an equal status is assigned to each of Ø-merged elements at some level, which is why conjoined structures such as bare nouns in conjunctions move as one constituent. In this sense, word order is a true reflection of the argument-centered syntactic primitive characterized by symmetry.

4.3. *Symmetrical Conjuncts*

The analysis under development shifts the focus from verb to the noun, from the propositional to the non-propositional logic of syntactic representations. The conclusion we have arrived at is that a minimal syntactic domain (phase) can be defined in non-propositional terms, such as a *relation between individuals*. The key requirement of the computational system of human language now includes an argument-centered configuration. As was already shown, a lower part $[_{XP} X]$ of $[_{VP} V [_{XP} X]]$ represents a phase in certain languages, contrary to what had been previously assumed.

In the present system of N-Law application, there is every reason to believe that a non-linear representation is characterized by *symmetry* of the basic form {{α, Ø}, {β, Ø}}.[35] Recall that Ø-Merge at the bottom level of the tree is necessitated by the requirement to induce a progressive cycle implemented by sums rather than single elements; {{α}, {β}} is preferred over {α, β}.[36] *Symmetrical conjuncts* are the core syntactic primitives, while displacement obeys the requirement to obtain a linear (asymmetric) order. Thus, the true representations underlying syntactic constructs can be characterized within a remarkably weak formalism of what we can call *conjunctivism*.[37]

---

[35] See Moro (2000) on the possibility of symmetry at base structure, resolved into asymmetry by Spell-Out. Kratzer's (1996) argumentation that the subject should be introduced by a separate predicate opposes the view presented here.

[36] Linguistic evidence suggests that certain lexical items that participate in conjunctions are Ø-branching projections, e.g., prepositional heads (*up and down the road*) and bare nouns (*cat and dog, knife and fork*). It is well known that conjuncts behave differently from other syntactic structures that can be derived from X-bar schema: Movement of a sub-part of a conjunct is prohibited.

[37] Conjunctivism says that *absolutely all* semantically relevant syntactic concatenation expresses conjunction (Pietroski 2005).

5.   Species-Specific Properties of FLN

Hauser *et al*. (2002) argue that FLN may have evolved for reasons other than language. Gallistel *et al*. (2006) arrive at the following conclusion:[38]

> [T]he nonverbal system for arithmetic reasoning with mental magnitudes precedes the verbal system both phylogenetically, and ontogenetically. […] The special role of the natural numbers in the cultural history of arithmetic is a consequence of the discrete character of human language, which picks out of the system of real numbers in the brain the discretely ordered subsets generated by the nonverbal counting process, and makes these the foundation of the linguistically mediated conception of number.
>
> (Gallistel *et al*. 2006: 270-271)

In this part, rather than trying to identify the driving force behind the evolution of FLN from a non-verbal to verbal form, we will continue approaching language as part of a general natural system, while continuing our search for the criteria that single out this particular computational mechanism as species-specific.

As previously discussed, an important property of FLN is recursion. Is it possible to have a non-recursive human language? Recently, a claim was made by Everett (2005) that Pirahã, a language spoken by approximately 250 speakers in Amazonas, Brazil, lacks a specific recursive property exemplified as embedded clauses in other languages. Nevins *et al*. (2007) argue that these grammatical "gaps" are incorrectly analyzed by Everett — most of the properties under discussion are familiar from languages whose speakers lack the cultural restrictions attributed to Pirahã, a language of the so-called *immediate experience* restriction. [39] Pirahã has possessive constructions such as in (30) but not in (31a); however, the same absence of constructions such as *John's mother's hat* can be found in German (31b). Furthermore, the language cannot be claimed to lack embedded clauses. In displaying VO word order where the object is a clause, Pirahã, an OV language, shows VO in a post-verbal clausal complement (32). This is a choice made by many other languages.

---

[38]   Blakemore & Frith (2005) observe that patients with an impaired system of calculation (summation, subtraction) still preserve the ability to estimate quantities, confirming the assumption that basic mental representations are continuous.

[39]   Pirahã uses a special copula (3rd person pronoun) to distinguish between individual- and stage-level predicates that express a distinction between permanent and temporary qualities, just as Hebrew does (Soschen 2003). It follows then that Pirahãns do differentiate between types of *experience.*

> (i)   *Pirahã*
>   Giopaíxi   hi   sabí-xi.
>   *dog   COP.3SG   wild*
>   'The dog is really wild.'
>
> (ii)   *Hebrew*
>   a. Dani   hu   nexmad.
>   *Dani   he.3SG   nice*
>   'Dani is a nice person (indeed).'
>   b. Dani   nexmad   haiom.
>   *Dani   nice   today*
>   'Dani is nice today.'

(30) a.   John's car   *English*
      b.   Hans–ens Auto   *German*

(31) a.   [John's car's] motor   *English*
      b.   *[Hans–ens Auto]–s Motor   *German*

(32) hi ob13–áaxáí   [kahaí   kai–sai]
      3   see/know–INTNS   arrow   make–NOMLZR
      'He really knows how to make arrows.'   (Nevins *et al.* 2007)

While the property of recursion most probably should be considered species-specific and thus attributed to all human languages, infinity is another feature that makes FLN crucially different from other discrete systems found in nature. There is no limit to the length of a meaningful string of words: There are ten-word sentences, twenty-word sentences, and so on, indefinitely. These properties are exemplified in a well-known nursery rhyme 'The House That Jack Built'. In the rhyme, each sentence X_k with a number of words *n* is succeeded by a sentence X_{k+1} with a number of words *n+m*: X_{k+1} (*n*) = X_k (*n+m*), X_2 (*n*) = X_1 (*n+4*), …, X_5 (*n*) = X_4 (*n+4*), X_6 (*n*) = X_5 (*n+8*), …[40,41] Language is also discrete: There are neither *n*-and-a-half words nor *n*-and-a-half-word sentences (Chomsky 2000). Syntactic units are also continuous: Once a constituent is formed, it cannot be broken up into separate elements. For example, *The dog chased the cat* is the basic representation; in a passive construction *The cat was chased* ~the cat~ *by the dog*, the NP *the cat* moves to the beginning of the sentence only as a constituent. Hence, *Cat was chased the* ~cat~ *by the dog* is ungrammatical.[42]

The application of N-Law logic to the analysis of syntax results in the re-evaluation of FLN as part of a larger mechanism designed for continuation of movement. A general physical law that appears in every living organism applies to the universal principles of grammar: FLN complies with the maximization requirement as well. The Fib-rule accounts for the limitations imposed on the number of arguments in thematic domains, and it also explains why syntactic derivations proceed by phases. Merge is an essential part of a unique recursive

---

[40]   "*This is the house that Jack built₁. This is the malt* That lay in the house that Jack built₂. *This is the rat*, That ate the malt That lay in the house that Jack built₃. *This is the cat*, That killed the rat, That ate the malt That lay in the house that Jack built₄. *This is the dog*, That worried the cat, That killed the rat, That ate the malt That lay in the house that Jack built₅. *This is the cow with the crumpled horn*, That tossed the dog, That worried the cat, That killed the rat, That ate the malt That lay in the house that Jack built₆."

[41]   In contrast, other biological systems exhibit finiteness. For each kind K of flower (a, b, c, d, e, …), there is a fixed number of petals X that corresponds to a Fib-number: K_a=X (3), K_b=X (5), K_c=X (8), K_d=X (13), K_e=X (21), K_f=X (34), … X (3) = X (3–1) + X(3–2) [lily, iris], X (5) [buttercup, wild rose, larkspur, columbine], X (8) [delphiniums], X (13) [ragwort, corn marigold, cineraria], X (21) [aster, black-eyed susan, chicory], X (34) [plantain, pytethrum], X (55), X (89) [michelmas daisies, the asteraceae family].

[42]   I am indebted to an anonymous reviewer for the following remark: "Saying that the ungrammaticality of *Cat was chased the by the dog* is due to the fact that *the cat* can move only as a constituent raises the question of why *cat* cannot behave the same way and move as a constituent". As one possible explanation, a Ø-merged element behaves as a constituent at the level of EM but not in IM. EM establishes hierarchical relations only; there is no movement in EM.

mechanism exemplified as phases in syntax.

In the present work, the *impenetrability* of already formed constituents (as the result of a specific type-shifting operation) is viewed as *the key requirement of FLN*. In contrast, segments comprising other GR-based systems of growth can in principle be separated from one another. Following from that, FLN as a subsystem of natural development based on optimal space filling can be represented graphically, representing both *discreteness and continuity of its constituents* (Figures 2 and 3 below).
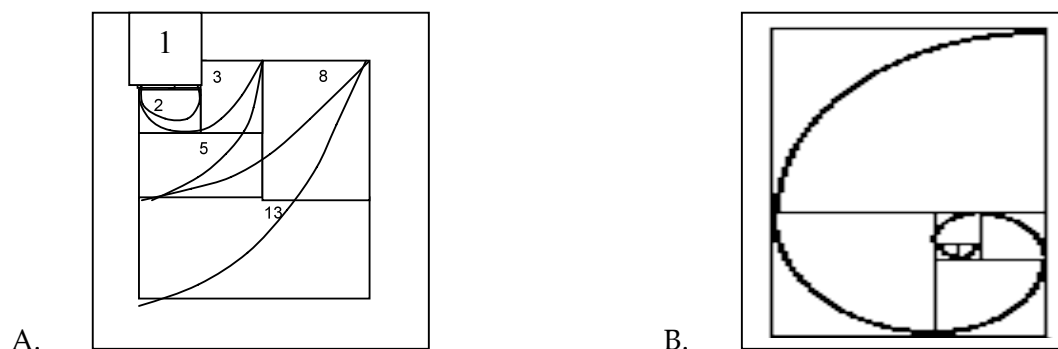
Figure 2: Pendulum- (A) vs. spiral-shaped (B) GR-based systems

Depending on whether the phase is complete or not, each constituent may appear either as a part of a larger unit or the sum of two elements. For example, one line that passes through the squares '3' and '2' connects '3' with its part '2'; the other line indicates that '3' as a whole is a part of '5'.

The *pendulum-shaped graph* to the left is contrasted with a non-linguistic model to the right where one line connects the preceding and the following elements in a spiral configuration of a sea-shell. This system does not comply with IC. For example, '3' is a sum of '2' and '1', while '2' is comprised of separate elements '1' and '1'. There is no line that connects '2', '3', and '5' in such a way that '2' as a whole is a part of '5' (Figures 2B, 3B).
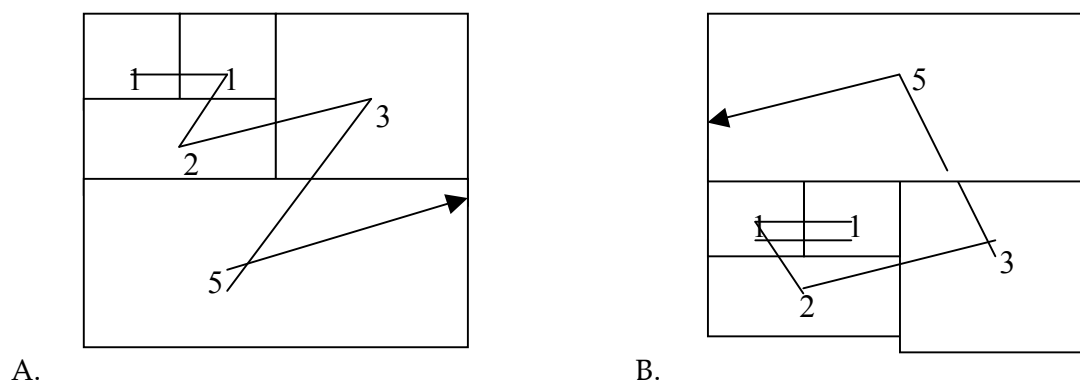

Figure 3: Configurations A and B (Figure 2) made explicit

The distance between the 'points of growth'/segments in the above representations can be measured according to GR, the requirement of optimization. The structure of FLN complies with N-Law; however, in contrast with other natural systems of growth, each element appears as either discrete (the sum of two elements) or continuous (part of a larger unit).

6.   Summary and Conclusions

This analysis, applied to the sequence of nodes in syntactic trees along the lines of N-Law, has focused on a functional explanation of binary branching, labeling, and the properties of the existing types of Merge. The optimization requirement justifies the basic principle of organization in both External and Internal Merge, the two forms of a basic Merge. EM either returns the same value as its input (Ø-Merge), or the cycle results in a new element (N-Merge). EM is responsible for the number of arguments, which corresponds to the number of positions available to the element adjoining a Fib-like tree. Maximal thematic domains incorporate all possible argument-based representations. This argument-centered approach shifts the focus from verb to noun, from the propositional to the non-propositional logic of grammar.[43] The minimal building block that enters into linguistic computation is identified as a symmetrical conjunct, which expresses a relation *between individuals* (rather than *between individuals and events*). As a result, the true structure of language is characterized within a remarkably weak formal system, which is expected to develop into a more complex one to handle a broader range of data.

IM is induced by the necessity that lexical items must obtain a linear (asymmetric) ordering. Movement depends on the qualification of phrases as phases. Any phrase can in principle constitute a phase. Phase heads are characterized by the ability to project specifier positions to ensure continuation of movement. Presumably all languages have maximal phases; in addition, synthetic (inflected) languages have minimal (i.e. Individual Applicative) phases. The label-free phases can be compared according to their configurations. As one example, this comparison provides an account of why languages with minimal phases lack ECM structures.

By developing the idea that linguistic structures have the properties of other biological systems, we have reached some conclusions concerning the underlying principles of the computational system of the human language. The Faculty of Language obeys the rule of optimization. However, in contrast with other GR-based natural systems of efficient growth, at some level each syntactic constituent may appear as either discrete or continuous. The impenetrability of already formed constituents — which in itself is a result of a unique type-shifting operation — is viewed as the key condition imposed upon FLN.

---

[43]   The argument-centered model of syntactic representations is experimentally supported in Soschen & Slavova (2008).

References

Blakemore, Sarah-Jayne & Uta Frith. 2005. *The Learning Brain: Lessons for Education*. Malden, MA: Blackwell.

Boeckx, Cedric. 2006. *Linguistic Minimalism: Origins, Concepts, Methods, and Aims.* New York: Oxford University Press.

Bošković, Željko. 2002. A-movement and the EPP. *Syntax* 5, 167-218.

Carnie, Andrew. 2000. On the definition of X^0 and XP. *Syntax* 3, 59-106.

Carnie, Andrew, David Medeiros & Cedric Boeckx. 2005. Some consequences of natural law in syntactic structure. Ms., University of Arizona, Tucson & Harvard University, Cambridge, MA.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen R. Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232-286. New York: Holt, Rinehart and Winston.

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28), 167-217. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. *The Architecture of Language*, Nirmalangslu Mukherji, Bibudhendra Narayan Patnaik & Rama Kant Agnihotri (eds.) New Delhi: Oxford University Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language* (Current Studies in Linguistics 36), 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *The Cartography of Syntactic Structures*, vol. 3: *Structures and Beyond* (Oxford Studies in Comparative Syntax), 104-131. New York: Oxford University Press.

Chomsky, Noam. 2007a. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax–Semantics* (Studies in Generative Grammar 89), 1-29. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2007b. Of minds and language. *Biolinguistics* 1, 9-27.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud* (Current Studies in Linguistics 47), 133-166. Cambridge, MA: MIT Press.

Collins, Chris. 2002. Eliminating labels. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program* (Generative Syntax 6), 42-64. Malden, MA: Blackwell.

Costa, João. 1998. L'opposition ser/estar en portugais. In Alain Rouveret (ed.) *"Etre" et "avoir" (syntaxe, sémantique, typologie)*, 139-153. Saint-Denis: Presses Universitaires de Vincennes.

Douady, Stephane & Yves Couder. 1992. Phyllotaxis as a physical self-organized growth process. *Physical Review Letters* 68, 2098-2101.

Dryer, Matthew S. 2005. Order of subject, object, and verb. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures*, 330-333. Oxford: Oxford University Press.

Epstein, Samuel David & T. Daniel Seely. 2002. Rule applications as cycles in a

level-free syntax. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program* (Generative Syntax 6), 65-89. Malden, MA: Blackwell.

Erteschik–Shir, Nomi. 1979. Discourse constraints on dative movement. In Talmy Givón (ed.), *Discourse and Syntax* (Syntax and Semantics 12), 441-467. New York: Academic Press.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, 95-126. Cambridge, MA: MIT Press.

Gallistel, C. Randy. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, C. Randy, Rochel Gelman & Sara Cordes. 2006. The cultural and evolutionary history of the real numbers. In Stephen C. Levinson & Pierre Jaisson (eds). *Evolution and Culture*, 247-274. Cambridge, MA: MIT Press.

Hale, Kenneth & Samuel Jay Keyser. 2002. *Prolegomenon to a Theory of Argument Structure* (Linguistic Inquiry Monograph 39. Cambridge, MA: MIT Press.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569-1579.

Hoeks, John & Petra Hendriks. 2005. Optimality theory and human sentence processing: Towards a cross-modular analysis of coordination. Ms., University of Groningen.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax* (Linguistic Inquiry Monographs 25). Cambridge, MA: MIT Press.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase Structure and the Lexicon* (Studies in Natural Language & Linguistic Theory 33), 109-137. Dordrecht: Kluwer.

Legate, Julie Anne. 2003. Some interface properties of the phase. *Linguistic Inquiry* 34, 506-516.

Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface* (Linguistic Inquiry Monograph 26). Cambridge, MA: MIT Press.

Lidz, Jeffrey & Lila R. Gleitman. 2004. Yes, we still need universal grammar. *Cognition* 94, 85-95.

Marantz, Alec. 1993. Implications of asymmetries in double object construction. In Sam Mchombo (ed.), *Theoretical Aspects of Bantu Grammar* 1 (CSLI Lecture Notes 38), 113-150. Stanford, CA: CSLI Publications.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4.2, 201-225.

McGinnis, Martha. 2001. Phases and the syntax of applicatives. *North East Linguistic Society (NELS)* 31, 333-349.

Moro, Andrea. 2000. *Dynamic Antisymmetry* (Linguistic Inquiry Monographs 38). Cambridge, MA: MIT Press.

Müller, Gereon. 2004. Phrase impenetrability and *wh*-intervention. In Arthur Stepanov, Gisbert Fanselow & Ralf Vogel (eds.), *Minimality Effects in Syntax*

(Studies in Generative Grammar 70), 289-326. Berlin: Mouton de Gruyter.

Nevins, Andrew Ira, David Pesetsky & Cilene Rodrigues. 2007. Pirahã exceptionality: A reassessment. *LingBuzz* 000411. [http://ling.auf.net/lingBuzz/000411 (18 September 2008).]

Pietroski, Paul M. 2005. *Events and Syntactic Architecture*. New York: Oxford University Press.

Pritchett, Bradley I. 1992. *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.

Pylkkänen, Liina. 2000. What applicative heads apply to. *Proceedings of the 24th Annual Penn Linguistics Colloquium — University of Pennsylvania Working Papers in Linguistics* 7.1, 197-210.

Pylkkänen, Liina. 2002. *Introducing Arguments*. Cambridge, MA: MIT dissertation.

Richards, Norvin. 2001. *Movement in Language: Interactions and Architectures.* New York: Oxford University Press.

Roodenburg, Jasper. 2004. French bare arguments are not extinct: The case of coordinated bare nouns. *Linguistic Inquiry* 35, 301-313.

Rubin, Edward. 2003. Determining pair-merge. *Linguistic Inquiry* 34, 660-668.

Snyder, Kieran M. 2003. *The Relationship between Form and Function in Ditransitive Constructions*. Philadelphia, PA: University of Pennsylvania dissertation.

Soschen, Alona. 2003. *On Subjects and Predicates in Russian*. Ottawa: University of Ottawa dissertation.

Soschen, Alona. 2005. Derivation by phase: Russian applicatives. *Proceedings of the 2005 CLA Annual Conference*. [http://ling.uwo.ca/publications/CLA-ACL/CLA-ACL2005.htm (18 September 2008).]

Soschen, Alona. 2006. Natural law and the dynamics of syntax (MP). *Optimality Theory and Minimalism: A Possible Convergence? — Linguistics in Potsdam* 25, 43-46.

Soschen, Alona & Velina Slavova. 2008. Experimental support of argument-based syntactic computation. *International Journal on Information Theories and Applications* 15, 1-11.

Suranyi, Balázs. 2004. The left periphery and cyclic spellout: The case of Hungarian. In David Adger, Cécile de Cat & George Tsoulas (eds.), *Peripheries: Syntactic Edges and Their Effects* (Studies in Natural Language & Linguistic Theory 59), 49-73. Dordrecht: Kluwer.

Uchechukwu, Chinedu. 2004. Igbo verb clusters: The ontology of the –*gbá* cluster. In Valery Solovyev & Vladimir Polyakow, *Text Processing and Cognitive Technologies*, 256-269. Moscow: MISA.

Uriagereka, Juan. 1998. *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.

Wexler, Kenneth. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106, 23-79.

Wexler, Kenneth. 2004. Theory of phasal development: Perfection in child grammar. *MIT Working Papers in Linguistics* 48, 159-209.

Alona Soschen
Massachusetts Institute of Technology
Department of Brain and Cognitive Sciences
& Department of Linguistics and Philosophy
Mass. Avenue 32–D808
Cambridge, MA 02139
USA

and

Technovation Systems NLP
alona@soschen.com

Also sprach Neanderthalis… Or Did She?

Antonio Benítez-Burraco, Víctor M. Longa,

Guillermo Lorenzo & Juan Uriagereka

1.   Introduction

Two Neanderthals from El Sidrón (Asturias, Spain; Rosas *et al.* 2006) have been recently analyzed by Krause *et al.* (henceforth K) for possible mutations in *FOXP2* (Krause *et al.* 2007), a gene involved in the faculty of language (Lai *et al.* 2001). Although these mutations were believed to be specific to modern humans (Enard *et al.* 2002), this investigation revealed otherwise. Other details of the genomic analysis of these specimens led K to the conclusion that "these two amino acid substitutions […] associated with the emergence of fully modern language ability" (Krause *et al.* 2007: 1908) were probably inherited both by Neanderthals and modern Sapiens from their last common ancestor (300,000 to 400,000 years B.P.).[1]

We argue that the data offered by K are compatible with less drastic interpretations, which we consider in three successive scenarios: (1) the mutations could be selected in Neanderthal's genetic endowment, but for some non-linguistic function; (2) they could be present, but unselected; or (3) they could be transferred into Neanderthals from modern humans through gene flow. Thus K's analysis does not confirm either the antiquity of the human faculty of language or the linguistic capabilities of Neanderthals, and more reliable data are still needed to settle this intriguing question.

The main conclusion that we have reached after discussing this paper is that K's data do not discard the idea that the faculty of language is an evolutionary innovation specific to anatomically modern humans. In fact, such a possibility is now supported by a recent study by Coop *et al.* (2008), and continues to be the most congenial with the behavioral asymmetry between Neanderthals and modern humans that the fossil record reflects (Klein with Edgar 2002: ch. 6, Mellars 2005, and Mithen 2006a). Whatever the origin for the mutations under discussion, they do not entail that Neanderthals from the

[1]   The amino acid substitutions under discussion are caused by nucleotide substitutions at positions 911 and 977 in exon 7 of the *FOXP2* gene, which change threonine to aspartic acid and arginine to serine residues, respectively.

relevant populations were capable of speaking in a modern way. This is quite simply because although *FOXP2* is arguably a necessary condition for language, it almost certainly is not a sufficient one, by any stretch of the imagination. Indeed, for this very reason any corroboration of K's analysis about the antiquity of the mutations in question would not entail that a modern faculty of language was accessible to humans before the evolutionary split leading to Neanderthals.

2.   First Scenario

K claim that the selective sweep on the evolutionary changes of *FOXP2* started before the split of the ancestral populations of Neanderthals and modern humans, some 300,000-400,000 years B.P. They also contend that the fixation of these mutations occurred within the last 260,000 years and were completed by 180,000 years B.P.

Based on an analysis of intronic regions (including the one investigated by K), a previous study by an overlapping team concluded that the modern mutations in *FOXP2* took place within the last 200,000 years, most probably around 125,000 years B.P. — thus concomitant with or subsequent to the emergence of modern humans (Enard *et al.* 2002).[2] K would do well to clarify in detail how their conclusions harmonize with the population reasoning offered in that earlier study.

That said, prospects other than Neanderthals having a complex human-like faculty of language are compatible with K's favored scenario. They themselves emphasize that uncertainties in *FOXP2*'s function in Neanderthals could only be cleared by a more complete sequence of the gene, which might uncover some further Neanderthal-specific substitutions. We agree with the skepticism this invites with regards to the putative existence of a complex form of language among Neanderthals, but our reasons are quite different.

A high degree of conservation of *FOXP2* orthologues among vertebrates has been independently established (Enard *et al.* 2002), which makes the existence of more substitutions within the complete sequence of the gene in Neanderthals rather improbable. However, the key to settle this question is not so much the complete sequence of *FOXP2* in Neanderthals, but attaining more information about the genetic context in which the gene displayed its regulatory function in this species. Unfortunately, this kind of information is rather sparse even in the case of modern humans (Spiteri *et al.* 2007, Vernes *et al.* 2007) and of course is completely non-existent in the case of Neanderthals.

Modern *FOXP2* could have become fixed in Neanderthals for reasons different from those operating in modern humans. Within a different genetic context, it could have helped regulate the development/execution of a symbolic but non-syntactic proto-language (Bickerton 1990), or some other form of quasi musical vocalizations (Mithen 2006b), among other conceivable possibilities. In

---

[2]   Coop *et al.* (2008: 1257) assert that the antiquity of the haplotype could be reduced up to 42,000 years B.P. using a different statistical procedure ('phylogenetic dating'). However, as they also alert that "there is considerable uncertainty associated with this estimate", we prefer not to make any statement based on this date.

fact, even identical mutant versions of *FOXP2* can correlate with different acoustic/prosodic phenotypes in the case of modern humans (Shriberg *et al.* 2006). Technically, this idea presupposes the existence of two parallel selective processes with identical molecular outcomes in two different species. It is difficult to assess the probability of such a scenario, but it may be a feasible one considering the genetic, anatomic and physiologic closeness of the two species, as well as the similar selective pressures they could be going through at a certain point of their evolutionary history.

Also relevant with regard to this idea is the fact that different *FOXP2* orthologues have been described for some species and associated with a distinctive ability in each case (see Haesler *et al.* 2004 and Haesler *et al.* 2007 on bird song, and Shu *et al.* 2005 and Fujita *et al.* 2008 on ultrasonic vocalization in mice). But it is worth reiterating again the high degree of conservation of this gene (Enard *et al.* 2002), which has undergone very few evolutionary changes among vertebrates.[3] Thus, in all likelihood the ability with which each *FOXP2* orthologue relates in each species is a function of the molecular context which the protein coded by the gene integrates in each particular case, and not of minor structural modifications experienced by its different variants.

Much more information is thus needed regarding the regulatory networks in which *FOXP2* is involved, and about its target genes in human development, before strong functional homologies in Neanderthals can be explored so that the 'Neanderthal language' question can be properly assessed from a paleogenetic point of view.

3.   Second Scenario

The relevant *FOXP2* haplotype could be present in the ancestral populations of Neanderthals and modern humans, but only positively selected in the latter. K reject this possibility because they detect a signal of selective sweep on the Neanderthal region under discussion.

The intronic region located in position 5 from the exon containing the modern mutations of *FOXP2* has been affected by a selective sweep. This is reflected in the low frequency of variants within different modern human populations, in a region otherwise subject to the frequency rates of a standard neutral mutation model. K observe that the analysis of some nucleotids from the same intronic regions of the El Sidrón specimens shows a high degree of identity with the predominant allele among humans. From this they conclude that the selective sweep on the region can be traced back to our last common ancestor with Neanderthals, around 260,000 years B.P.

However, it is important to note that the fact that two Neanderthal variants of the intronic region under discussion are similar to the modern allele does not necessarily entail that all Neanderthal variation concentrate on the same modern-

---

[3]   An exception seems to be the case of some species of echolocating bats, which present massive variants of *FOXP2*. Li *et al.* (2007), who relate the gene with a function in sensorimotor coordination, argue that this fact is to be explained by the divergent selective pressures these species have been subjected to in the evolution of echolocation.

like allele. Thus the signal of the selective sweep on Neanderthal *FOXP2* will only be confirmed by the analysis of a representative sample of individuals of this species.

Furthermore, Coop *et al.* (2008: 1257) argue that the following lines of evidence rule out the possibility of such an early selective process:

> (1) The persistence of ancestral alleles for close to 300,000 years in both Neanderthal and human lineages is unlikely, given that low frequency variants will tend to be rapidly lost from the population by genetic drift. Actually, ancestral alleles among modern humans are found in the intronic region under examination, as well as the ancestral allele in some intronic markers of the Neanderthal sample ; and

> (2) If the selective sweep was close to completion 300,000 years ago, the selected haplotype should have accumulated more mutations since. Actually, what is noticeable is the scarcity of divergences added up to the haplotype.

If Coop *et al.*'s conclusion is on the right track, how could one explain that the mutations under discussion are present in the genetic pools of both Neanderthals and modern humans but only selected in the latter species? The key to answer this question may be that modifications on the concerted action of *FOXP2* with other genes, in the development of an innovative cognitive structure (Piattelli–Palmarini & Uriagereka 2005), could underlie the selective sweep on the modern mutations of this regulatory gene (Lai *et al.* 2001). Selection operating on a complete module of coordinated genes (Oldham *et al.* 2006, Spiteri *et al.* 2007) in modern humans, but not in Neanderthals, could in principle co-exist with a lack of selective sweep on *FOXP2* in the latter species, even if the modern mutations are relevantly confirmed.

4.   Third Scenario

Gene flow could be the source of the two evolutionary changes in *FOXP2* that Neanderthals from El Sidrón share with modern humans. K reject this possibility on the basis of previous analyses of Neanderthal mtDNA (Krings *et al.* 1997, Krings *et al.* 1999, Hofreiter *et al.* 2001, Serre *et al.* 2004) plus their own analysis of the Y chromosomes in the two specimens from El Sidrón.

We are aware that the admixture thesis has been controversial ever since its proposal by Green *et al.* (2006).[4] Nevertheless, it is in principle possible that this state-of-affairs did take place, an thus we should consider the possibility, remote as it may be, of a scenario along these lines for the mutations that concern us here.[5] Actually, this is the preferred scenario of Coop *et al.*, who conclude that K's

---

[4]   See Noonan *et al.* (2006), and Wall & Kim (2007) for a useful comparison between the two theses.

[5]   Regarding K's arguments against admixture, we would like to briefly note the following facts:

> (1) Maternally inherited mtDNA cannot settle the question, as Neanderthal mitochondrial

results "may reflect gene flow between modern human and Neanderthal populations" (Coop *et al.* 2008: 1257).

It is true that the antiquity of the specimens from El Sidrón (around 43,000 years old) is at the limit for such a possibility. In order to shed light on this question, more analysis of specimens from different locations seems imperative, ideally earlier ones than those found in El Sidrón (in the 46,000–50,000 B.P. range or older, taking this as the approximate date of arrival of modern humans to Europe; Oppenheimer 2003).

5. Conclusion

The significance of K's finding cannot be overemphasized. However, even though they are not equally probable, none of the three scenarios commented on here, attempting to explicate such an important discovery, can be summarily discarded. Therefore, we consider that the interpretation of the facts advanced by this team is premature. Many questions still await an answer, and crucial data need to be uncovered, before anyone can assert whether Neanderthals spoke or not.

References

Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.

Coop, Graham, Kevin Bullaughey, Francesca Luca & Molly Przeworski. 2008. The timing of selection at the human *FOXP2* gene. *Molecular Biology and Evolution* 25(7), 1257-1259.

Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S.L. Lai, Victor Wiebe, Takashi Kitano, Anthony P. Monaco & Svante Pääbo. 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* 418, 869-872.

Fujita, Eriko, Yuko Tanabe, Akira Shiota, Masatsugu Ueda, Kiyotaka Suwa, Mariko Y. Momoi & Takashi Momoi. 2008. Ultrasonic vocalization impairment of *Foxp2* (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proceedings of the National Academy of Sciences of the United States of America* 105, 3117-3122.

---

haplotypes would be left unchanged by a modern-males-to-Neanderthal-females genetic channel (actually, Green *et al.* 2006 contend that gene flow must have predominantly stemmed from modern human males); and

(2) The analysis of Y chromosome fragments in the El Sidrón specimens is not conclusive in this regard. Clear results have been obtained from only one specimen, and are not identical in each of the laboratories the team worked with (Krause *et al.* 2007, Supplemental Data). While the outcome of the analysis does suggest that no derived alleles exist in the Y chromosome of that one specimen, the possibility cannot be excluded with such sparse data.

Green, Richard E., Johannes Krause, Susan E. Ptak, Adrian W. Briggs, Michael T. Ronan, Jan F. Simons, Lei Du, Michael Egholm, Jonathan M. Rothberg, Maja Paunovic & Svante Pääbo. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330-336.

Haesler, Sebastian, Christelle Rochefort, Benjamin Georgi, Pawel Licznerski, Pavel Osten & Constance Scharff. 2007. Incomplete and inaccurate vocal imitation after knockdown of *FoxP2* in songbird basal ganglia nucleus Area X. *Public Library of Science Biology* 5(12), e321, doi:10.1371/journal.pbio. 0050321.

Haesler, Sebastian, Kazuhiro Wada, A. Nshdejan, Edward E. Morrisey, Thierry Lints, Eric D. Jarvis & Constance Scharff. 2004. *Foxp2* expression in avian vocal learners and non-learners. *Journal of Neuroscience* 24, 3164-3175.

Hofreiter, Michael, David Serre, Hendrik Poinar, Melanie Kuch & Svante Pääbo. 2001. Ancient DNA. *Nature Reviews* 2, 353-359.

Klein, Richard G. with Blake Edgar. 2002. *The Dawn of Human Culture*. New York: John Wiley and Sons.

Krause, Johannes, Charles Lalueza–Fox, Ludovic Orlando, Wolfgang Enard, Richard E. Green, Hernán A. Burbano, Jean-Jacques Hublin, Catherine Hänni, Javier Fortea, Marco de la Rasilla, Jaume Bertranpetit, Antonio Rosas & Svante Pääbo. 2007. The derived *FOXP2* variant of modern humans was shared with Neandertals. *Current Biology* 17, 1908-1912.

Krings, Matthias, Helga Geisert, Ralf W. Schmitz, Heike Krainitzki & Svante Pääbo. 1999. DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. *Proceedings of the National Academy of Sciences of the United States of America* 99, 5581-5585.

Krings, Matthias, Anne Stone, Ralf W. Schmitz, Heike Krainitzki, Mark Stoneking & Svante Pääbo. 1997. Neandertal DNA sequences and the origins of modern humans. *Cell* 90, 19-30.

Lai, Cecilia S.L., Simon E. Fisher, Jane A. Hurst, Faraneh Vargha-Khadem & Anthony P. Monaco. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519-523.

Li, Gang, Jinhong Wang, Stephen J. Rossiter, Gareth Jones & Shuyi Zhang. 2007. Accelerated *FoxP2* evolution in echolocating bats. *Public Library of Science ONE* 2(9), e900, doi:10.1371/journal.pone.0000900.

Mellars, Paul. 2005. The impossible coincidence: A single-species model for the origins of modern human behavior. *Evolutionary Anthropology* 14, 12-27.

Mithen, Steven. 2006a. General intellectual ability. In Steven W. Gangestad & Jeffrey A. Simpson (eds.), *The Evolution of Mind: Fundamental Questions and Controversies*, 319-324. New York: The Guilford Press.

Mithen, Steven. 2006b. *The Singing Neandertal. The Origins of Music, Language, Mind, and Body*. Cambridge: Cambridge University Press.

Noonan, James P., Graham Coop, Sridhar Kudaravalli, Doug Smith, Johannes Krause, Joe Alessi, Feng Chen, Darren Platt, Svante Pääbo, Jonathan K. Pritchard & Edward M. Rubin. 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* 314, 1113-1118.

Oldham, Michael C., Steve Horvath & Daniel H. Geschwind. 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee

brains. *Proceedings of the National Academy of Sciences of the United States of America* 103, 17973-17978.

Oppenheimer, Stephen. 2003. *Out of Eden. The peopling of the world*. Robinson Publishing, London.

Piattelli–Palmarini, Massimo & Juan Uriagereka. 2005. The evolution of the narrow faculty of language: the skeptical view and a reasonable conjecture. *Lingue e Linguaggio* IV.1, 27-79.

Rosas, Antonio, Cayetana Martínez–Maza, Markus Bastir, Antonio García–Tabernero, Carles Lalueza–Foz, Rosa Huget, José Eugenio Ortiz, Ramón Juliá, Vicente Soler, Trinidad de Torres, Enrique Martínez, Juan Carlos Cañaveras, Sergio Sánchez–Moral, Soledad Cuezva, Javier Lario, David Santamaría, Marco de la Rasilla & Javier Fortea. 2006. Paleobiology and comparative morphology of a late Neandertal sample from El Sidrón, Asturias, Spain. *Proceedings of the National Academy of Sciences of the United States of America* 103, 19266-19271.

Serre, David, André Langaney, Mario Chech, Maria Teschler–Nichola, Maja Paunovic, Philippe Mennecier, Michael Hofreiter, Göran Possnert & Svante Pääbo. 2004. No Evidence of Neandertal mtDNA contribution to Early Modern Humans. *Public Library of Science Biology* 2, 313-317.

Shriberg, Laurence D., Kirrie J. Ballard, J. Bruce Tomblin, Joseph R. Duffy, Katherine H. Odell & Charles A. Williams. 2006. Speech, prosody, and voice characteristics of a mother and daughter with a 7,13 translocation affecting *FOXP2*. *Journal of Speech, Language and Hearing Research* 49, 500-525.

Shu, Weiguo, Julie Y. Cho, Yuhui Jiang, Minhua Zhang, Donald Weisz, Gregory A. Elder, James Schmeidler, Rita De Gasperi, Miguel A. Gama Sosa, Donald Rabidou, Anthony C. Santucci, Daniel Perl, Edward Morrosey & Joseph D. Buxbaum. 2005. Altered ultrasonic vocalization in mice with a disruption in *Foxp2* gene. *Proceedings of the National Academy of Sciences of the United States of America* 102, 9643-9648.

Spiteri, Elisabeth, Genevieve Konopka, Giovanni Coppola, Jamee Bomar, Michael Oldham, Jing Ou, Sonja C. Vernes, Simon E. Fisher, Bing Ren & Daniel H. Geschwind. 2007. Identification of the transcriptional targets of *FOXP2*, a gene linked to speech and language, in developing human brain. *American Journal of Human Genetics* 81, 1144-1157.

Vernes, Sonja C., Elisabeth Spiteri, Jérôme Nicod, Matthias Groszer, Jennifer M. Taylor, Kay E. Davies, Daniel H. Geschwind & Simon E. Fisher. 2007. High-throughput analysis of promoter occupancy reveals direct neural targets of *FOXP2*, a gene mutated in speech and language disorders. *American Journal of Human Genetics* 81, 1232-1250.

Wall, Jefrrey D. & Sung K. Kim. 2007. Inconsistencies in Neanderthal genomic DNA sequences. *Public Library of Science Genetics* 3(10), 1862-1866.

Antonio Benítez-Burraco
Universidad de Oviedo
Área de Lingüística
Campus El Milán s/n
33011 Oviedo
Spain
benitezantonio.uo@uniovi.es

Guillermo Lorenzo
Universidad de Oviedo
Área de Lingüística
Campus El Milán s/n
33011 Oviedo
Spain
glorenzo@uniovi.es

Víctor M. Longa
Universidade de Santiago de Compostela
Área de Lingüística
Praza Isabel A Católica 2-2E
36204 Vigo
Spain
victormanuel.longa@usc.es

Juan Uriagereka
University of Maryland
Department of Linguistics
1401 Marie Mount Hall
College Park, MD 20742
USA
juan@umd.edu

Combinatorics for Metrical Feet

William J. Idsardi

1. Introduction

Halle & Vergnaud (1987) propose a convention on the parsing of elements into metrical feet — the Exhaustivity Condition — that requires all elements to belong to some foot, except for certain principled cases of extrametricality. However, the general consensus now prevailing is that even internal metrical elements can remain unparsed, failing to belong to any foot, generalizing the notion of extra-metricality. Hayes (1995), Halle & Idsardi (1995), and Kager (1999), among many others, explicitly reject the Exhaustivity Condition. Hayes's comments are given in (1), Halle & Idsardi's are given in (2).

(1)   "The upshot seems to be that in our present state of knowledge, it would be aprioristic to adhere firmly to a rigid principle of exhaustive prosodic parsing […]."   (Hayes 1995: 110)

(2)   "We also deviate from previous metrical theories by not requiring exhaust-ive parsing of the sequence of elements, that is we do not require that every element belong to some constituent […]."   (Halle & Idsardi 1995: 440)

In this squib I will prove that the number of possible metrical parsings into feet under these assumptions for a string of n elements is $Fib(2n)$ where $Fib(n)$ is the nth Fibonacci number.

2. Initial Observations

Disregarding prominence relations within the feet (that is, headedness), the possible footings for strings up to a length of three elements are shown in (3). Feet are indicated here by matching parentheses; elements not contained within parentheses are unfooted (that is, 'unparsed' in Optimality Theory terminology).

(3)   a.   1 element, 2 possible parsings: (x), x
       b.   2 elements, 5 possible parsings: (xx), (x)(x), (x)x, x(x), xx
       c.   3 elements, 13 possible parsings: (xxx), (xx)(x), (xx)x, (x)(xx), x(xx),
                                                  (x)(x)(x), (x)(x)x, (x)x(x), x(x)(x),
                                                  (x)xx, x(x)x, xx(x), xxx

The number of possible footings is equal to every other member of the Fibonacci sequence, illustrated and defined as a recurrence relation in (4); see, for example, Cameron (1994).

(4)   Fibonacci sequence: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, …
        Fib(0) = Fib(1) = 1; for *n > 1 Fib(n) = Fib(n–1) + Fib(n–2)*

There is only one possible footing of a string of zero elements, so that it is also the case that the number of footings of zero elements is equal to *Fib(0).*

3.   Proof

Let *f(n)* be the number of parsings of a string of *n* elements into metrical feet, not subject to the Exhaustivity Condition. We can derive a recurrence relation for the number of metrical feet in a string of length *n+1* by dividing the string after the places where an initial foot could occur, as shown in (5).

(5)   a.   no initial foot: x | …, *n* elements left, therefore *f(n)* footings
        b.   1–element foot: (x) | …, *n* elements left, therefore *f(n)* footings
        c.   2–element foot: (xx) | …, *n–1* elements left, therefore *f(n–1)* footings
        d.   3–element foot: (xxx) | …, *n–2* elements left, therefore *f(n–2)* footings
              …
        e.   n–element foot: (x…x) | , 0 elements left, therefore *f(0)* = 1 footing

        Generally then, $f(n+1) = f(n) + \sum_{i=0}^{n} f(i)$

We then prove the general relation by induction on *n*. That is, we have shown by direct calculation that the relation holds for *n* = 0, 1, 2 and 3, (3), and now, assuming that *f(i) = Fib(2i)* for *i* up to and including *n*, we will prove that *f(n+1) = Fib(2n+2)*. We begin with the recurrence relation derived in (5), pulling out the *n*[th] term of the summation, shown in (6).

(6) $f(n+1) = f(n) + f(n) + \sum_{i=0}^{n-1} f(i)$

Substituting for *f(n)* using the induction assumption gives (7).

(7) $f(n+1) = Fib(2n) + Fib(2n) + \sum_{i=0}^{n-1} f(i)$

Substituting for *Fib(2n)* using the Fibonacci recurrence relation gives (8).

(8) $f(n+1) = Fib(2n) + Fib(2n-1) + Fib(2n-2) + \sum_{i=0}^{n-1} f(i)$

Substituting for *Fib(2n–2)* again using the induction assumption gives (9).

(9) $f(n+1) = Fib(2n) + Fib(2n-1) + f(n-1) + \sum_{i=0}^{n-1} f(i)$

Substituting for the last two terms using the *f(n)* recurrence relation gives (10).

(10) $f(n+1) = Fib(2n) + Fib(2n-1) + f(n)$

Substituting for *f(n)* again using the induction assumption gives (11).

(11) $f(n+1) = Fib(2n) + Fib(2n-1) + Fib(2n)$

Substituting the first two terms using the Fibonacci recurrence relation gives (12).

(12) $f(n+1) = Fib(2n+1) + Fib(2n)$

Again substituting using the Fibonacci recurrence relation gives (13), as required.

(13) $f(n+1) = Fib(2n+2)$ Q.E.D.

Having proved that if *f(n) = Fib(2n)* then *f(n+1) = Fib(2n+2)* for *n* > 1, and having *f(0) = Fib(0)* and *f(1) = Fib(2)*, we have proved the relation for all non-negative *n*.

4.   A Corollary

Given the above proof, substituting into the footing recurrence relation gives (14).

(14) $f(n+1) = f(n) + \sum_{i=0}^{n} f(i)$

$Fib(2n+2) = Fib(2n) + \sum_{i=0}^{n} Fib(2i)$

And, since from the Fibonacci recurrence relation we have *Fib(2n+2) = Fib(2n+1) + Fib(2n)*, therefore we derive (15).

(15) $Fib(2n+1) = \sum_{i=0}^{n} Fib(2i)$

That is, for example, *Fib(7) = Fib(6) + Fib(4) + Fib(2) + Fib(0)* = 13 + 5 + 2 + 1 = 21.

5.   Conclusion

The number of non-exhaustive parsings of n elements into metrical feet (i.e. the number of non-exhaustive partitions of n elements) has been proven to be equal to *Fib(2n)*, the $2n^{\text{th}}$ Fibonacci number.

References

Cameron, Peter J. 1994. *Combinatorics: Topics, Techniques, Algorithms*. Cambridge: Cambridge University Press.

Halle, Morris & William J. Idsardi. 1995. General properties of stress and metrical structure. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory* (Blackwell Handbooks in Linguistics 1), 403-443. Malden, MA: Blackwell.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An Essay on Stress* (Current Studies in Linguistics 15). Cambridge, MA: MIT Press.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies.* Chicago: University of Chicago Press.

Kager, Rene. 1999. *Optimality Theory* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

William J. Idsardi
University of Maryland
Department of Linguistics
1401 Marie Mount Hall
College Park, MD 20742
USA
idsardi@umd.edu

★ REVIEWS ★

Novel Tools at the Service of Old Ideas

By Massimo Piattelli–Palmarini

I wholeheartedly endorse one central idea in this book and the motivation behind it. Eva Jablonka and Marion J. Lamb (henceforth J&L) make it very clear that a multiplicity of stunning advances in biology and in evolutionary theory in the last several years have so completely reshaped the standard neo-Darwinian picture that, indeed, cognitive scientists should pay attention and re-think many of their ideas about the evolution of cognition. The main facts and ideas of this new biology are explained very well by J&L, as they are in two recent excellent books, also fully accessible to a lay audience (Kirschner & Gerhart 2005 and Carroll 2005). There is a lot to be learned in this essay about new ideas in biology and in modern evolutionary theory. Having said this, I wish to trace a sharp divide between J&L's excellent exposition of biology and their objectionable picture of language evolution. Before I explain why, I need to insert one important consideration.

1.   A Missing Dimension (the 5th?)

All the ideas and experiments in biology that are detailed in this book are the right ones. There is not one of them that I would have liked to see left out. There is, however, a glaring lacuna: no mention of the powerful return of the laws of form in biology, of the central role that physico-chemical and computational factors play in the optimization of biological functions and assemblies. J&L's pages dedicated to Waddington could have been the right entry into this domain, but they are focussed on Waddington's interesting ideas about development and complex patterns of selection. Emphasis on global invariants and on the morpho-genetic power of the laws of physics and chemistry goes back to Wentworth D'Arcy Thompson and Alan M. Turing (Thompson 1917/1992, Turing 1952), but it has come back in force in the last few years. In J&L's tally, it should be conceived as the fifth dimension in evolution.

There is only so much that the 25,000 or so genes in the human genome *can* do to assemble a human being. Sure, as J&L explain in detail, there are multiple gene regulations and networks of interactions, and morphogenetic attractors, and

epigenetic modifications, and a complex interaction with culture. But this is far from being enough. Among other complex structures, tens of millions of kinds of antibodies have to be produced, and 10^{11} neurons and 10^{13} synapses to be developed and fixated, and about 60,000 miles of veins, arteries and capillaries to be exactly placed in each of our bodies. Christopher Cherniak has introduced the notion of "non-genomic nativism" and has shown by means of extensive computer calculations that the wiring of the cerebral cortex is the most efficient among, literally, billions of conceivable alternatives (see Cherniak *et al.* 2004, Cherniak 2005). The maximization of connection density in the cerebral cortex is even better than in the best industrial micro-chips.

On a different, but converging, front, West, Brown & Enquist (1997, 1999) have shown that the "multiples of 1/4th" power laws that govern the scaling of metabolic activities, membrane fluxes, heart beat, blood circulation lifetime, and life span, from unicellular organisms all the way up to whales, can only be explained by universal fractal laws. Symptomatically, they also used, years before J&L, in their title, the expression "4th dimension of life", explaining that natural selection has exploited variations on this fractal theme to produce the incredible variety of biological form and function. There are genes, of course, but also severe geometric and physical constraints on metabolic processes.

A brief list of discoveries in this fast progressing sector must also include the work by Bejan & Marden (2006) on universal invariants of locomotion. Starting with general principles of physics and engineering, they have shown that the optimal speed and frequency of locomotion (be it walking, swimming, crawling, or flying), for unit of biological energy spent, scales linearly with the size of animals from fruit flies to whales. Other interesting applications of general physical principles to biological functions and structures cover optimal foraging in bees (Dechaume–Moncharmont *et al*. 2005), the neuronal regulation of singing in birds (Trevisan, Mindlin & Goller 2006), and the optimal character of the genetic code. Among thousands of possible alternatives, the genetic code as we know it is optimal for minimizing the effect of frame-shift mutations and minimizing the energy wasted in synthesizing the start of anomalous protein sequences (Itzkovitz & Alon 2007). It is perhaps ungracious to reproach a lacuna to the authors of such a rich and diverse book, but their complete neglect of this entire crucial dimension of evolution (the 4th or 5th, depending on how you count them) deserves to be signaled and lamented. Neglect of this dimension also reverberates negatively onto J&L's treatment of language and evolution.

2.   Symbols? Oh, No, Please!

As of Chapter 6, I start to disagree with J&L. They follow a very old script, one that opens up with the appearance of symbolic systems. They duly acknowledge that language is special, with respect to other symbolic communication systems found in animals, essentially because of the subtlety of syntax. That is correct, but there is more to be said. Other crucial differences are to be found already at the level of the lexicon. It's not just syntax that makes human language special, but also the nature of individual words and the way they connect with each other and with the world. There are at least four major differences between words and

all non-linguistic symbols: (A) aspectual reference, (B) headedness, (C) internal structure, and (D) edge features. Briefly about each one in turn:

(A) *Buy* and *sell*, *fear* and *frighten*, and a huge variety of such oppositions, in all languages, refer to a same objective, physical, filmable, state of affairs, but have transparently different meanings. The same applies to nouns (*destruction* vs. *demolition*, *gift* versus *theft*) and to adjectives (*thrifty* vs. *stingy*, *abundant* versus *excessive*, and so on). Even apparently innocent words like *city* embody an aspectual component, a point of view. Words refer only under specific itineraries of mental access (a city can be said to be chaotic, polluted, expensive, mostly Victorian, each expression obviously referring to very different objective features; cf. Chomsky 2005). Word meanings are through and through intensional. No symbol used in animal communication systems has this property. Also many non-linguistic symbols used by humans to communicate lack it, unless they are transparently parasitic on language.

(B) 'The California highway commissioner report' is a report. 'The world trade exchange bank' is a bank. 'The spy who came in from the cold' is a spy. The rightmost noun (in English, the leftmost in other languages) heads all nominal compounds. A noun with a determiner (such as *the spy*) heads the Determiner Phrase, even when the DP contains a whole sentence (*who came in from the cold*). Headedness also applies to Verb Phrases (in a more complicated way which need not detain us here; see below). The property of headedness is conserved by the syntactic derivation, from start to finish, and cannot be altered. It's a crucial combinatorial valency of lexical entries, determining the category to which they belong and how the syntactic machinery *must* treat them. There are, of course, many ways to make a certain symbol particularly salient in a string of non-verbal symbols (size, color, etc.), but headedness is unique to words.

(C) Words have a rich internal structure. Thematic roles are probably the most conspicuous such structures. There was *the destruction of Carthage by Scipio*, but there cannot be **the sleep of the bed by Scipio*. Together with headedness, thematic roles are crucial valencies for combination into larger expressions. Morphological domains within words are also central, with relations of dominance and asymmetry. Vast, subtle, and ramified consequences of this internal structures ensue for syntax and semantics (Halle & Marantz 1993, di Sciullo 2005). No other system of non-linguistic symbols has any semblance of such property.

(D) Very simply said, words are "sticky" and so are phrasal constituents obtained by merging two of them, and then merging this compound with other words, again and again, recursively and hierarchically. (The technical term for this intrinsic combinatorial power of words and phrasal constituents in the minimalist program is "edge features"; Chomsky has rightly stressed that the appearance of edge features has been one of the central events in the evolution of language.) Whole linguistic expressions, and

sentences in particular, are not lists of words, not even *ordered* lists of words. The point I wish to emphasize here is that words have the intrinsic capacity to project structure "upwards" onto larger compounds. Verbs offer the richest case, but not the only one. Verbs project a stratification of "shells" in a fixed hierarchical order, specifying the place where to insert the actants, the auxiliaries, the checking of tense, Case and agreement, and more (ever since the seminal work of Richard Larson — cf. Larson 1988).

All in all, therefore, contrary to spontaneous intuition, contrary to the whole domain of semiotics, and contrary to what Chapter 6 and Chapter 9 of J&L suggest, there is *no gain in our understanding of language* by assimilating it to a system of symbols. Any attempt to reconstruct language evolution as the evolution of a symbolic system leads us badly astray. Words are, of course, in some sense, symbols, and they enter into the system of language, but the unique properties summarized here above make words stand radically apart from all other symbolic systems. J&L, unbeknownst to them, seal this radical separation in the last line of their table on p. 234, when they state that the "range of variation" of symbolic systems is "unlimited". I doubt that they are right even about symbolic systems, but surely this does not apply to language. The range of variation for language is quite severely limited, as J&L sketch in Chapter 8, sort of noncommittally, when speaking of the "principles and parameters" model (Baker 2001, 2003). Symbolic systems are not relevant to language, and they cannot be offered as an intermediate step in language evolution.

3.   Culture and Language

J&L embrace a thesis that several other authors also have tried to promote: the shaping of language by culture and history. Their critique of the innatist, modularist, and highly specific nature of language has, as is often the case with those who adopt their position, a possibilistic attitude: Why could we not, one day, explain a lot in language by means of cultural and historical factors, communicative functions, motor control, and general intelligence? This line was offered over 30 years ago already by Jean Piaget to Noam Chomsky, in a direct debate (Piattelli–Palmarini 1980). The answer is today what it was then: No one can exclude this possibility, as a remote possibility. It is, however, eminently rational to expect that it will *not* happen. The task seems even more hopeless today than it seemed 35 years ago, because we know a lot more about language than we did then. For instance, none of the properties of words that I have sketched above can be explained in terms of culture or history, or motor control, or factors of general intelligence.

On p. 218, J&L venture into a minefield, quite similar to the one into which Michael Arbib also ventured in *BBS* recently (Arbib 2005) — a parallel between language and mathematics:

> Although the speed and ease of learning [of language by the child] may indicate that there are some preexisting specifically selected neural mechanisms, the same properties could also be due to a culturally evolved system that is well adapted to the brain, and therefore makes learning easy. For example, think how difficult it was 1200 years ago for someone in Europe to

> divide one number by another. Say they wanted to divide 3712 by 116 […]
> [*they point to the impracticality of the Roman numerals — MPP*] Today, with our
> Arabic notation system (and the useful zero), it would take the average ten-
> year-old only minutes to get the answer 32.

No genetic change, no brain change, but rather a cultural invention that has
become common knowledge. J&L advocate (like Arbib and Deacon and
Tomasello) a co-evolution of brain and language and do not advocate a *purely*
cultural-evolution explanation of the language capacity. Well, anyway, their
analogy with the numerical division is totally irrelevant. No sentence in any
language requires "minutes" to be understood by a ten-year-old, or by anyone at
any age. Aside from the fact that ten years is a very old age for language,
sentences are processed in fractions of seconds, not minutes, today just as they
were 1200 years ago, or earlier. Moreover, the number system and the rules for
dividing numbers have to be explicitly and painfully taught. No three-year-old
child today can make that division, while he or she can well understand quite
subtle syntactic constructions, exactly like a child could already in ancient Egypt.
The analogy is infelicitous, because language is in a completely different ball-
park. Like this one, many analogies and thought-experiments offered by J&L in
the domain of language are inconsequential or misleading, unlike those that deal
with biology proper.

4.   New Biology and Old Reflexes

A most puzzling aspect of this book is that, after having pleaded persuasively for
a major expansion of concepts and models in evolutionary theory, J&L fall back
onto a basically classic, neo-Darwinian, functionalist explanation of the evolution
of language. Just as an example, on p. 339 we read:

> Two related sets of conditions seem to have pushed our ancestors along the
> route to language. The first was an altered ecological and social
> environment, which provided a strong and persistent motivation for better
> communication […]. The second and related set of conditions has to do with
> anatomy and physiology. […] It was probably the increased motor control
> over hand movements and vocalizations, and the ability to imitate both
> gestures and vocal sounds.

They are in excellent and very old company in making these hypotheses, from
Darwin himself, to Jean Piaget, Philip Liberman, Steven Pinker, Paul Bloom,
Michael Arbib, and Derek Bickerton, just to name a few. Yet, all that we have
learned from the new biology, and from this very book, should make any such
functionalist hypothesis unnecessary or even suspect. Master regulatory genes
with pleiotropic effects, transposons, gene duplications, histone modification,
and alternative gene splicing (just to mention a few) offer manifold evolutionary
mechanisms that make progressive functional adaptation quite marginal. But
J&L insist, venturing into "non-genetic inheritance" to explain how "various
features of the emerging language system *that were initially culturally transmitted*
were later genetically assimilated" (p. 340, my emphasis). I have no qualm with
non-genetic inheritance, amply attested in experiments well explained in their
own previous chapters and also endorsed by Cherniak's "non-genomic nativism"

(which J&L ignore — see *supra*), but I strongly object to the cultural transmission hypothesis.

Hauser, Chomsky & Fitch (2002) have rightly insisted on the uniqueness of the capacity of humans to acquire a lexicon, and on the presence in humans of syntactic computational powers that are conspicuously absent in other primates (Fitch & Hauser 2004). Together with the very special properties of words seen above, these are quantum changes in cognitive powers, both qualitatively and quantitatively, impossible to reconstruct by piecemeal functional adaptation. Cultural interactions among humans that are allowed by language presuppose them and cannot explain their gradualistic adaptive origin. The new evolutionary mechanisms presented in this book could have finally dispensed us from exploring again an old dead-end.

The surprising reappearance of old, standard neo-Darwinism is also to be witnessed when J&L criticize the approach promoted by Hauser, Chomsky & Fitch in an already famous (or infamous, for some; cf. Pinker & Jackendoff 2005) paper published in 2002 (Hauser, Chomsky & Fitch 2002). They surprisingly repeat *en passant* the most routine neo-Darwinian objections.

I must also point out that in Chapter 9, J&L choose to tell us the story of the chimp Kanzi and the data collected by Sue Savage–Rumbaugh, allegedly showing important continuity between the symbolic system mastered by apes (after long training) and human language. They fail to even mention the case of the chimpanzee Nim Chimpsky which led to drastically opposite conclusions. After several years of daily cohabitation and of daily sessions of several hours trying to teach Nim American Sign Language, Laura Petitto, Herbert Terrace, and Thomas G. Bever concluded that no real progress had been made. This momentous piece of work (Terrace *et al.* 1979) as well as the papers and book by David Premack (Premack 1972, 1986), that for many of us closed the chapter of the search for animal language, should at least have been presented, if only to criticize them.

5.   Language

On the basis of previous work by Eva Jablonka and Daniel Dor, a variant of the co-evolution of brain and language, or rather (very importantly to J&L) language, brain, *and culture* is offered. As usual, in this kind of literature, they indulge in imagining various spiraling interactions between social organization, individual cognition, brain evolution, and language. Michael Arbib has given us his spirals, J&L now give us theirs. The problem, again and again, is that, if you take just any article at random, say, in the journal *Linguistic Inquiry* over the last 20 years or so, and look at the data, just the data (forget about the explanations), there is no hope whatsoever for J&L not only of explaining those data, but even of saying something that is remotely relevant.

While many interesting details are provided about experiments in biology, no specific data are presented in the case of language. Nowhere are we told *how* cultural transmission and the function of communication and general intelligence and motor control *can* have shaped language as we know it. On p. 305 we come as close to a specific hypothesis as their approach allows:

> […] Dror and other linguists have found [that] the grammatical structure of phrases and sentences is associated with the types of concepts the words in sentences embody […]. For example, the grammatical patterns we use depend on whether the participant in an event are active or inactive, on whether an action leads to a change in state of the object or it does not; on whether events are factual or hypothesized; on whether things are countable or not countable […] and so on.

J&L then point out, correctly, that "although there are endless ways of classifying things, events, properties, and so on, the categories that are reflected in differences in grammatical patterns are only a small set of all those that we could use".

It's hard to disagree with this. The paucity of syntactic theta-roles, with respect to all the things we are interested in in our life, is one of the central observations in linguistics (the most insightful and influential treatment is Hale & Keyser 1993, 2002). Several deep explanations have been given in generative grammar (theta theory, X-bar theory, the semantics of count ad mass terms, event semantics, the theory of aspect, the theory of telicity, internal structures in lexical semantics, and so on, not to mention the rich theory of concepts and of concept acquisition by the child).

The rub comes next (p. 306):

> What Dror concludes from this is that language is structurally designed to communicate some things better than others. Its design enables it to deal well with messages that are grounded in a rather constrained set of categories having to do with events and situations, their time and place, and the participants in them, *all of which are reflected in grammatical structures* [my emphasis — MPP].

Sorry, but it's not so. Just to take a few signal examples, the sources of objects, the motivations of actions, the banality versus the exceptionality of events — all things we do care a lot about — are *not* reflected in grammatical structures. The endpoint of an action and the culmination of an event are routinely and subtly encoded in syntax, but no syntactic device exists, in any language, to encode the beginning of an action or the initial event. We can talk about them, of course, but no structure in grammar "reflects" them. Grammatical structure is *only* sensitive to actor, patient (or theme, more generally) and in some cases the instrument or the modality of action. Period. *Bottle the wine, shelve the books,* and similar verbs incorporate the instrument or the modality. *Climb, hop, drag, attain* incorporate the path or the telos or the modality of motion. Marginal, but admissible, constructions like *we laughed the bad actor off the scene, John smiled the girl into his house*, and similar ones allow to syntactically encode modality or causality. Grammar has no place for more than this. For everything else, we have to go paratactic (use adjunctions, circumlocutions, add further separate sentences, develop a whole discourse, and so on). Grammatical structures do not "reflect" what Dror and J&L want us to believe.

Moreover, in many cases, grammar is a *hindrance* to communication. There are things we would very much like to say, but grammar does not allow us to:

(1)   a.   *Who was it apparent yesterday that Jay saw?
       b.   *Who do you wonder how solved the problem?
       c.   *This is the student who I wonder what bought.

It would be nice to be able to communicate such simple thoughts in such simple ways, but grammar blocks these constructions. Many examples of how different languages manage to *overcome* these straightjackets of grammar are to be found in (Lightfoot 2000).

Another glaring case is ambiguity, a severe hindrance to communication. Not only grammar cannot resolve it in many cases, but sometimes *forces* it on us. It can do nothing to obviate the ambiguity of sentences like:

(2)   To who did you say we should tell the truth?

Is the question about the saying or about truth-telling? Grammar bars the quick insertion of disambiguation. We cannot say either (3a) or (3b):

(3)   a.   * To who did you say to who we should tell the truth?
       b.   * To who did you say we should tell the truth to who?

Many other examples are abundant in all languages. The explanation of this impossibility is strictly grammatical, and deep and complex (Rizzi 2004, Folli & Harley 2006). Grammar often clashes with our needs to communicate, and so be it. Communication must bow to grammar, not vice versa. Grammar does not "reflect" the narrow sub-set of thinkables we especially care for. It shapes a further sub-sub-set of these, in ways that are proprietary, letting general thoughts, culture, and history fend for themselves.

6.   Summing Up

The prima facie appealing and almost irresistible hypothesis that the need to communicate has shaped the evolution of language is countered by a huge corpus of data collected in many languages and dialects. The deep and complex and detailed (and far from final) explanations advanced for these subtle facts about language in generative grammar (but also, competitively, advanced in neighboring fields such as Head-Driven Phrase Structure Grammar, Lexical Functional Grammar, Tree-Adjoining Grammar, and even, to some extent, Culicover and Jackendoff's "Simpler Syntax") are alien to all conjectures based on cultural transmission, pressures from communicability or general intelligence.

The wonderful developments of the new biology should have suggested that something else can and should be sought. This book, alas, shows that even accurate knowledge of the new biology is not sufficient to urge a radical re-conceptualization of the evolution of language. J&L use their panoply of new evolutionary mechanisms only to try to improve the most canonical hypotheses about language evolution. It remains to be hoped that the readers of this fine exposition of the new biology will use the many eye-openers to be found in it to explore on their own quite different avenues to the evolution of language.

References

Arbib, Michael A. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and*

Brain Sciences 28, 105-167.

Baker, Mark C. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.

Baker, Mark. C. 2003. Linguistic differences and language design. *Trends in Cognitive Science* 7(8), 349-353.

Bejan, Adrian & James H. Marden. 2006. Unifying constructal theory for scale effects in running, swimming and flying. *The Journal of Experimental Biology* 209, 238-248.

Carroll, Sean B. 2005. *Forms Most Beautiful: The New Science of Evo Devo*. New York: Norton.

Cherniak, Christopher, Zekeria Mokhtarzada, Raul Rodriguez–Esteban & Kelly Changizi. 2004. Global optimization of cerebral cortex layout. *Proceedings National Academy of Sciences* 101, 1081-1086.

Cherniak, Christopher. 2005. Innateness and brain-wiring optimization: Non-genomic nativism. In Antonio Zilhao (ed.), *Cognition, Evolution, and Rationality*, 103-112. London: Routledge.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1-22.

Dechaume–Moncharmont, Francois-Xavier, Anna Dornhaus, Alasdair I. Houston, John M. McNamara, Edmund J. Collins & Nigel R. Franks. 2005. The hidden cost of information in collective foraging. *Proceedings of the Royal Society, B: Biological Sciences* 272(1573), 1689-1695.

Di Sciullo, Anna-Maria. 2005 *Asymmetry in Morphology* (Linguistic Inquiry Monograph 46). Cambridge, MA: MIT Press.

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303(5656), 377-380.

Folli, Raffaella & Heidi Harley. 2006. What language says about the psychology of events. *Trends in Cognitive Sciences* 10(3), 91-92.

Hale, Kenneth & Samuel Jay Keyser. 1993. On argument structure and the lexical representation of semantic relations. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (Current Studies in Linguistics 24), 53-110. Cambridge, MA: MIT Press.

Hale, Kenneth & Samuel Jay Keyser. 2002. *Prolegomenon to a Theory of Argument Structure* (Linguistic Inquiry Monograph 35). Cambridge, MA: MIT Press.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (Current Studies in Linguistics 24), 111-176. Cambridge, MA: MIT Press.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What it is, who has it, and how did it evolve? *Science* 298, 1569-1579.

Itzkovitz, Shalev & Uri Alon. 2007. The genetic code is nearly optimal for allowing arbitrary additional information within protein-coding sequences. *Genome Research*, doi: 10.1101/gr.5987307.

Kirschner, Marc W. & John C. Gerhart. 2005. *The Plausibility of Life: Resolving Darwin's Dilemma*. New Haven, CT: Yale University Press.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19, 335-391.

Lightfoot, David. 2000. The spandrels of the linguistic genotype. In Chris Knight & Michael Studdert–Kennedy & James C. Hurford (eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, 231-247. Cambridge: Cambridge University Press.

Piattelli–Palmarini, Massimo (ed.). 1980. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95, 201-236.

Premack, David. 1972. Language in chimpanzees? *Science* 172, 808-822.

Premack, David. 1986. *Gavagai! or the Future History of the Animal Language Controversy* (Learning, Development, and Conceptual Change). Cambridge, MA: MIT Press.

Rizzi, Luigi. 2004. On the study of the language faculty: Results, developments and perspectives. *The Linguistic Review* 21(3-4), 323-344.

Terrace, Herbert S., Laura-Ann Petitto, R.J. Sanders & Thomas G. Bever. 1979. Can an ape create a sentence? *Science* 206, 891-902.

Thompson, D'Arcy Wentworth. 1917/1992. *On Growth and Form* [abridged edn., prepared by John Tyler Bonner]. Cambridge: Cambridge University Press.

Trevisan, Marcos A., Gabriel B. Mindlin & Franz Goller. 2006. Nonlinear model predicts diverse respiratory patterns of birdsongs. *Physical Review Letters 96* (10 January 2006), 058103-058101-058104.

Turing, Alan M. 1952. The chemical bases of morphogenesis. *Philosophical Transactions of the Royal Society* B 237, 37-72. [Reprinted in Turing, Alan M. 1992. *Morphogenesis*. Amsterdam: North Holland.]

West, Geoffrey B., James H. Brown & Brian J. Enquist. 1997. A general model for the allometric scaling laws in biology. *Science* 276, 122-126.

West, Geoffrey B., James H. Brown & Brian J. Enquist. 1999. The fourth dimension of life: Fractal geometry and allometric scaling of organisms. *Science* 284, 1677-1679.

Massimo Piattelli–Palmarini
University of Arizona
Department of Cognitive Science
& Department of Linguistics
& Department of Psychology
P.O. Box 210028
Tucson, AZ 85721
USA
massimo@u.arizona.edu

Some Questions about Determining Causal Inference and Criteria for Evidence: Response to Ladd, Dediu & Kinsella (2008)

Joshua Bowles

1. Introduction

Ladd, Dediu & Kinsella (2008; LDK from now on) and Dediu & Ladd (2007; DL from now on) are excellent examples of contemporary biolinguistic research and convey what Chomsky (2000: 27) calls the primary goal of bringing "the bodies of doctrine concerning language into closer relation with those emerging from the brain sciences and other perspectives". The articles also shed light on what Chomsky (2005, 2007) calls the "three factors" of language design: (i) genetic factors and UG, (ii) experience and variation within narrow parameters, and (iii) principles not specific to Language such as efficient computation. Additionally, they point out what Boeckx & Grohmann (2007: 2) define as the sense of 'weak' and 'strong' biolinguistics. In particular, LDK's investigation of correlations between populations exhibiting a low frequency of certain allele combinations with populations exhibiting a specific type of language feature — tone systems[1] — concretely puts into practice observational analysis of possible genetic factors related to UG principles and parameters and the notion that variations from UG principles must be narrow in range. Here, the narrow variations (parameters) are part of both the 'physical' and 'abstract' properties of linguistic inquiry and implicate consequences for both the physical brain and the abstract-theoretical structure of grammatical systems. Of course, LDK and DL only present an observed correlation that could be the result of chance — as any correlation between X and Y may be the result of chance with no underlying causation between X and Y. The goal of my response is to highlight some questions that could potentially be useful for issues of deriving inferences from the observed correlations to a degree of causation (see Clark 2000, Shipley 2000, and Thagard 1998 for discussion of causality and correlation). I also ask some questions about

---

[1] To be entirely accurate about LDK and DL's idea about the direction of bias — whether the muted allele pairs bias toward tone systems or non-muted allele pairs bias toward non-tone systems I quote DL (2007: 4): "Finally, note that this bias could be either for or against tone, but the fact that nontonality is associated with the derived haplogroups […] suggests that tone is phylogenetically older and that the bias favors nontonality".

what kinds of evidence and/or counter-examples are needed to potentially support an inference from gene-tone correlation to causation.

2.   The Problem

Despite the pioneering new ground that biolinguistic inquiry is starting to cover, it still faces classic problems related to the issues of correlation and causality, evidence, counterexample, and refutation. The LDK and DL papers are no exception to these problems. Here I ask questions about general problems of adequate evidence/counterexample and causal-inference-from-correlation in gene-language studies. My questions originate from what I perceive to be a possible problem of simultaneity in correlating genetic and linguistic features (see below). A related problem that arises in making causal inferences from observed correlations between genes and languages has to do with the fact that populations of speakers can change languages or language features — consciously or not — quite rapidly when compared to the time it takes for genetic change (see Campbell 2006 for this basic idea applied to gene populations and language families, as well as criticisms of many of the gene-language identi-fication approaches). In other words, a homogenous genetic population can, over time, come to represent a heterogeneous linguistic population by "random chance" of history, culture, and demography. The gene-language relations in these cases are coincidental and no complex causal chain of inference can be established. LDK are vigilant in responding to these 'spurious' relations that are most likely the result of chance and not causality, and thus, their correlational observation seems to not be the result of chance. It is worth quoting LDK (2008: 117) in full.

> The statistical analysis showed that the distribution of the correlations between genetic and linguistic features strongly supports the hypothesized connection between *ASPM-D/MCPH-D* and tone. To rule out the likelihood that this correlation is of the spurious type discussed above, i.e. due entirely to underlying demographic and linguistic processes, Dediu & Ladd computed the correlation between tone and the two derived haplogroups while simultaneously controlling for geographic distances between populations (a proxy for population contact and dispersal) and historical linguistic affiliation between languages (a proxy for similarity through common descent); the proportion explained by these factors turned out to be minimal (again, details are to be found in Dediu & Ladd 2007 and Dediu 2007). It seems, therefore, that the relationship between tone and the derived haplogroups is not due to these standard factors; instead, it could reflect a *causal* relationship between the inter-population genetic and linguistic diversities.

But I have a question. Does correlating a typological feature, such as tone, with a (muted) genetic feature in a population assume that the typological feature has been around approximately as long as the genetic feature it correlates with? That is, if we observe a correlation between tone and a low frequency of alleles in specific populations, and we want to try to infer some complex causal chain wherein the genetic features are part of a complex network of causal factors

for the emergence of tone, then the genetic features and the typological features *should* be simultaneously existent at some point in time. However, the correlated typological feature need not be active — nor does it need to be fully developed. But this leads to a problem for LDK and DL's observations. It is true in general that the absence of evidence (for a feature or property) is not evidence of absence (for that feature or property). But let us assume that a human being does not need the proposed genetic feature in order to acquire or use a natural language tone system — which LDK and DL do. There seems to be no clear way in which to distinguish the natural development of tonogenesis in languages with speakers who *do not have* the muted allele pairs from the development of tonogenesis in languages with speakers who *do have* the muted allele pairs, also assuming tonogenesis proceeds the same in both population groups (which if LDK and DL are right then it should not). Additionally, if one assumes that tone languages arise only in populations with the muted allele pairs, then that is begging the question. Furthermore, LDK state that the particular focus of their discussion "is the recent claim (Dediu & Ladd 2007) that there is a causal relationship between genetic and linguistic diversities at the population level, involving brain growth-related genes and linguistic tone" (p. 114). If one is going to draw a 'causal' implication between gene populations and a typological feature, then is one also arguing that the language feature has stayed *somewhat* actively constant in the target population? LDK allow for the masking of the typological feature by other features or factors, but if the correlation is viewed as somehow 'causal' in any degree then such a typological feature *should* have the capacity to resurface systematically in at least some of the populations that have a predisposition for it. But given the rapidity and frequency with which language populations can, and usually do, alter typological features (due to reanalysis–borrowing–extension and intergenerational parameter shifts — see Harris & Campbell 1995, Lightfoot 1979, 1991, and Roberts 2007 for these two partly conflicting views on language change) it would be rare that any population would consistently retain such features for a substantial period of time, say 6,000 years.[2] The rarity of long-term retention of tone-systems (in at least some of the distinct populations exhibiting the gene-typology correlation), would add credence to the possibility of a neuro-genetic bias. Additionally, long-term retention seems to point out a possible *direction* for dismissing the gene-typology correlation: Show that the typological feature has not stayed constant in the target population it is supposed to be correlated with. Of course, there are mitigating circumstances and one instance of a counter-example would not be enough to dismiss LDK's suggestions. It has also been pointed out by a reviewer that there is no problem in showing that factor X contributes to the prevalence of phenomenon Y, and then observing that in given populations Y has disappeared while X is still existent — in this case other factors W, A, Z suppress the effect of X. As the reviewer points out, this is an essential notion to what a correlation is. But I argue that a systematic instability or non-continuity of the supposed typological feature in *all* the target populations would be adequate evidence for

---

[2]   Dediu & Ladd (2007: 2) give the age of the haplogroup *ASPM-D* as approximately 5.8 thousand years while *MCPH-D* is around 37 thousand years old.

questioning the validity of LDK and DL's correlation leading to a causal explanation.

Imagine a scenario where speakers of language T are transplanted to a population of speakers of language I. In this scenario the children of following T generations will acquire perfect I, becoming I^T. But if the ethnically T descendents who speak I (= I^T) were consequently isolated from the original I population for an adequate amount of time, the LDK view would seem to predict that a possible tonal neuro-genetic bias *could* again trigger some kind of tonogenesis in the newly acquired and now geographically isolated I^T. This kind of 'natural experiment' could surely provide some evidence, but it would not be easy to observe and is probably never likely to be observed. Instead it serves its purpose as an appropriate scenario, or 'thought experiment', and brings to light another question I have. Exactly what kind of counter-example (or what kind of systematic instability) of the possible neuro-genetically biased typological feature would count as dismissive?[3] If a (muted) genetic feature in specific populations can be correlated with a typological feature of those same populations, then should we not expect that typological feature to be prevalent in the languages of those populations? And if not always prevalent or stable across a sufficient period of time — a very likely probability as the 'thought experiment' informally shows — then what kind of instability of the predicted typological feature would count as a genuine counter-example to a possible LDK-type hypothesis? In other words, if there existed a population of speakers with the muted allele pairs that had never acquired or used a tone system, then how would one explain this? If the explanation was the systematic suppression of the typological feature of tone by other factors W, Z, A, then what evidence would count as showing the systematic non-expression of a feature that is correlated with a genetic predisposition for it? To put it another way, how does one measure the suppression of a typological feature?

One might argue that (i) a causal inference from the correlation of a genetic feature with a typological feature does not imply that the typological feature should be approximately as old as the genetic feature; whether active or not. In this case, an external or internal stimulus could 'trigger' the rapid development through the mechanism of intergenerational transmission of the typological feature throughout the target genetic population. One might also argue that (ii), assuming a causal inference from the correlation, the development of the typological feature took a very long time to reach its present state, and thus, there is no need to say that the typological feature had been around for 6,000 years. Instead, it is a more recent innovation with a long historical development now facilitated by the mechanism of intergenerational transmission. But (i) and (ii) are both complications that need to be verified empirically. In the first case of the 'trigger' (i), what would serve as an adequate stimulus? In the second case of the development scenario (ii), it seems one would be hard pressed to show why this development is not different than any other kind of language structure

---

3   I admit that this is simplistic, as counter-examples need not dismiss, destroy, or falsify a hypothesis, or a theory built from hypotheses, based on empirical observations — assuming a theory can develop from the gene-tone correlation. I merely intend here to ask what would constitute a genuine counter-example.

development that occurs over time: What makes it so unique that it can be causally linked with a correlated neuro-genetic feature? I argue that if LDK's correlational observation is going to yield a causal explanation then it cannot escape the implication that the 'life' of the typological feature (whether active or not) should be roughly simultaneous with the appearance or 'activation' of the genetic feature. Providing evidence for this simultaneity is another issue.

Lastly, if a typological feature could be proved to be stable for a specific population of speakers, and this population appeared to have some unique genetic feature that could be shown to correlate with the language feature in question, then there will be a discrepancy between the time-depth of reliable information between language and gene datum — as historical-descriptive linguistics generally has only a reliable 6,000 year time-depth, while genetic information can exceed this limit by a substantial amount.[4] It is not clear if this poses any real problems to causal inferences for genetic and typological features, but it is surely a factor in considering the kinds of evidence used for establishing genetic and typological relations.

In (1) I repeat the questions asked above; though I accept the risk that they may not be coherent out of the context in which they were asked and there may be some redundancy. Following (1) are a few more questions, in (2), that might be relevant to both the general method of gene-language correlation and the specific observations of LDK and DL. I could not possibly begin to sketch answers to these questions in a short response, but will try to give very short answers to those in (2).

(1)   a.   Does correlating a typological feature, such as tone, with a (muted) genetic feature in a population assume that the typological feature has been around approximately as long as the genetic feature it correlates with?

      b.   If one is going to draw a "causal" implication between gene populations and a typological feature, then is one also arguing that the language feature has stayed *somewhat* actively constant in the target population?

      c.   Exactly what kind of counter-example, or systematic instability, of the possible neuro-genetically biased typological feature would count as dismissive?

      d.   If a (muted) genetic feature in specific populations can be correlated with a typological feature of those same populations, then should we not expect that typological feature to be prevalent in the languages of those populations?

      e.   What kind of instability of the predicted typological feature would count as a genuine counter-example to a possible LDK-type hypothesis?

---

[4]   Where *ASPM-D* is about 5.8 thousand years old and *MCPH-D* is 37 thousand years old. Perhaps a moot point here because the correlation crucially involves the *pair* of haplogroups and so any typological feature correlated with the pair can only be as old as the earliest instance of the pair.

f.   If there existed a population of speakers with the muted allele pairs that had never acquired or used a tone system, then how would one explain this?

g.   What evidence would count as showing the systematic non-expression of feature that is correlated with a genetic predisposition for it?

h.   How does one measure the suppression of a typological feature?

(2) a.   Are tone systems really prone to regular historical change, or are they somehow more resilient to change (in the sense that if tone systems are very stable in themselves, then long-term retention of them may not be due to a possible genetic bias but to the abstract nature of the typological feature itself)?

b.   Can we confidently show that populations suggested to exhibit genetic factors that increase the likelihood of having tone systems based on certain muted allele pairs have historically stable tone systems — and what are the linguistic factors contributing to loss or gain of tone systems in these populations?

c.   What kind of unstable, or discontinuous, appearance of the typological feature in the target population counts as a genuine counter-example — or does the criterion of "appearance of the feature" even qualify as relevant to establishing the parameters for counter-examples to LDK's research?

d.   What kinds of assumptions about simultaneity of genetic properties/features and typological properties/features are operative when discussing issues of the neuro-genetic bases of natural human languages?

I think the answer in (2a) is fairly straightforward: Tone systems *are* prone to regular change and do not show any more stability than other structures (Gussenhoven 2004, Yip 2002). But with this answer comes more questions about certain facts of tone. For example, if the target population has a predisposition or bias to acquiring and using tone systems, then do they also have a bias for what are commonly recognized as the phonemic/phonetic precursors to tone (Fromkin 1978, Hombert, Ohala & Ewan 1979, Matisoff 1973)? (Of course, see footnote 2.) The answer to (2b) would take some time, but I believe that it is a productive direction towards compiling linguistic data sets relevant to LDK's research. Of course, it has its strict limits — namely that even with written records going back 6,000 years the evidence of a tone system in a language that old is not easy (or impossible) to substantiate. As for (2c), also (1c) and (1e), I have no adequate answer, but it seems to be an important and relevant question to specific issues in LDK and DL *if* one assumes that the goal is to derive causal inferences from the observed correlations *and* the problem of simultaneity is a real problem. As for (2d), it is a general question relevant to the methodological aims and practices of biolinguistic research specifically aimed at deriving causal inferences from

correlational observations about genes and typology; it can only be answered through the process of research, investigation, and critical inquiry and can, I believe, potentially have what Chomsky (1995: 232) attributes to the Minimalist Program — "a certain therapeutic value".

3.   Conclusion

Unless a causal link between gene-language or genetic feature and typological feature can be established, then an observed correlation does not seem to be very useful. LDK and DL are clearly committed to a research strategy that seeks to discover a causal link; although it is overwhelmingly clear that this link *should not* be direct or deterministic and is likely not to be. Any degree of causality here, I think, is generally expected to be of a complex, multifactorial nature. In fact, Paul Thagard's Causal Network Instantiation (CNI) model (1998) for making causal inferences from observed correlations in medical scientific explanations for diseases seems like a good fit with the LDK and DL research. As Thagard (1998: 76) himself says,

> I expect, however, that there are many fields such as evolutionary biology, ecology, genetics, psychology, and sociology in which explanatory practice fits the CNI model. For example, the possession of a feature or behavior by members of a particular species can be explained   in   terms   of   a   causal network involving mechanisms of genetics and natural selection. Similarly, the possession of a trait or behavior by a human can be understood in terms of a causal network of hereditary, environmental, and psychological factors. In psychology as in medicine, explanation is complex and multifactorial in ways well characterized as causal network instantiation.

In pushing any research to reveal potentially useful inferences from correlation to causation one almost heuristically demands that there is a causal link, once chance has been somewhat ruled out (and while trying to rule out other causes), and then one works to establish the most likely complex causal path. This should be true also in the search for causal paths, webs, or networks from genes to languages — or populations with specific muted allele pairs to populations who are predisposed to acquire, use, or generate tone languages. Shipley (2000) argues that in most cases correlation implies an *unresolved* causal structure — unresolved in that we have not yet discovered cause, effect, and/or other variables. Shipley (2000: 3) says that "[i]n fact, with few exceptions, correlation does imply causation. If we observe a systematic relationship between two variables, and we have ruled out the likelihood that this is simply due to random coincidence, then *something* must be causing this relationship." Precariously, the assumptions needed for discovering inferences from correlation to causation may turn out to be as complex as the phenomenon under investigation. As Chomsky (1995: 233) notes, "[i]t is all too easy to succumb to the temptation to offer a purported explanation for some phenomenon on the basis of assumptions that are roughly the order of complexity of what is to be explained". LDK and DL seem to me to be cautious about not succumbing to the 'temptation'. And even though the assumptions needed to discover inferences

from correlation to causation may be complex, and the criteria of evidence for measuring the neuro-genetic bias or predisposition that a person and population may have for exhibiting some linguistic trait may not seem clear (whether that trait is ever expressed or not), it is well to remember what Boeckx (2006: 91) points out about rigor and maturation in research programs: "Programs take time to mature, and rigor cannot be required in the beginning". The expectation of solid evidence of some causal link between muted allele pairs and tone systems is premature and stifles the hard-won creativity in research that the Minimalist Program, and by extension Biolinguistics, has achieved. New areas of scientific research are messy, and this messiness should not cloud our vision of what kind of order may reveal itself over time. But this does not mean we should not ask a variety of questions and expect some answers — or at least a direction towards answers. Whether an inference from correlation to causation in LDK and DL will ultimately be found, or the questions asked here are useful or relevant, the lesson is that there is at least a "therapeutic" value to biolinguistic research through eliminating questions and trying to establish causal inferences.

References

Boeckx, Cedric. 2006. *Linguistic Minimalism: Origins, Concepts, Methods, and Aims*. Oxford: Oxford University Press.

Boeckx, Cedric & Kleanthes K. Grohmann. 2007. The *Biolinguistics* manifesto. *Biolinguistics* 1, 1-8.

Campbell, Lyle. 2006. Languages and genes in collaboration: Some practical matters. Paper presented at *Language and Genes: An Interdisciplinary Conference*, Santa Barbara. [University of California, 8-10 September 2006.]

Clark, Andy. 2000. Twisted tales: Causal complexity and cognitive scientific explanation. In Frank C. Keil & Robert A. Wilson (eds.), *Explanation and Cognition*, 145 -166. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28), 167-217. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Linguistics and brain science. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, Language, and Brain*, 13-28. Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1-22.

Chomsky, Noam. 2007. On minds and language. *Biolinguistics* 1, 9-27.

Dediu, Dan & D. Robert Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. *Proceedings of the National Academy of Sciences of the United States of America* 104, 10944-10949.

Fromkin, Victoria (ed.). 1978. *Tone: A Linguistic Survey*. New York: Academic Press.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation* (Research Surveys in Linguistics). Cambridge: Cambridge University Press.

Harris, Alice & Lyle Campbell. 1995. *Historical Syntax in a Cross-Linguistic*

Perspective (Cambridge Studies in Linguistics 74). Cambridge: Cambridge University Press.

Hombert, Jean-Marie, John Ohala & William Ewan. 1979. Phonetic explanations for the development of tone. *Language* 55, 37-58.

Ladd, D. Robert, Dan Dediu & Anna R. Kinsella. 2008. Languages and genes: Reflections on biolinguistics and the nature-nurture question. *Biolinguistics* 2, 114-126.

Lightfoot, David. 1979. *Principles of Diachronic Syntax* (Cambridge Studies in Linguistics 23). Cambridge: Cambridge University Press.

Lightfoot, David. 1991. *How to Set Parameters: Arguments from Language Change*. Cambridge, MA: MIT Press.

Matisoff, James. 1973. Tonogenesis in Southeast Asia. *Southern California Occasional Papers in Linguistics* 1, 71-95.

Roberts, Ian. 2007. *Diachronic Syntax* (Oxford Textbooks in Linguistics). Oxford: Oxford University Press.

Thagard, Paul. 1998. Explaining disease: Correlations, causes, and mechanisms. *Minds and Machines* 8, 61-78.

Shipley, Bill. 2000. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations, and Causal Inference*. Cambridge: Cambridge University Press.

Yip, Moira. 2002. *Tone* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Joshua Bowles
University of Utah
Department of Linguistics
255 South Central Campus Drive, Room 2300
Salt Lake City, UT 84112
bowles.j.ling@utah.edu

★ FORUM ★

Reply to Bowles (2008)

D. Robert Ladd, Dan Dediu & Anna R. Kinsella

In response to Bowles (2008), we wish above all to reiterate one of the main points of our original discussion (henceforth LDK 2008), which is the following: If there is to be a field of biolinguistics that makes a useful contribution to our understanding of the human language faculty, then it is important to adopt Boeckx & Grohmann's (2007) 'strong' sense of biolinguistics. We are naturally pleased that our work has elicited detailed comment from an adherent of 'weak biolinguistics', but we feel that, in his eagerness to equate biolinguistics with Minimalist research in formal linguistics and to evaluate our work according to the standards of that research paradigm, Bowles has missed our point about the need for genuinely interdisciplinary investigation. We are well aware of the logical problems associated with conclusions based on correlations. However, as we tried to make clear in LDK, the consequence we draw is that we need to look for evidence in other sources of data, not (as Bowles does) merely think harder about the logic of our claims.

When Bowles says (p. 247) that biolinguistic research "still faces classic problems related to the issues of correlation and causality, evidence, counterexample, and refutation", he ignores our general suggestion (LDK 2008: 122) that we need to bring together "linguists and others *in equal measure*, making use of their respective methodologies with a full understanding of their assumptions, and trying to resolve any incompatibilities using shared standards of falsifiability and argumentation". When he rightly points out (p. 248) that there is "no clear way in which to distinguish the natural development of tonogenesis in languages with speakers who do not have the muted [sic] allele pairs from the development of tonogenesis in languages with speakers who do have the muted allele pairs", he ignores the fact that nothing in our work suggests that there should be. When he wonders (p. 249, fn. 3, and again p. 250, questions (1c) and (1e)) "what would constitute a genuine counter-example" to our claim, he is thinking in terms of the kinds of theoretical enquiry in which counter-examples play an important role in shaping conclusions; he ignores Dediu & Ladd's (2007: 10947) explicit suggestion that their correlational finding "warrants future experimental work, which will help test and refine the hypothesis of a causal effect".

A more concrete problem with Bowles's discussion is that, despite his disclaimers, many of his points seem to be based on the assumption that there are specific genes that code for specific linguistic features in the individual. Among

the issues he considers at some length is whether positing a correlation between a typological feature and a genetic feature "assume[s] that the typological feature has been around approximately as long as the genetic feature it correlates with" (p. 250, question (1a)). This question reveals a profound misunderstanding of how genetics works. It seems pretty clear, for example, that the *FOXP2* gene makes some essential contribution to human linguistic abilities, but *FOXP2* has been "around" for millions of years and is found in many other species without allowing any of those other species to talk. The phenotypic effects of a gene are highly dependent on context, where context includes the rest of the genome, the physical environment, and (in the case of humans) culture. This is part of the reason that it is still far from clear exactly what *FOXP2* does to facilitate language in humans.

Note in this connection that Dediu & Ladd and LDK suggest a number of general cognitive and perceptual differences that might be relevant to a bias for or against linguistic tone, including phonological working memory, low-level pitch tracking, and the ability to process rapid sequences of sounds. These do not appear to be the kinds of traits Bowles has in mind when he talks about "possible genetic factors related to UG principles and parameters" (p. 246) or about linguistic features being "expressed" (p. 252) or "somewhat actively constant" (p. 250, question (1b)) in a population. They are, however, the kinds of differences that can be investigated experimentally and related to observable differences in brain anatomy and physiology, and are biologically far more plausible candidates for the substance of the hypothesized bias than a specific instruction to the language acquirer to assume that the language they are exposed to is tonal. We also note that Bowles seems not to appreciate the importance of the fact that the correlation under discussion is between genetic variation and linguistic variation *in populations*. In both the original Dediu & Ladd paper and in LDK, we went out of our way to emphasize that the contribution of intergenerational transmission of language is essential to any proposed link between population genetics and linguistic typology. No specific *linguistic* predictions about individuals are implied by our work.

Nevertheless, Bowles does raise one important issue that is primarily linguistic, concerning the historical stability of typological features and specifically the historical stability of tone (p. 251, question (2a)). If the distribution of tone (or any other typological feature) is affected by a genetically-mediated bias, it is reasonable to expect that it may be more stable over time. That is, once tone is present in a genetically predisposed group, it should be less likely to disappear through the ordinary mechanisms of language change; by the same token, a language that lacks tone should be less likely to acquire it through those mechanisms it if it is spoken by a group genetically disposed against it. Bowles argues that these expectations are not met: Tonogenesis and tone loss, he says, are as common as any other historical change, and the idea of a genetically-mediated bias is therefore problematical. But the idea that tone comes and goes like any other typological feature is actually open to discussion, *pace* Bowles and the authorities he cites. For one thing, the languages of sub-Saharan Africa, across three major language phyla, are overwhelmingly tonal, and for most of them there is no evidence that they have ever been anything else. Loss of tone in

Swahili, for example, is relatively recent and almost certainly related to contact and use as a lingua franca. More generally, it may be important to distinguish between the structural pressures that bring about tonogenesis and the long-term historical developments that follow. In East and Southeast Asia, it is generally accepted that many previously non-tonal languages rapidly became tonal two or three thousand years ago (e.g., Haudricourt 1954), and tone is now central to the phonology of most of these languages. In Northern Europe, by contrast, it is similarly uncontroversial that some sort of tonogenesis took place about 800 years ago, yet tone remains marginal. Norwegian probably has the best claim of any European language to be called tonal, but it is an obvious typological oversimplification to put Norwegian in a class with Chinese, and there are researchers (e.g., Morén 2005) who argue that the Scandinavian languages do not actually have lexically-specified tone at all.

These considerations suggest a refinement of what Bowles says about the historical stability of tone: Tonogenesis itself may indeed be a rather ordinary historical process of phonologization or secondary split, but the thoroughgoing incorporation of tone into a language depends heavily on other factors — almost certainly including areal language contact, and possibly including genetically-mediated biases. The idea of drawing such a distinction — between structural triggers for phonologization of phonetic differences and the long-term establishment of new phonemic contrasts — is discussed by Kiparsky (1995: 655ff.), who specifically (citing Svantesson 1989) mentions tone as a likely case in point. If some such distinction is valid, then Dediu & Ladd's hypothesis suggests a historical account along the following lines. Tonogenesis 'happened' in Southeast Asia and in Northern Europe, in both cases through well-established mechanisms of diachronic change. In Southeast Asia, the population genetic environment was favourable, and tone took hold and spread to become a thoroughly ingrained feature of the phonology of the languages involved. In Northern Europe, the population genetic environment was unfavourable, and tone remained marginal and continues to struggle to this day.

It is thus possible that typological change involving tone is different from typological change in, say, word order. This is a matter that can best be studied on the basis of descriptive and historical linguistic work, and typological theorising about the nature of tone. But such research is not biolinguistics: A finding that tone is exceptionally stable in Africa, or that tonogenesis happens regularly everywhere but only catches on in certain areas, might be *consistent* with the Dediu-Ladd hypothesis, but on its own would do nothing to prove it. If we are serious about learning more about the biological foundations of language, we have to integrate what we know about language with what we know about biology. Research into the formal properties of language is useful and important, but describing it as "biolinguistics" is just wishful thinking.

References

Boeckx, Cedric & Kleanthes K. Grohmann. 2007. The *Biolinguistics* manifesto. *Biolinguistics* 1, 1-8.

Bowles, Joshua. 2008. Some questions about determining causal inference and criteria for evidence: Response to Ladd, Dediu & Kinsella (2008). *Biolinguistics* 2.2–3, 246-254.

Dediu, Dan & D. Robert Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and *Microcephalin*. *Proceedings of the National Academy of Sciences of the United States of America* 104, 10944-10949.

Haudricourt, André. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242, 69-82.

Kiparsky, Paul. 1995. The phonological basis of sound change. In John Goldsmith (ed.), *The Handbook of Phonological Theory* (Blackwell Handbooks in Linguistics), 640-670. Oxford: Blackwell.

Ladd, D. Robert, Dan Dediu & Anna R. Kinsella. 2008. Languages and genes: Reflections on biolinguistics and the nature-nurture question. *Biolinguistics* 2, 114-126.

Morén, Bruce. 2005. Danish *stød* and Eastern Norwegian pitch accent: The myth of lexical tones. Paper presented at the *13th Manchester Phonology Meeting*, Manchester. [Hulme Hall, 26-28 May 2005.]

Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in northern Mon-Khmer. *Phonetica* 46, 60-79.

D. Robert Ladd & Anna R. Kinsella
The University of Edinburgh
School of Philosophy, Psychology and Language Sciences
Dugald Stewart Building
3 Charles Street
Edinburgh EH8 9AD
United Kingdom
bob.ladd@ed.ac.uk & anna.kinsella@ed.ac.uk

Dan Dediu
Max Planck Institute for Psycholinguistics
Wundtlaan 1
6525 XD Nijmegen
The Netherlands
dan.dediu@mpi.nl

Biolinguistics

Volume 2, Issue 2–3

Spring–Summer 2008

TABLE OF CONTENTS