In Defense of Vacuous Projections in Bare Phrase Structure

MAXIMILIANO GUIMARÃES

Abstract

According to the current minimalist view, Bare Phrase Structure Theory (Chomsky 1994, 1995) implies the impossibility of vacuous (non-branching) projections. The problem with that is that the existence of structures with two heads in symmetric c-command relation at the very bottom of the phrase marker is inevitable, making impossible the linearization of the terminals according to the LCA (Kayne 1994). Assuming the Bare Phrase Structure Theory in all its aspects, I argue in this paper that vacuous projections are totally compatible with the minimalist *desiderata*, being available for the system as a legitimate (last resort) strategy to avoid symmetric c-command. Consequently, the linearization problem is just illusory.

Introduction

This paper is an exercise in grammatical formalism. My goal here is to show that one of the central claims of Chomsky's (1994, 1995) Bare Phrase Structure Theory – namely, that vacuous projections do not exist – is not as trivial as it seems to be at first blush. Eventually, I argue against this claim, showing that vacuous projections are a natural way of reconciling the Bare Phrase Structure Theory with Kayne's (1994) Linear Correspondence Axiom (henceforth, LCA). Perhaps, I am taking the technical apparatus of the theory too seriously and consequently my arguments might be nullified as some details are worked out. But even if that is the case, I consider this work relevant anyway, to the extent that it shows that there are some points in the theory which have been taken for granted even though there is a lot of "homework" to do before taking them as natural and obvious consequences of the *desiderata* of Bare Phrase Structure Theory. The paper is organized as follows. In section 1, I summarize the goals and the core assumptions of Bare Phrase Structure Theory. In section 2, I briefly discuss how the LCA (Kayne 1994) has been adapted in order to fit into Bare Phrase Structure Theory. In section 3, I recall a problematic incompatibility between both theories, and argue that current attempts to solve the problem are empirically and conceptually inadequate. In section 4, I argue that when we look at the problem more closely and carefully, we realize that it is just illusory, and that there is nothing in the Bare Phrase Structure Theory itself that rules out vacuous projections. I further show how vacuous projections are generated by the very same mechanism that generates branching projections. In section 5, I take a closer look at vacuous projections from the point of view of economy of derivations and representations. Finally, section 6 concludes the paper.

1 *Inutilia Truncat*

One of the main advances of the Minimalist Program with respect to previous versions of the Theory of Principles & Parameters is the abandonment of X-Bar Theory (Chomsky

University of Maryland Working Papers in Linguistics, vol. 9, 90-115.

1970, 1981, 1986a, 1986b; Jackendoff 1977; Stowell 1981; *inter alia*) in favor of Bare Phrase Structure Theory (Chomsky 1994, 1995). The distinction between head and terminal node is eliminated, and the concepts of maximal, minimal & intermediate projections, on the one hand, and those of head, complement, specifier & adjunct, on the other hand, are defined over set-theoretic objects in derivational and relational terms (*i.e.* with respect to each derivational step, and with respect to other syntactic objects of the phrase marker), instead of taken as primitive notions based on a primitive structural skeleton. This makes Bare Phrase Structure Theory significantly less arbitrary and more natural than X-Bar Theory, although the former is still a "cleaned-up" version of the latter.

In (1), I present the definition of syntactic object (*i.e.* phrase) according to this framework. The operation that recursively generates syntactic objects is Merge, defined as in (2). The two fundamental syntactic relations between syntactic objects are Dominance and C-Command, defined as in (3) and (4), respectively. All four definitions below are mine, and, apart from small technical details, they pretty much represent the standard view in the literature.[1]

(1) SYNTACTIC OBJECT: Σ is a syntactic object if and only if either (i) Σ is a lexical item,[2] or (ii) Σ is the set $K = \{\gamma, \{\alpha, \beta\}\}$, such that both α and β are syntactic objects, and γ is the label of K.[3]

(2) MERGE:[4] input: α & β, such that both α and β are syntactic objects;
output: $K = \{\gamma, \{\alpha, \beta\}\}$, such that γ (the label of K) corresponds to either the head of α or the head of β.

(3) DOMINANCE: Given a syntactic object $K = \{\gamma, \{\alpha, \beta\}\}$, K dominates a syntactic object δ if and only if either (i) $\exists L \mid \delta \in L \,\&\, L \in K$, or (ii) $\exists M \mid K$ dominates M & M dominates δ.

---

[1] See Nunes & Thompson (1998) for a detailed discussion on the matter.

[2] If the hypothesis of feature-movement (Chomsky 1995: 261-271) is correct, then the condition (i) in (1) needs to be redefined as "Σ is a lexical item or the set of formal features of a lexical item". One possible way of stating this in technical terms is assuming that lexical items have an internal structure like the one specified in (1-ii), as suggested by Nunes (2000b, note 24).

[3] I'm abstracting away from adjunction, as well as from issues related to the precise nature of labels (*i.e.* whether they are entities (with a substantive content) or just relations between two objects that undergo Merge (hence, a formal property of syntactic objects); whether they are types or tokens; whether they have internal substructure; or even whether they really exist to begin with). For different takes on these issues, see Chomsky (1995: 244-249; 1998), Moro (1997), Baker & Stewart (1999), Collins (1997, 1999), and Hornstein & Uriagereka (1999), *inter alia*. From now on, I indicate the labels of syntactic objects by underlining them. This is a mere expository device, not intended to express any theoretical position.

[4] This definition of Merge does not contemplate adjunction. I refer the reader to Nunes (1995b: 67-70), Gärtner (1997, sections B.6.2 & B.6.3) and Nunes & Thompson (1998: 499-503) for some insightful conceptions of adjunction under Bare Phrase Structure Theory.

(4) C-COMMAND:[5] Given two maximal and/or minimal projections α & β,[6]
α c-commands β if and only if (i) α ≠ β, (ii) α does not dominate β[7] & (iii) every category that dominates α also dominates β.

These four primitives are sufficient to define all relevant syntactic relations.[8] A minimal projection is a syntactic object that does not dominate any other syntactic object. A maximal projection is a syntactic object whose label is distinct from the one of the syntactic object that immediately dominates it. An intermediate projection is a syntactic object that is neither a minimal projection nor a maximal projection. The head of a minimal projection is the minimal projection itself (what else could it be?). The head of a non-minimal projection is the dominated minimal projection that determines its label. A phrase α is the complement of β if and only if α is the sister of β, and is β is a minimal non-maximal projection. A phrase α is the specifier of β if and only if α is the sister of an

[5] A strict derivational conception of c-command from the point of view of Bare Phrase Structure Theory is given by Epstein (1999), without making reference to dominance. From this point of view, c-command is simply the direct reflex of successive applications of Merge along the history of the derivation, defined as follows: *"X c-commands all and only the terms of the category Y with which X was paired by Merge (or by Move) in the course of the derivation"*. Notice, however, that it is inevitable to make reference to the notion of "*term of*", which is nothing but "reflexive dominance". See Drury (1998b) for arguments against Epstein's (1999) view. A more sophisticated version of Epstein's (1999) idea is given by Epstein, Groat, Kawashima & Kitahara (1998: chapter 6), which, in my opinion, ends up being equivalent to a representational approach due to its global nature and its appeal to "derivational memory".

[6] The assumption that c-command is a relation that does not apply to intermediate projections is a necessary stipulation in Bare Phrase Structure Theory; otherwise it is impossible to get the asymmetry that is necessary for linearizing the terminals according to the LCA (Chomsky 1995: 338-340). Chomsky states this assumption by stipulating that intermediate projections are "invisible for the computational system". This stipulation is paradoxical, however, to the extent that intermediate projections are relevant to calculating the c-command relations for other syntactic objects in the phrase marker, which means that intermediate projections cannot be "invisible", strictly speaking. Therefore, we have to built this stipulation into the definition of c-command itself, stating that c-command does not hold for intermediate projections (*i.e.* either intermediate projections do not participate in c-command relations at all, or, at least, they cannot be commanders). See Drury (1998b) for some discussion on this matter.

[7] In order to account for adjunction, the condition (ii) in (4) must be redefined as *"no segment of the category α dominates β"*; otherwise adjuncts and the phrases they are adjoined to cannot be linearized with respect to each other, since there is no asymmetric c-command relation between them (see (6) and (7) below). Attempts to derive/eliminate the segment/category distinction are made by Nunes & Thompson (1998) and Frank & Vijay-Shanker (1999).

[8] Actually, this is an oversimplification. In order for the definitions I give here to work, first we need to formulate a formal and precise definition of a fifth syntactic relation: sisterhood, which I am taking for granted here. Abstracting away from adjunction, a simple way of defining it is taking α and β to be sisters if and only if both α and β are dominated by the same phrases. From this perspective, the notion of sisterhood is just an expository device rather than a real fifth primitive of the system.

intermediate projection **γ**, and β is the head of **γ**.[9] By these definitions, a syntactic object can be simultaneously a minimal and a maximal projection (an idea that goes back to Muysken (1982) and Speas (1990)). Therefore, minimal projections can behave as complements or specifiers, as long as they are also maximal projections. This point constitutes one of the main differences between Bare Phrase Structure Theory and X-Bar Theory. In this paper I am concerned specifically with the consequences that minimal projections in complement positions bring to Bare Phrase Structure Theory, and, more generally, to the minimalist conception of grammar.

2 The Relation Between Phrase Structure and Word Order

Along the pre-minimalist generative tradition, the notion of order was considered to be an inherent property of syntactic structures. The standard definition of phrase marker itself incorporates the precedence relation as one of the primitive syntactic notions[10] (cf. Wall 1972: 149; Partee, ter Meulen & Wall 1993: 441-442). Under this view, conditions on derivations and representations should be sensitive to properties like "precedence", "subsequence", "adjacency" or "edge/periphery". The fact that human languages differ with respect to the order of the constituents for a syntactic structure was treated in terms of movements and parameters of order.

A new and revolutionary approach to the matter was offered by Kayne (1994). He observed that, if word order variations were really due to parameters of order (combined with movements), we would expect to find many cross-language symmetries in pairs of structures that scope and binding tests show to be in fact asymmetric. Therefore, the apparent mirror-image effect is just illusory (*i.e.* a mere misleading PF fact). In face of this, the author concludes that there are no directionality parameters, and that the linear order of terminal constituents directly follows from the syntactic hierarchy. This correspondence between hierarchy and order is taken to be determined by an invariable principle of Universal Grammar: the *Linear Correspondence Axiom* (LCA).

> Consider the set *A* of ordered pairs $<X_j,Y_j>$ such that for each j, X_j asymmetrically c-commands Y_j. Let us further take *A* to be the maximal such set; that is, *A* contains all pairs of nonterminals such that the first asymmetrically c-commands the second. Then the central proposal I would like to make is the following (for a given phrase marker P, with *T* the set of terminals and *A* as just given). *Linear Correspondence Axiom:* d(*A*) is a linear ordering of *T*.[11] (Kayne 1994: 5-6)

---

[9] Notice that, according to this definition, there can be multiple specifiers per head, as opposed to X-Bar Theory, which allows only one specifier per each head, according to an axiomatic skeleton.

[10] See Chametzky (1996: chapter 1) for arguments against the standard idea that precedence is a formal primitive of syntax. See also Drury (1998a, 1998b, 1999) and Guimarães (1999a) for an approach in which precedence is THE grammatical principle, constituting the basis for defining most grammatical notions.

[11] *"I will refer to nonterminal-to-terminal dominance relation as d. This relation d is a many-to-many mapping from nonterminals to terminals. For a given nonterminal X, let us call d(X) the set of terminals that X dominates. d(X) can be said to be the 'image' under d of X"* (Kayne 1994: 5).

According to the LCA, all members of *T* (*i.e.* all terminal elements of a given syntactic structure) must be linearly ordered with respect to each other. That is, there must be some kind of relation *R* holding of all terminals, such that *R* is a linear order on *T*, *i.e. R* must have the three properties in (5) (cf. Partee, ter Meulen & Wall 1993: 206-211).

(5) FORMAL PROPERTIES OF A LINEAR ORDER:[12]

a. ***transitivity*** $=_{def} \forall\alpha, \forall\beta, \forall\gamma\ [[[<\alpha,\beta> \in R] \& [<\beta,\gamma> \in R]] \rightarrow [<\alpha,\gamma> \in R]]$

b. ***asymmetry*** $=_{def} \forall\alpha, \forall\beta\ [[<\alpha,\beta> \in R] \rightarrow [<\beta,\alpha> \notin R]]$

c. ***totality/connectedness***: $=_{def} \forall\alpha, \forall\beta\ [[<\alpha,\beta> \in R] \vee [<\beta,\alpha> \in R]]$

In addition to this formal claim, Kayne (1994) makes the substantive claim that the relation *R* just mentioned is *Precedence*.[13] Therefore, the LCA can be stated as the mapping function in (6) below.

(6) LINEAR CORRESPONDENCE AXIOM (Kayne 1994: 33)

Given two nonterminal nodes A & B, and two terminal nodes α & β, such that A dominates α, and B dominates β, if A asymmetrically c-commands B, then α precedes β.

By (6), a phrase marker is well-formed if and only if none of the c-command and dominance relations it presents incur in a violation of the required linear order of terminals. That is, all requirements in (5), which are inherent to the *Precedence* relation, must hold among the terminals of every phrase marker. This theory has two radical immediate consequences:[14] (i) different order patterns necessarily correspond to different hierarchical structure patterns, and (ii) There is no rightward movement (under the standard assumption that movement is always to a c-commanding position).[15]

---

[12] The fourth formal property inherent to a linear order is *irreflexivity* (called *aliorelativity* by Russel 1919: chapter IV), defined as follows: ***irreflexivity*** $=_{def} \forall\alpha, \forall\beta\ [[<\alpha,\beta> \in R] \rightarrow [\alpha \neq \beta]]$ (*i.e.* $\forall\alpha, [<\alpha,\alpha> \notin R]$). Actually, it is not necessary to make this fourth property explicit, since every asymmetric relation is, by definition, an irreflexive one. Alternatively, we can do the other way around, making *irreflexivity* explicit and omitting (pressuposing) *asymmetry*, since, among transitive relations, all irreflexive ones are also asymmetric, and *vice versa* (see Russel 1919: chapter IV, *inter alia*).

[13] This is not the only possibility. In principle the relation *R* could be *Subsequence*. See Kayne (1994: 36-38) for an attempt to derive why *R* must be *Precedence* rather than *Subsequence*. For an insightful (but inconclusive) discussion on the matter, see Uriagereka (1999).

[14] In Kayne's system, there is a third radical consequence, namely, that every specifier is in fact an adjunct (Kayne 1994: 15-17). In the standard version of Bare Phrase Structure Theory, given the invisibility of intermediate projections (cf. note 6), Chomsky (1995: 248-249) assumes that the traditional distinction between specifier and adjunct is real.

[15] Movement is always leftwards, unless we assume lowering operations, which are ruled out in standards versions of of Bare Phrase Structure Theory due to the Extension Requirement (Chomsky 1995: 190, 327-328; Kitahara 1995; Watanabe 1995). See also Uriagereka (1998: chapter 4), Bobaljik & Brown (1997), Nunes (1995b, 2000b, 2000c), and Guimarães (1998: 75-

Through this formalism, Kayne (1994) derives the essential properties of X-Bar Theory, such as endocentricity and binary branching. The author endorses the traditional position that linear order is an inherent property of syntactic structures. However, he radically departures from the standard view in two important aspects. First of all, he assumes that linear order is not established for all constituents of a sentence, but only for terminal elements. Moreover, he claims that linear order is not just a mere property of phrase markers among many others (Partee, ter Meulen & Wall 1993: 441-442). Rather, it is the syntactic notion per excellence: the axiom from which other properties of X-Bar Theory are theorematically derived. Thus, the LCA is a well-formedness principle of syntax (actually, THE syntactic well-formedness principle) to be satisfied in all derivational steps and at all levels of representation (Kayne 1994: 48-49).

Kayne's proposal has been incorporated, with some adjustments, to Bare Phrase Structure Theory. Chomsky (1995: 334) claims that syntactic computation is insensitive to linear order of constituents, because there are just hierarchical relations among constituents in the syntax, but no linear order relations, which only exist after the syntax-phonology mapping, between Spell-Out and PF, when precedence relations among terminals are established on the basis of asymmetric c-command (and dominance).

> There is no clear evidence that order plays a role at LF or in the computation from N[umeration] to LF. Let us assume that it does not. Then ordering is part of the phonological component, a proposal that has been put forth over the years in various forms. If so, then it might take quite a different form without affecting C_{HL} if language use involved greater expressive dimensionality or no sensorimotor manifestation at all. (Chomsky 1995: 334)

This mapping algorithm that linearizes the terminals corresponds, *mutatis mutandis*, to the LCA proposed by Kayne (1994). Chomsky (1995) considers that the organization of words (*i.e.* terminals) in the temporal axis is a necessary condition for the construction of a PF representation. Since the speech is instantiated in real time, and the nature of the Articulatory-Perceptual (A-P) system imposes the restriction that it is impossible to pronounce two or more words simultaneously, it is necessary to establish a temporal linear order on the set of words of the sentence in order to obtain an *output* that is legible by the A-P system (*i.e.* pronounceable). From this perspective, the linear order of words is a *bare output condition*: *i.e.* a requirement imposed upon the level of representation PF by the A-P performance system, which is supposed to read it, interpret it, use it (an idea that goes back to Higginbotham (1983)).[16] The null hypothesis is that only terminals with phonological features are visible for the LCA, because, by the Full Interpretation Principle (or Economy of Representation (cf. Chomsky 1995: 150-151)),

---

78) for how head adjunction and covert movement (or feature movement) can be accommodated by means of Copy and Merge across unconnected phrase-markers.

[16] Actually, this assumption is not so uncontroversial as it seems to be at first sight. See Guimarães (1998: 53-55; 1999) for discussion and criticism.

only these elements have to be represented at PF, and pronounced/heared by the A-P system.[17]

In Kayne's (1994) approach, the necessity for assuming that the LCA applies at all levels of representation comes from the nature of the theory of phrase structure he assumed (*i.e.* X-Bar Theory), and from previous assumptions about the architecture of grammar (namely, that some operations take place at the "covert syntax", with no reflex at PF). He defends that the essential properties of X-Bar Theory actually follow from the LCA. Therefore, if the LCA were only applicable at PF, it would be impossible to block the overgeneration of phrases that violate X-Bar requirements (*i.e.* headless phrases, multi-headed phrases, multi-branching phrases, *etc*.) in the course of the derivation between Spell-Out and LF. In Chomsky's (1995: 334-340) view, once X-Bar Theory has been replaced with Bare Phrase Structure Theory, there is no longer any reason for taking that the LCA applies outside the phonological component. Some principles of X-Bar Theory are rejected (*e.g.* no minimal projection can occupy a specifier or complement position), while other ones continue to be assumed (*e.g.* binary branching, endocentricity, *etc*.), but do not need to be derived from the LCA, because they are taken to follow from other independent assumptions.

In (7), I present the reformulation of the LCA in terms of Bare Phrase Structure Theory. This modification is due to the elimination of the distinction between "terminal element" and "head".

(7) LINEAR CORRESPONDENCE AXIOM: Given any phonologically active minimal projections α & β, α precedes β if and only if either (i) α asymmetrically c-commands β, or (ii) $\exists\gamma \mid \gamma$ dominates α & γ asymmetrically c-commands β.

Although the idea of restricting the LCA to PF-active terminals has a "null-hypothesis" flavor, fitting well into the whole idea of linearization as a *bare output condition* imposed by the A-P system, it is not entirely obvious that it is the right way to

[17] Many works on the syntax-phonology interface have suggested that some syntactic boundaries are relevant to defining prosodic domains (*e.g.* Selkirk 1984, 1986; Nespor & Vogel 1982, 1986; Inkelas & Zec 1990, 1995; Truckendbrodt 1995, 1999; Guimarães 1997, 1998, 1999a, 1999b, *inter alia*), which made Gärtner (1998) hypothesize that *"the mapping into prosodic constituents has to start from ordered syntactic constituents, so the strategy of constructing and evaluating precedence among subtrees rather than precedence among terminals (...) would be vindicated on these grounds"*. If such mapping is to be done in accordance with LCA, this view is problematic. It is impossible to establish a strict linear order among all constintuents (terminals and non-terminals) of a given phrase-marker, because the property of totality cannot be satisfied, since non-terminals will never precede nor follow whatever they dominate or dominates them (*modulo* the non-totality of the c-command relation). Therefore, the LCA would always be violated. Redefining the LCA in terms of partial (non-total) order makes it too weak to retain the idea that all terminals have to be linearized with respect to each other. Also, the fact that syntactic constituency plays a role at PF does not necessarily entails that PF exhibits non-terminal syntactic phrases as such. All it means is that the mapping procedure that establishes prosodic boundaries is sensitive to syntactic constituency, and I have shown in Guimarães (1997, 1998, 1999a, 1999b) that it is possible to define such mapping without having non-terminals in the output from the linearization procedure.

go. Phenomena like *wanna contraction* and the like actually point to the opposite direction. Therefore, we may alternatively think of the LCA as a dumb automaton that blindly linearizes all terminals of the phrase marker, without worrying about their phonological content, which becomes relevant only at a later derivational stage. This makes even more sense if we assume some version of Distributive Morphology (Halle & Marantz 1993) in which the morpho-phonological instantiation of terminals applies over a string of formatives rather than over a phrase marker.[18] Moreover, recent works by Nunes (1995a) and Hornstein (forthcoming) suggest that some form of linearization that piggybacks on asymmetric c-command holds at LF too,[19] which derives many LF-phenomena, like reconstruction, idiom interpretation, quantifier scope, and binding principles. If this is on the right track, we should redefine the LCA as in (7'), in which all terminals have to be linearized, not only those that have an effect at PF. It is this second view that I am endorsing in what follows.

(7') LINEAR CORRESPONDENCE AXIOM: Given any minimal projections α & β, α precedes β if and only if either (i) α asymmetrically c-commands β, or (ii) $\exists\gamma \mid \gamma$ dominates α & γ asymmetrically c-commands β.

3 Dealing with Symmetric C-Command

3.1 Getting Rid of Vacuous Projections Because of the abandonment of X-Bar Theory in favor of Bare Phrase Structure Theory, *a priori* defined structural positions (*i.e.* empty slots, bar levels, *etc.*) are no longer available. None of these notions is necessary to define the relevant syntactic relations (*i.e.* complement, head, specifier, adjunct, maximal, minimal and intermediate projection). Chomsky (1995: 246) categorically claims that one of the consequences of this new approach is the non-existence of any kind of vacuous projection.

> The operation Merge(α,β) is asymmetric, projecting either α or β, the head of the object that projects becoming the label of the complex formed. If α projects, we can refer to it as the *target* of the operation (...). There is no such thing as a nonbranching projection. In particular, there is no way to project from a lexical item α a subelement H(α) consisting of the category α and whatever enters into further computation, H(α) being the actual "head" and α the lexical element itself; nor can such "partial projections" be constructed from larger elements. (Chomsky 1995: 246; my emphasis [MG])

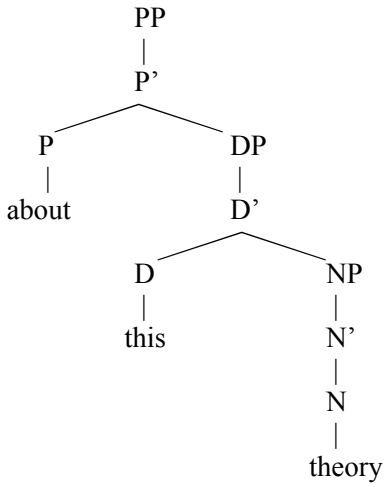
Therefore, an X-Bar structure like (8a) reduces to (8b) in Bare Phrase Structure Theory.

---

[18] I am thankful to Juan Carlos Castillo for pointing out this alternative to me.

[19] In this LF-Linearization, the relation *R* with the properties in (5) is not *Precedence*, but *Scope*.

(8) a.



b. {about {about, {this, {this, theory}}}}

3.2 The Puzzle Notice that **'this'** & **'theory'** mutually c-command each other in (8b), which makes the structure impossible to linearize, given that the LCA is defined in terms of asymmetric c-command. In face of this, either the derivation is cancelled or some additional operation is required to repair the structure, making it linearizable.[20] It is exactly at this point that, according to Chomsky (1995: 337), morphology plays a crucial role. He suggests three ways of handling the problem of linearization of terminals that mutually c-command each other. Two of them are shown below.

> Let us return to the case of L = *m*P with the single-terminal complement *p*, both minimal and maximal. Since neither *m* nor *p* asymmetrically c-commands the other, no ordering is assigned to *m*,*p*; the assigned ordering is not total, and the structure violates the LCA. That leaves two possibilities. Either we weaken the LCA so that nontotal orderings (but not 'contradictory' orderings) are admissible under certain conditions, or we conclude that the derivation crashes unless the structure N = [$_L$ *m p*] has changed by the time the LCA applies so that its internal structure is irrelevant; perhaps N is converted by Morphology to a 'phonological word' not subjected internally to the LCA, assuming that the LCA is an operation that applies after Morphology. (Chomsky 1995: 337)

[20] If we take the LCA to be restricted to PF-active terminals, one way of handling this is to assume that there is a phonologically null functional category (say, light-*n*), projected between DP and NP. The presence of this category creates an asymmetry in the c-command relation between **'this'** & **'theory'**, so that they can be linearized with respect to each other. This seems plausible given that a good amount of evidence for functional projections in all levels of structure has been found over the years. It may be the case that this "bottom-of-the-tree-problem" never really arises at PF because the functional categories conspire to make the LCA fly. However, that might be tricky to defend sometimes, since the motivation for the existence of some categories might be the linearization itself, which makes the argument circular. On the other hand, if we take linearization to apply to all terminals (as I do), any extra functional structure is useless.

The first suggestion certainly is not the null hypothesis if we want to retain the idea that the LCA plays a central role in the grammar. The idea of partial/underspecified/defective orderings that are completed/fixed after the LCA is plausible if at least one of the two elements that stand in a mutual c-command relation is lexically specified as proclitic or enclitic. For example, given the phrase {$\underline{\alpha}$, {α, β}}, where both α and β are terminals, if α is enclitic (and/or β is proclitic), then the order $\beta^{\frown}\alpha$ is established. Conversely, if α is proclitic (and/or β is enclitic then the order $\alpha^{\frown}\beta$ is established. In cases where neither α nor β are phonological clitics, this kind of explanation is problematic, because it predicts an optionality between $\alpha^{\frown}\beta$ & $\beta^{\frown}\alpha$, contrary to the empirical facts (at least in the concrete cases that I know of (*e.g. "let's talk about music" versus "*let's talk music about"*), unless one assumes extra *ad hoc* ordering rules. As far as I can see, these additional rules would be nothing but a value for an ordering parameter (*i.e.* in English, prepositions must precedes their complements), therefore having nothing to do with morphology in the strict sense. If this is correct, the LCA ends up not just being weakened, but also nullified.

The second possibility seems to be the one in which Chomsky is more inclined to adopt, as we can gather from another part of the same text.

> It seems natural to suppose that ordering applies to the output of Morphology, assigning a linear (temporal, left-to-right) order to the elements it forms, all of them X^{0}s though not necessarily lexical items. If correct, these assumptions lend further reason to suppose that there is no linear order in the N→LF computation, assuming that it has no access to the output of Morphology. (Chomsky 1995: 334-335)

The central idea is that, between Spell-Out and the lineaization proceduure, unlinearizable syntactic objects (*i.e.* the ones exhibiting symmetric c-command at the very bottom of the tree) are submitted to morphological transformational rules that convert the two embeddedmost terminals into a single word, making the original terminals two *quasi*-morphemes (ordered with respect to each other by independent, LCA-insensitive, morphological principles) of the brand-new derived *quasi*-word. I will explore this possibility in details in the next section.

The third suggestion made by Chomsky (1995: 337) is that we may consider the hypothesis that there are some instantiations of movement that are triggered by the necessity of linearization.

> In short, if the complement is a single-terminal XP, then it must raise overtly. (Chomsky 1995:337)

I refer the reader to Moro (1997) and Uriagereka (1998: 220-222) for two ways of exploring this idea. For now, I would just like to point out that, in order for this to work, it is necessary to adopt a more general version of *last resort* (*i.e.* not in terms of feature checking, but in terms of necessity for convergence). Actually, I do not think this assumption is problematic by itself. However, in most cases, it has the consequence that it is impossible to predict which one of the two heads standing in mutual c-command

relation must undergo movement (violating either *last resort* (in the strict sense) or *procrastinate*).

Finally, it is worth mentioning that the second and the third possibilities may collapse under a more general movement-based approach if the establishment of complex post-lexical words at the very bottom of the tree is viewed as an incorporation procedure, as suggested by Uriagereka (1998:215-224).

3.3 Where's Morphology? Let us now consider in detail Chomsky's hypothesis of collapsing the two embeddedmost terminals at the very bottom of the tree into a single derived terminal.

First of all, it is worth taking a closer look at the empirical arguments usually offered by those who advocate for this view. Consider the interesting and well-known pattern in (9) below.

(9) a. * I like her.
b. I lik'er.
c. I like HER (not you).

The generalization that we can gather from these data is that English accusative pronouns must cliticize to the verb unless focused. If the pronoun is fully pronounced, the sentence is ungrammatical, like in (9a); but the ungrammaticality disappears once the pronoun is reduced and cliticized to the verb, like in (9b). Accusative pronouns can surface in their unreduced and uncliticized forms if and only if the intended meaning involves a narrow focus on the pronoun, like in (9c).

At first sight, it seems that Chomsky's hypothesis shown in the previous section straightforwardly accounts for these data.[21] The structure behind (9a) would be the one in (10a), where no remorphologization took place to get rid of the symmetric c-command. Consequently, the linearization algorithm failed to establish a precedence relation between **'like'** and **'her'**, leading to the cancellation of the derivation. The structure behind (9b) would be the one in (10b), where the pair of words **'like'** & **'her'** were turned into the word ***'lik'er'***. Presumably, the output of the remorphologization rule is ***'lik'er'*** instead of ***'herlike'*** due to purely morphological reasons (*i.e.* **'her'** is lexically specified as enclitic). Finally, the structure behind (9c) would be the one in (10c), where there is an extra category, the Focus Phrase, intervening between the verb and the pronoun. Besides being responsible for the focus interpretation at LF, the presence of this extra category triggers remorphologization of the mutual c-commanding terminals **'F'** and **'her'**, making them collapse into a single terminal ***HER*** (stressed **'her'**), as in (10d), where nothing cliticizes into the verb.[22]

[21] I refer the reader to Uriagereka (1998: 215-224) for a slightly different version of the same idea, in which the cliticization of the pronoun involves head-movement.

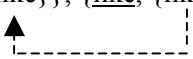
[22] If we assume that the LCA only sees PF-active terminals, the presence of FP causes **'like'** to asymmetrically c-command **'her'**. Consequently, **'like'** must precede **'her'**, and no cliticization is necessary (actually, it should be ruled out on economy grounds, unless it is required for extra – non-LCA-related – reasons).

(10) a. VP = {<u>like</u>, {like, her}}
b. VP = {<u>like</u>, {*lik'er*}}
c. VP = {<u>like</u>, {like, {<u>F</u>, {F, her}}}}
d. VP = {<u>like</u>, {like, {<u>F</u>, {*HER*}}}}

So far so good, but this is not the whole story. First of all, as pointed out to me by John Drury (personal communication), the intuition behind the generalization in (9) is not shared by all native speakers of English.

Moreover, there is evidence that, in English, V moves at least to light-*v* (*e.g.* structures with ditransitive verbs),[23] which implies that the actual structure behind (9a) is a configuration where **'like'** asymmetrically c-commands **'her'** to begin with, like in (11). If so, we cannot take symmetric c-command as the trigger of cliticization.[24]

(11) *v*P = {<u>*v*</u>, {{<u>≤*v*,*v*></u>, {*v*, like}}, {<u>like</u>, {~~like~~, her}}}}



Finally, it is a robust fact about English that, within the subset of those speakers whose judgments correspond to (9) above, all of them have crystal clear intuitions that the very same pattern of cliticization is found wherever there is an accusative pronoun[25]. It is irrelevant if the pronoun stands in symmetric c-command relation with the previous word or not, as shown in (12).[26]

(12) a. * I believe her to be guilty.
b. I believ'er to be guilty.
c. I believe HER to be guilty (not you).

The facts force us to conclude that, in the dialects that exhibit the pattern in (9) & (12), the cliticization is due to some independent and intrinsic prosodic property of the pronoun, and has nothing to do with symmetric c-command.[27/28] Therefore, the idea that morphology plays a role in linearization lacks empirical support.

---

[23] The actual substantive properties of the higher V are orthogonal to this point. The reader can take his/her favorite version of the VP shell hypothesis. The relevant fact is that *"John gave the book to Mary"* is grammatical, whereas *"*John the book gave to Mary"* is not.

[24] See Cardinaletti & Starke (1994) and Castillo (1999) for a more sophisticated theory of pronouns, in which pronoun cliticization to verbs has to do with structural deficiency.

[25] I am thankful to Norbert Hornstein for discussion on this topic, and for the judgments too.

[26] In these examples, the relevant structure is **[^{TP} I believe [^{TP} her to be guilty]]**, where **'believe'** asymmetrically c-commands **'her'**.

[27] Here there are two possibilities. Either the cliticizations in (9b) and (12b) are both the PF reflex of a head adjunction movement (*i.e.* the pronoun adjoining to the verb), triggered by some other factor completely independent from linearization; or they are the result of a late phonological process that collapses terminal symbols, operating not on the phrase marker, but on strings of words generated by the linearization procedure (cf. next note). Either way, the cliticization is not triggered for linearization purposes.

Let us now move from empirical considerations to conceptual issues regarding Chomsky's story about the linearization of heads at the very bottom of the tree. Consider again the structure in (8b), repeated below as (13).

(13) K = {about {about, {this, {this, theory}}}}

Before linearization takes place, the constituent that immediately dominates **'this'** and **'theory'** undergoes a transformational rule of the morphological subcomponent. After this, its internal structure is lost (or, at least, becomes irrelevant, *i.e.* invisible to the linearization algorithm). Therefore, the constituent **L = {this, {this, theory}}** is – roughly speaking – converted into the *quasi*-word ***'thistheory'***, so that **'this'** & **'theory'** turn out to be the *quasi*-morphemes of this *quasi*-word, and the ordering among them is given by independent principles of the morphological subcomponent, to which the LCA is completely indifferent.[29]

Making use of the tools available in Chomsky's (1995) system, the null hypothesis about how this transformation of phrases into words takes place is the following. Before being subjected to linearization, the phrase marker L is rewritten by the morphological subcomponent, such that its two terms are collapsed into a single one, while its label remains intact.[30] Consequently, K is automatically rewritten as K',[31] which has the structure in (14).

---

[28] Notice that the analysis in (10) also needs to assume that **'her'** is lexically specified as enclitic, otherwise it is impossible to predict why the derived word is ***'lik'er'*** rather than of *'**herlike'***. Usually, syntacticians tend to view phonological cliticization as a one-to-one correlation between syntactic heads (*i.e.* X^0) and prosodic words. I do not share the same opinion. In Guimarães (1998; 1999a), I suggest a particular version of Prosodic Phonology from a minimalist perspective, in which the phonological component has its own principles and mechanisms of cliticization. That is, given a string of words $\alpha^\frown\beta^\frown\gamma^\frown\delta$, it may be the case that two (or more) adjacent symbols (say, β & γ) behave like a single prosodic word (*i.e.* the substring $\beta^\frown\gamma$ contains only one primary stress) even though they do not correspond to a single X^0 at the syntactic level. This does not mean that syntactic information is irrelevant. In fact, I claim that syntactic constituency (specifically, asymmetric c-command) plays a crucial role, determining which substrings of distinct syntactic words can behave as a single prosodic words, or as phonological phrases, or as intonational phrases.

[29] The same logic of the argument I presented against the idea of incomplete linearizations applies here. That is, if α is proclitic (and/or β is enclitic), then $\{\underline{\alpha}, \{\alpha, \beta\}\}$ turns into the *quasi*-word $\#\alpha\beta\#$; and if α is enclitic (and/or β is proclitic), then $\{\underline{\alpha}, \{\alpha, \beta\}\}$ turns into the *quasi*-word $\#\beta\alpha\#$. On the other hand, if neither α nor β are phonological clitics, we wrongly predict an optionality between $\#\alpha\beta\#$ & $\#\beta\alpha\#$ (again, unless one assumes extra *ad hoc* ordering rules).

[30] Notice that, by the definition of syntactic object in (1), ***'thistheory'*** does not qualify as a syntactic object. Consequently, every set in which ***'thistheory'*** is embedded is not a syntactic object either. Moreover, depending on one's previous assumptions about the ontology of labels, and about the technical details of the definitions of syntactic object and Merge (cf. (1) and (2) above), the new phrase **L' = {this, {*thistheory*}}** may not qualify as a legitimate syntactic object either, even if we take ***'thistheory'*** to be a syntactic object,. In (14), the label of **L'** not does not equal the head of one of its terms. Probably, if **L'** does not qualify as a legitimate syntactic object to begin with, linearization fails to apply. Consequently, the terminals of **L'** cannot be linearized,

(14) K' = {about {about, **{this, {*thistheory*}}**}}

Notice however that this formalism implies the generation of a vacuous projection. The original phrase is **L = {this, {this, theory}}**, whose subconstituents are **'this'** & **'theory'**. After the transformation, the new phrase **L' = {this, {*thistheory*}}** obtains. It dominates only one syntactic object, namely: ***'thistheory'***.[32] In the new configuration **K'** in (14), there are no longer three terminals, but only two, namely **'about'** and ***'thistheory'***. Given that **'about'** asymmetrically c-commands ***'thistheory'***, the phrase marker is now perfectly linearizable. The output from linearization is the string **about**$^{\frown}$***thistheory***.

Consider now a slightly different version of this idea. If we assume that the morphological subcomponent turns **L = {this, {this, theory}}** into ***'thistheory'***, affecting not only the terms of K, but also its label, then (15) is generated from (13).

(15) K' = {about {about, ***thistheory***}}

This kind of transformation is even more radical than the one in (14), because the output of the morphological operation is a completely atomic element from the syntactic point of view, *i.e.* a minimal-&-maximal projection, with neither a label nor internal syntactic hierarchy (although there may be some non-syntactic, purely morphological, hierarchy).

Notice that this strategy solves one linearization problem by creating another one. As in (14), the configuration in (15) no longer has three terminals, but only two, namely **'about'** and ***'thistheory'***. The problem is that **'about'** and ***'thistheory'*** symmetrically c-command each other in (15). Consequently, the phrase marker remains unlinearizable.

Interestingly, the reason why (14) works while (15) doesn't is the very presence of a vacuous projection, bringing the asymmetry that makes the linearization possible. Otherwise, the LCA is violated. Remind that vacuous projection are taken to be impossible to generate via Merge in the core syntax, according to Chomsky (1995: 246).

---

causing the derivation to be cancelled. If this is true, we should rethink the product of the morphological transformation as being **L' = {*thistheory*, {*thistheory*}}**. Perhaps, these are not real problems. Since (14) is constructed outside the core syntax, it may be true that, within the morphological subcomponent, certain small details about phrase structure are irrelevant or undistinguishable, so that a more flexible notion of syntactic object holds at that level.

[31] Strictly speaking, not only K and L are rewritten, but also all other phrases that happen to dominate L.

[32] If we take morphological fusion to be an instance of head incorporation, we first get (14)', in which **'this+theory'** is not a terminal, and then (14)'', in which remorphologization has applied. But since the lower copy of **'theory'** is deleted at some point (before or after remorphologization), eventually what we actually get is (14)''', which is the same structure as (14) above (but see note 38).

(14') K' = {about {about, {this, {{<this,this>, {this, theory}}, theory}}}}

(14'') K'' = {about {about, {this, {*thistheory*, theory}}}}

(14''') K''' = {about {about, {this, {*thistheory*}}}}

If some version of the analysis described in (14) is correct, we are making the claim that morphology precedes the linearization procedure in the course of derivation, somewhere between Spell-Out and PF. A problem with this assumption is that empirical facts lead us to assume that grammars work the other way around, with morphology applying to the output from linearization.

Very clear counter-evidence for the idea of pre-linearization morphology can be found in the distribution of supletive possessives in French. In this language, possessives must agree both in gender and number with the head of the NP they are related to. Thus the masculine form **'mon'** (*"my"*) is not legitimate in (16a) because **'soeur'** (*"sister"*) is a feminine noun; while, in (16b), the feminine form **'ma'** (*"my"*) is not legitimate because **'frère'** (*"brother"*) is a masculine noun. However, although in principle feminine nouns require the possessive to be inflected as feminine, a supletive form homophonous to the masculine one is the only legitimate option in cases where the possessive immediately precedes a word that starts with a vowel. Such word can be a pre-nominal modifier, as in (16c), or even the feminine noun itself, like **'actrice'** (*"actress"*) in (16d). The possessive agrees in gender with the noun if the pre-nominal modifier in between the possessive and the feminine noun starts with a consonant, as in (16e).

(16) a. ✓ma/*mon soeur s'appelle Marie.
✓*my*[FEM]/**my*[MASC] *sister REFL-calls Mary*
'My sister calls Mary'

b. *ma/✓mon frère s'appelle Jean.
**my*[FEM]/✓*my*[MASC] *brother REFL-calls John*
'My brother calls John'

c. *ma/✓mon admirable soeur s'appelle Marie.
**my*[FEM]/✓*my*[MASC] *admirable sister REFL-calls Mary*
'My admirable sister calls Mary'

d. *ma/✓mon actrice préférée est Danica McKellar.
**my*[FEM]/✓*my*[MASC] *actress preferred is Danica McKellar*
'My favorite actress is Danica McKellar'

e. ✓ma/*mon jeune actrice préférée est Danica McKellar.
✓*my*[FEM]/**my*[MASC] *young actress preferred is Danica McKellar*
'My favorite young actress is Danica McKellar'

Clearly this phenomenon is sensitive to adjacency. The (il)legitimacy of a supletive form exclusively depends on the phonological properties of the next symbol in the string of words. It is obvious that any operation or filter that is sensitive to adjacency is also sensitive to linear order, since there can be no adjacency at all without linear order. Consequently, whatever the morphophonological mechanism involved in this supletism phenomenon is, it takes place in some derivational stage after the linearization. This does

not necessarily mean that all morphology must happen after linearization, but I take this to be the null hypothesis.[33]

Once we assume that morphologization follows the linearization algorithm, we are forced to deny the idea that terminals standing in a symmetric c-command relation are "turned into a word" by the morphological component before being linearized. We, then, leave ourselves with the task of giving an alternative solution for the puzzle of how to linearize mutually c-commanding heads at the very bottom of the tree. This is what I do in the next section, by reincorporating the notion of vacuous projection into the theory of phrase structure.

4 Self-Merge and Vacuous Projections

All the conceptual and empirical points discussed in the previous section strongly suggest that only vacuous projections are the most natural way of making possible the linearization of the two embeddedmost heads of the phrase marker. If this is true, there is something awkward about the model of grammar. The theory of phrase structure predicts that vacuous projections cannot exist. Nevertheless, there are many cases in which the output of the syntactic component has to be transformed into a structure that contains a vacuous projection, otherwise the linearization is impossible. Intuitively, if vacuous projections are really necessary, it would seem more natural if they were cyclically and monotonically constructed by the syntactic component itself, instead of being the product of an extra transformational rule. Therefore, none of the problems discussed above would exist if the a structure like (17), containing a vacuous projection (*i.e.* **{theory, {theory}}**), could be a legitimate output from Spell-Out.

(17) K = {about {about, {this, {this, **{theory, {theory}}**}}}}

The most natural way to avoid the problem, then, is to assume once for all that the syntactic component can generate vacuous projections. This seems to corroborate Kayne's (1994: 8) original position that the complement of a head cannot be another head. Should this be taken as a definitive argument against Bare Phrase Structure Theory? My goal here is to show that, contrary to what it may appear at first sight, vacuous projections are not incompatible with Bare Phrase Structure Theory.

---

[33] In fact, Nunes (1999, 2000c) shows some evidence for his hypothesis that multiple pronunciation of chain links is a reflex of pre-linearization morphological transformations, which turn a complex phrase into a post-lexical derived terminal that contains – as one of its *quasi*-morphemes – a word that has a copy somewhere else in the tree. The author claims that the LCA cannot see inside terminals, therefore a copy that is hidden inside a post-lexical derived terminal is not deleted by the chain-reduction algorithm, and is pronounced together with its other copy/copies. Although Nunes's analysis is perfectly coherent with the facts, I remain skeptical about it because, in order to explain those extra-ordinary linearization patterns, it requires some extra assumptions to account for ordinary chains formed by head movement, otherwise it incorrectly predicts multiple pronunciation of heads all over the place. Of course, there might be a way out of this problem; and, if what Nunes says turns out to be true, then most of what I am saying needs to undergo major revision.

In principle, nothing in Chomsky's system prohibits a syntactic object α from merging with itself. The only explicit statement is that the operation Merge takes two syntactic objects **α** & **β** as the input, and yields a more complex structure by projecting either **α** or **β**, generating the syntactic object **K = {γ, {α, β}}** as the *output*, where **γ** (the label of **K**) corresponds to either the head of **α** or the head of **β**, depending on which of the two merged elements projects (Chomsky 1995: 241-249). By this very definition, nothing precludes **α** from being identical to **β**, which means that Merge is being applied to a single element.

As Collins (1997) admits, the idea/claim that such possibility is incompatible with Bare Phrase Structure Theory has no independent motivation, and, thus, relies on a stipulation.

> We must stipulate that Merge has as a requirement that the set Σ' it takes as an argument has two or more elements. This kind of stipulation, since it follows from the most basic consideration about the empirical adequacy of the theory and since it must be made universally, seems to be a rather minimal departure from a system with no assumptions at all about the form of Merge. (Collins 1997: 81; my emphasis [MG])

Here I propose to get rid of this stipulation. The conceptual motivation for this abandonment is obvious. A stipulation should only be assumed as a last resort, if it is the only way to account for the empirical facts. In the case under discussion, the stipulation is the only responsible for the linearization puzzle at the very bottom of the tree. By getting rid of this stipulation, we get the possibility of vacuous projections back, which solves/eliminates the linearization puzzle.

Let us now see why and how vacuous projections can be generated without changing anything in Chomsky's (1995) system (except the stipulation just mentioned). Consider that there are three isolated heads (lexical items): **'about'**, **'this'** & **'theory'**. The computational system, then, merges **'this'** with **'theory'**, projecting **'this'**, and generating the output **{this, {this, theory}}**, as in (18). In the next step, **'about'** is merged with the phrase **{this, {theory}}**, such that **'about'** projects. The output of this second Merge is K = **{about, {about, {this, {this, theory}}}}**, as in (19).

(18) input: α = **this** & β = **theory**
output: L = **{this, {this, theory}}**

(19) input: α = **about** & β = L = **{this, {this, theory}}**
output: K = **{about, {about, {this, {this, theory}}}}**

Actually, there is nothing new in what I just said. Now look how this very same formalism can generate a vacuous projection. As I have already pointed out, there is absolutely nothing in the theory requiring that α & β must be distinct (unless we introduce this prohibition as an additional unmotivated axiom). Consider the following derivation. In the first derivational step, there are three isolated heads (lexical items): **'about'**, **'this'** & **'theory'**. Then the computational system merges **'theory'** with **'theory'** itself, projecting **'theory'**. I call this operation Self-Merge. The output of this

instance of Self-Merge is **M = {theory, {theory, theory}}**. Notice that this structure includes the set **{theory, theory}**. According to the Extensionality Axiom of Set Theory, the set **{theory, theory}** is exactly identical to the set **{theory}**, since both have exactly the same member[34] (cf. Partee, ter Meulen & Wall 1993: 8-9; *inter alia*). Consequently, the phrase just generated is **M = {theory, {theory}}**. In the next derivational step, the computational system merges the syntactic objects **'this'** & **{theory, {theory}}**, projecting **'this'**. The output of this Merge operation is the phrase **L = {this, {this, {theory, {theory}}}}**. After this, the constituents **'about'** & **{this, {this, {theory, {theory}}}}** are merged, and **'about'** is projected. The new phrase **K = {about, {about, {this, {this, {theory, {theory}}}}}}** is, then, generated. This derivation is described in (20-22) below.

(20) input: α = **theory** & β = **theory**
output: M = **{theory, {theory}}**

(21) input: α = **this** & β = M = **{theory, {theory}}**
output: L = **{this, {this, {theory, {theory}}}}**

(22) input: α = **about** & β = L = **{this, {this, {theory, {theory}}}}**
output: K = **{about, {about, {this, {this, {theory, {theory}}}}}}**

We can easily identify in K the existence of a vacuous projection, namely: M = **{theory, {theory}}**. This is what ensures the asymmetry of c-command relations among **'about'**, **'this'** & **'theory'**. No linearization problem arises, and the string ***about*$^\frown$*this*$^\frown$*theory*** is generated in the phonological component from K, according to the LCA.[35]

---

[34] According to the principles of (Cantorian) Set-Theory, set-membership is defined in such a way that an element α can be a member of a set Z only once (cf. Partee, ter Meulen & Wall 1993: 6; *inter alia*). Given any object α and any set Z, either α is a member of Z or not. There is no such thing as partial, multiple or gradual set-membership. Therefore, $\{\alpha\}$, $\{\alpha, \alpha\}$, $\{\alpha, \alpha, \alpha\}$, $\{\alpha, \alpha, \alpha, \alpha\}$, *etc*... are just notational variants for the very same set (but see Zadeh (1987) for a relativization of the notion of set-membership, under the framework of Fuzzy Set Theory).

[35] According to what I have said so far, if the syntactic component is supposed to built a DP from **'this'** & **'theory'**, it has to vacuously project **'theory'** first, and then it merges the resulting vacuous projection, *i.e.* **{theory, {theory}}**, with **'this'**, generating **{this, {this, {theory, {theory}}}}**. This looks trivial. However, as pointed out to me by John Drury (personal communication), there is another logical possibility that we should consider. Instead of doing what I just described, the system may self-merge **'this'**, and then merge the resulting vacuous projection, *i.e.* **{this, {this}}**, with **'theory'**, generating **{this, {theory, {this, {this}}}}**, in which is **'theory'** is a specifier of **'this'**, yielding the (unattested) word order **theory$^\frown$this**. What rules this out? Very tough question! The mechanics of Self-Merge alone does not excludes this possibility. Perhaps, this has to do with some interpretive principles at LF, which require configurations in which some kinds of terminals "scope over" (*i.e.* asymmetrically c-command) other kinds of terminals, in case they merge with each other, as a way of encoding the asymmetry of their

I would like to emphasize that Self-Merge does not have any independent theoretical status, being nothing but one particular instance of the Merge operation itself. The terminology Self-Merge is only a mere expository device. Another point that deserves to be emphasized is that Self-Merge is the merging of a given syntactic object α with α itself, and not the merging of α with a copy of α. According to the formalism here assumed (based on the Extensionality Axiom of Set Theory), no copying operation is necessary. Following this line of reasoning, it is not necessary either to assume that a vacuous projection **{$\underline{\alpha}$, {α}}** is formed from two occurrences/tokens of the lexical item α in the Numeration. If there are neither copies nor distinct occurrences, then no deletion operation is necessary.

In order to understand the logic of the Self-Merge operation, it is necessary not to get mislead by the fact that there is a lexical item that appears twice in the notation. An analogy with arithmetic might be helpful here. Given the equation **x + y = z**, if **x = 1** & **y = 3**, then **z = 4**, since **x + y = 1 + 3 = 4**. Given the same equation **x + y = z**, if **x = 3** & **y = 3**, then **z = 6**, since **x + y = 3 + 3 = 6**. In the later case, the number **3** appears twice in the notation (**3 + 3**), but it does not mean that two copies or two occurrences of the number **3** have been added to one another. Strictly speaking, there is only one number **3**, which has been added to itself. We can call this "self-addition" just for expository reasons, but this new concept has no independent theoretical status at all. In fact, it is the very same addition operation involved in **1 + 3 = 4**. It is exactly this kind of reasoning that I want the reader to have in mind when I say that **'theory'** is merged with **'theory'**, generating M = **{<u>theory</u>, {theory}}** in (20). Strictly speaking, **'theory'** is merged with itself.[36]

5 Some Notes on Economy

In the system I am assuming here, there is a vacuous projection at the very bottom of every (sub)tree, and nowhere else. Such vacuous projections are generated by Self-Merge, which is an instance of Merge triggered by linearization, instead of feature-checking or selectional requirements. This raises many questions regarding the architecture of the grammar as a whole. I would like to briefly address three of them below.

5.1 Blocking Overgeneration of Vacuous Projections Notice that any vacuous projection generated by Self-Merge plays a role at the interfaces only when it is absolutely required in order for linearization to successfully apply. Under minimalist assumptions, any other instance of Self-Merge is excluded on economy grounds. For example, vacuous projections are always illegitimate if they occupy a specifier position, as in (23), because they involve an unnecessary additional derivational step (*i.e.* an instance of Merge). There is no problem with making a minimal-&-maximal projection a

---

selectional requirements. Of course, this is just an informal speculation, not a solution to the problem. I have no satisfactory story to tell up to now.

[36] From this perspective, then, it is misleading to say that Merge is an asymmetric (therefore, irreflexive) operation, as Chomsky (1995: 246) does. Rather, it is more accurate to say that Merge is non-symmetric, because it may or may not apply to a single element, in a reflexive fashion, depending on the context.

specifier (*e.g.* **'he'** in (24)), because it will asymmetrically c-command whichever terminals its sister may dominate[37] (*e.g.* **'he'** asymmetrically c-commands **'writes'**, **'about'** and **'minimalism'**).

(23) * {writes, {**{he, {he}}**, {writes, {writes, {about, {about, {minimalism, {minimalism}}}}}}}}

(24) {writes, {**he**, {writes, {writes, {about, {about, {minimalism, {minimalism}}}}}}}}

If Self-Merge is nothing but an instance of ordinary Merge applied to a single syntactic object, then it must be subjected to derivational economy conditions as any other operation. Thus, either an instance of Self-Merge is obligatory or it is prohibited.

We can approach this economy issue from two perspectives: representational and/or derivational (cf. Chomsky 1995: 150-151).

> [D]erivations must be as economical as possible: there is no superfluous rule application. (...) The analogous principle for representations would stipulate that, just as there can be no superfluous steps in derivations, so there can be no superfluous symbols in representations. This is the intuitive content of the notion of Full Interpretation (FI), which holds that an element can appear in a representation only if it is properly "licensed". (Chomsky 1995: 150-151)

These two perspectives are intrinsically related, because vacuous projections are the representational reflex of applications of Self-Merge in the derivation. Thus, Self-Merge and vacuous projections are like two sides of the same coin.[38] From the representational viewpoint, vacuous projections are legitimate if and only if they are

---

[37] Interestingly, Kayne (1994), in his project to derive X-Bar Theory, has succeed in demonstrating that complements must not be heads, but failed in demonstrating that specifiers must not be heads (see Kayne (1994: 30-32) for his stipulative argumentation). But the point here is not just conceptual. Rather, it has interesting empirical consequences. In Guimarães (1998: chapters III & IV), I show that there are phonological processes that apply across word boundaries, which – among other things – require c-command between the two adjacent words whose edges are affected (these are basically the processes that have the Phonological Phrase as their domain of application, as usually claimed in Prosodic Phonology). Stress shift in Brazilian Portuguese and *liaison* in French are examples of such processes. What is interesting about them is that, in typical SVO structures, they never apply across the last word of the subject and the following verb, except when the subject is pronominal. This strongly suggests that: (i) nominative pronouns in Romance are atomic constituents; and (ii) they asymmetrically c-command the following verb, which means that they do not vacuously project.

[38] In fact, every application of Self-Merge yields a vacuous projection. But it is does not necessarily mean that every vacuous projection is the product of a Self-Merge operation. Consider the syntactic object Y = {$\underline{\alpha}$, {α, β}} after β has been moved out from Y. The question is: is the structure Y' = {$\underline{\alpha}$, {α, ~~β~~}} (generated outside the core syntax, after the deletion of the original copy of β) identical to the structure Z = {$\underline{\alpha}$, {α}}, and therefore a vacuous projection? The answer to this question depends on a prior theoretical position regarding what are the actual effects of the deletion operation, and whether or not both Delete & Erasure exist as two distinct operations (cf. Nunes 2000a).

strictly necessary for getting the asymmetric c-command between the two embeddedmost terminals of the phrase marker, meeting the requirements of the LCA. Otherwise, vacuous projections constitute a violation of representational economy, because they contain extra symbols that play no role at the interfaces. From the derivational viewpoint, applications of the Self-Merge operation are legitimate if and only if they are absolutely necessary to construct a vacuous projection that does not violate representational economy (in the sense just mentioned above). Otherwise, any instance of Self-Merge violates derivational economy, because they introduce an additional derivational step for no rhyme or reason. In these cases, the evaluation metrics determines that the derivation with Self-Merge must be excluded in favor of a competitor derivation which does not involve such Self-Merge operation.

5.2 Last Resort Recently, Collins (1997) and Chomsky (1998) have proposed that economy is evaluated locally, at each and every instance of Merge. Therefore, Merge is not a costless operation. Rather, it has a cost, and, like any other operation, is also subjected to economy principles. They claim that when α merges with β, at least one property of α or β is being satisfied. If this is not the case for a particular instance of Merge, then it is ruled out on economy grounds. If we take this approach to be correct, does it mean that Self-Merge is impossible?

One way of implementing the idea of Merge as a *last resort* is in terms of feature-checking. That is, it is legitimate to merge α with β if and only if at least one feature α or β is being checked at this step, and there is no other way of checking that feature. Certainly, if we define *last resort* solely in terms of feature-checking, then the hypothesis of Self-Merge has to be wrong, because we have absolutely no evidence that a feature of α is being checked when α merges with itself.

Notice, however, that this conception of *last resort* is different from what both Collins (1997) and Chomsky (1998) had in mind.

Collins (1997: 65-67) explicitly says that *"last resort should make reference to any property that is relevant to the internal operations of the syntactic computation"*. Therefore, feature-checking is only one possible trigger for Merge, among others.

Chomsky (1998) is less explicit about this issue, but his system clearly requires a broader conception of *last resort*, since he assumes – as Collin (1997) also does – the standard view that theta-role assignment is a result of interpretive procedures over configurations (cf. Halle & Keyser 1993) instead of feature-checking (*contra* Bošković 1994; Bošković & Takahashi 1998; Hornstein 1999, forthcoming), which makes him think of it in terms of satisfaction of selectional properties of predicates. Moreover, in Chomsky's (1998) system, set-merge (substitution) is greedy, but pair-merge (adjunction) is not.

According to this more flexible conception of *last resort*, it is natural to assume that an instance of Self-Merge is legitimate if and only if it is the only way of making the structure linearizable. But this raises a further point, which I address in the next section.

5.3 Look-ahead From the minimalist point of view, it is preferable to have a system in which economy is calculated in a local fashion, based only upon the information present

in the input to the operation that is being executed. However, if linearization is an extra-syntactic mechanism that only takes place when the phrase marker is shipped to the interfaces, why should syntactic derivations care about this? How does the syntactic component know that terminals are supposed to be linearized with respect to each other at PF/LF, and how does it know that applying Self-Merge in the very beginning of the derivation will make linearization possible?[39] This implies that the system needs to know that the first two terminals to undergo Merge will eventually have to be linearized with respect to each other. But this is a global property of linguistic objects at the interfaces, not a property of the terminals themselves. Therefore, in order for the system to apply Self-Merge, it has to forsee/anticipate the interfaces; whereas ordinary Merge does not entail look-ahead, since it is mechanically/automatically triggered by feature-checking, on the basis of information that can be locally accessed.

This look-ahead problem disappears once we find a way to make the syntactic component able to evaluate convergence for pieces of structures incrementally, as the derivation goes, in a local fashion, without proliferating intermediate levels of representations, nor anticipating interface levels, as suggested by Chomsky (1998).

> The property *converges at [an] I[nterface] L[evel]* may hold of an expression formed in the course of a derivation that then proceeds on to IL. If, say, particles or adverbs have only LF-interpretable features, then they converge at LF when extracted from the lexicon and at every subsequent stage of derivation to LF. Similarly, an embedded clause may converge, for example, the bracketed subpart of "John thinks [it is raining]". The phrase "converge at an interface" should not mislead: convergence is an internal property of an expression, detectable by inspection. (Chomsky 1998: section 2)

Following this intuition, Boeckx (1999) develops the notion of "invasive interfaces" – adapted from Epstein, Groat, Kawashima & Kitahara (1998) – in the context of Uriagereka's (1999) Multiple Spell-Out model.

> I would like to see the interfaces as opportunistic, capitalizing on a syntactic configuration as soon as they can – that is, as soon as it is convergent for them. It is important to distinguish this sense of convergence from the more familiar one, which sees convergence as a sanction by the interface put on a derivation. Familiar convergence is understood representationally (a characteristic of the whole derivation), which I have argued elsewhere (...) is misleading. Convergence as understood here applies to portions of the derivation. Interestingly, then, it is possible for a derivational chunk to converge for some aspect/subcomponent of the interface. (...) Here, Spell-Out is not a syntactic operation (not even 'final' Spell-Out, when the derivation has exhausted all the material in the numeration); the interface is active. (Boeckx 1999)

If something along those lines is correct, we can assume that the system works in a radically dynamic fashion, with linearization being an operation of the phonological

[39] So, it makes us wonder whether it implies that the LCA is somehow embodied within the core syntax itself (as first suggested by Kayne 1994), rather than a mapping procedure external to the core syntax (as Chomsky (1995) proposes).

component that "invades" syntactic objects incrementally, after every merge, or even in cascades, like in Uriagereka's (1999) Multiple Spell-Out model, or in "phases" like in Chomsky's (1998, 1999) system. In parallel with the step-by-step construction of the phrase maker from the bottom upwards, the PF-component generates a string of sounds step-by-step, in a right-to-left fashion, and the LF-component generates a linear LF-object encoding scope relations step-by-step. That way, the computational complexity of Self-Merge is minimal (or even null), because the decision on when to apply it is made locally. The choice between merging a syntactic object α with itself or with some other syntactic object β is made at that point, solely on the basis of the input and the output of that particular instance of Merge, with no look-ahead at all.

6 Conclusion

The generation of vacuous projections via Self-Merge solves the linearization problem faced by the standard version of Bare Phrase Structure Theory in a straightforward way. The most important aspect of this move is that it changes absolutely no theoretical assumptions with respect to the standard version, except for the abandonment of the stipulation that the two constituents being merged cannot be identical. Such stipulation plays no role in the theory but causing linearization problems.

Acknowledgments

This paper is a development of chapter V of my Master's Thesis (Guimarães 1998). Some parts of previous versions of this study have been presented at the *XLVI Seminário do GEL* (UNESP,São José do Rio Preto, July 27th 1998), and at the *University of Maryland Fall 1998 Linguistics Department Student Conference* (UMD, October 22nd 1998). I am thankful to those audiences, and in particular to Thaïs Chaves, John Drury, Juan Carlos Castillo and Norbert Hornstein for their many insightful comments and suggestions, in various stages of this research. The present version was written under financial support from CAPES (Brazilian Government), grant # 2049/97-7, and from the Graduate School of the University of Maryland at College Park.

References

Baker, M. & O. T. Stewart. 1999. *On Double-Headness and the Anatomy of the Clause*. Unpublished. Rutgers University.

Bobaljik, J. & S. Brown. 1997. Inter-Arboreal Operations: Head Movement and the Extension Requirement. *Linguistic Inquiry*, 28: 345-356.

Boeckx, C. 1999. *Relativizing Spell-Out*. Unpublished. Storrs: University of Connecticut.

Bošković, Z. 1994. D-Structure, Theta-Criterion, and Movement into Theta Positions. *Linguistic Analysis*, 24: 247-286.

Bošković, Z & D. Takahashi. 1998. Scrambling and Last Resort. *Linguistic Inquiry*, 29: 347-336.

Cardinaletti, A. & M. Starke. 1994. *The Typology of Structural Deficiency: On the Tree Grammatical Classes*. ms. University of Geneva.

Castillo, J. C. 1999. From Latin to Romance: The Tripartition of Pronouns. in: S. Aoshima, J. Drury & T. Neuvonen (eds.) *University of Maryland Working Papers in Linguistics*, 8: 43-65.

Chametzky, R. 1996. *A Theory of Phrase Markers and the Extended Base*. Albany, NY: State University of New York Press.

Chomsky, N. 1970. Remarks on Nominalization. In: R. Jacobs & P. Rosembaum (eds.), *Readings in English Transformational Grammar.* Waltham: Ginn and Company.

Chomsky, N. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Berlin/New York: Mouton de Gruyter.

Chomsky, N. 1986a. *Barriers*. Cambridge: The MIT Press.

Chomsky, N. 1986b. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Chomsky, N. 1994. *Bare Phrase Structure*. Cambridge, Massachusetts: MIT Occasional Papers in Linguistics 5. MITWPL, Department of Linguistics and Philosophy, MIT [Published in G. Webelhuth (ed.) 1995. *Government and Binding Theory and the Minimalist Program*. Oxford: Blackwell].

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

Chomsky, N. 1998. *Minimalist Inquires: The Framework.* MIT Occasional Papers in Linguistics, 15. [Published in R. Martin, D. Michaels & J. Uriagereka (eds.) 2000. *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 89-155. Cambridge: The MIT Press.]

Chomsky, N. 1999. *Derivation by Phase*. MIT Occasional Papers in Linguistics, 18.

Collins, C. 1997. *Local Economy*. Cambridge: The MIT Press.

Collins, C. 1999. *Eliminating Labels*. ms. Cornell University.

Drury, J. 1998a. *Root-First Derivations: Multiple Spell-Out, Atomic Merge, and the Coresidence Theory of Movement*. Doctoral Research Paper. University of Maryland, College Park.

Drury, J. 1998b. The Promise of Derivations: Atomic Merge and Multiple Spell-out. In: Abraham, W. (ed.) *Groninger Arbeiten zur Germanistischen Linguistik* 42: 61-108. Rijksuniversiteit Groningen: Germanistisch Institut.

Drury, J. 1999. Movement as Remerge and C-Command as Sub-Derivational Precedence. Paper presented at *the GLOW 1999 Workshops in Postdam: Technical Aspects of Movement*. March 29th-April 1st.

Epstein, S. D. 1999. Un-Principled Syntax and the Derivation of Syntactic Relations. In: S. Epstein & N. Hornstein (eds.) *Working Minimalism*, 317-345. Cambridge: The MIT Press.

Epstein, S. D., E. Groat, R. Kawashima & H. Kitahara. 1998. *A Derivational Approach to Syntactic Relations*. New York/Oxford: Oxford University Press.

Frank, R. & K. Vijay-Shanker. 1999. *Primitive C-Command*. ms. Baltimore: Johns Hopkins University; Newark: University of Delaware. [available at http://128.220.29.11/faculty/frank]

Gärtner, H-M. 1997. *Generalized Transformations and Beyond (reflections on minimalist syntax)*. Ph.D. dissertation. Frankfurt: Johan-Wolfgang-Goethe Universität.

Gärtner, H-M. 1998. Review of "The Copy Theory of Movement and Linearization of Chains in the Minimalist Program, by Jairo Nunes". *Glot International*, 3 (18): 15-19.

Guimarães, M. 1997. Unifying LCA & Prosodic Phrasing in the Minimalist Program. Talk given at the *Workshop on Theory of Grammar: Problems at PF and LF Interface Levels*. Campinas: Universidade Estadual de Campinas. August 5th-7th.

Guimarães, M. 1998. *Repensando a Interface Sintaxe-Fonologia a partir do Axioma de Correspondência Linear*. Master's Thesis. Campinas: Universidade Estadual de Campinas.

Guimarães, M. 1999a. *Deriving Prosodic Structure from Dynamic Top-Down Syntax*. ms. University of Maryland at College Park.

Guimarães, M. 1999b. *Phonological Cascades & Intonational Structure in Dynamic Top-Down Syntax*. ms. University of Maryland at College Park. [available at http://www.inform.umd.edu/EdRes/Colleges/ARHU/Depts/Linguistics/Events/StudentConference/1999.html]

Hale, K. & S. J. Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In: K. Hale & S. J. Keyser (eds.), *The View from Building 20: Essays in Honor of Sylvain Bromberger*, 53-110. Cambridge: The MIT Press.

Halle, M. & A. Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In: K. Hale & S. J. Keyser (eds.), *The View from Building 20: Essays in Honor of Sylvain Bromberger*, 111-176. Cambridge: The MIT Press.

Higginbotham, J. 1983. A Note on Phrase Markers. *Revue Quebecoise de Linguistique*, 13: 147-166.

Hornstein, N. 1999. Movement and Control. *Linguistic Inquiry*, 30: 69-96.

Hornstein, N. forthcoming. *Move!* Oxford: Blackwell.

Hornstein, N. & J. Uriagereka. 1999. Labels and Projections: A Note on the Syntax of Quantifiers. In: S. Aoshima, J. Drury & T. Neuvonen (eds.) *University of Maryland Working Papers in Linguistics*, 8: 249-270.

Inkelas, S. & D. Zec (eds.). 1990. *The Phonology-Syntax Connection.* Chicago: The University of Chicago Press.

Inkelas, S. & D. Zec. 1995. Syntax-Phonology Interface. In: J. Goldsmith (org.) *The Handbook of Phonological Theory.* Oxford: Blackwell.

Jackendoff, R. 1977. *X'-Syntax: A Study of Phrase Structure*. Cambridge: The MIT Press.

Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kitahara, H. 1995. Target α: Deducing Strict Cyclicity from Derivational Economy. *Linguistic Inquiry*, 26: 47-77.

Moro, A. 1997. Dynamic Antisymmetry: Movement as a Symmetry-Breaking Phenomenon. *Studia Linguistica,* 52: 50-76.

Muysken, P. 1982. Parametrizing the Notion 'Head'. *Journal of Linguistics Research*, 2: 57-75.

Nespor, M. & I. Vogel. 1982. Prosodic Domains of External Sandhi Rules. In: H. van der Hulst & N. Smith (eds.). *The Structure of Phonological Representation. Part I*, 225-255. Dordrecht: Foris.

Nespor, M. & I. Vogel. 1986. *Prosodic Phonology.* Foris: Dordrecht.

Nunes, J. 1995a. *Linearization of Chains at LF*. ms. College Park: University of Maryland.
Nunes, J. 1995b. *The Copy Theory of Movement and Linearization of Chains in the Minimalist Program*. Ph.D. dissertation. University of Maryland, College Park.
Nunes, J. 1999. *Linearization of Chains and Phonetic Realization of Chain Links*. In: S. D. Epstein & N. Hornstein (eds.) *Working Minimalism*, 217-249. Cambridge: The MIT Press.
Nunes, J. 2000a. *Erasing Erasure*. ms. Campinas: Universidade Estadual de Campinas.
Nunes, J. 2000b. *Sideward Movement*. ms. Campinas: Universidade Estadual de Campinas.
Nunes, J. 2000c. *Linearization of Chains and Sideward Movement*. ms. Campinas: Universidade Estadual de Campinas.
Nunes, J. & E. Thompson. 1998. Formal Appendix. in: J. Uriagereka. *Rhyme and Reason: An Introduction to Minimalist Syntax.* Cambridge: The MIT Press.
Partee, B, A. ter Meulen & R. Wall. 1993. *Mathematical Methods in Linguistics*. Boston, London: Kluwer-Dordrecht.
Russel, B. 1919. *Introduction to Mathematical Philosophy*. London: George Allen & Unwin. New York: The Macmillan Co.
Selkirk, E. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge: The MIT Press.
Selkirk, E. 1986. On Derived Domains in Sentence Phonology. *Phonology Yearbook*, 3: 371-405.
Speas, M. 1990. *Phrase Structure in Natural Language*. Dordrecht: Kluwer.
Stowell, T. 1981. *Origins of Phrase Structure*. Ph.D. dissertation. Massachusetts Institute of Technology.
Truckendbrodt, H. 1995. *Phonological Phrases: Their Relation to Syntax, Focus and Prominence*. Ph.D. dissertation. Massachusetts Institute of Technology.
Truckendbrodt, H. 1999. On the Relation Between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry*, 30: 219-255.
Uriagereka, J. 1998. *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge: The MIT Press.
Uriagereka, J. 1999. Multiple Spell-Out. In: S. D. Epstein & N. Hornstein (eds.) *Working Minimalism*, 251-282. Cambridge: The MIT Press.
Wall, R. 1972. *Introduction to Mathematical Linguistics*. Englewood Cliffs, NJ: Prentice-Hall.
Watanabe, A. 1995. Conceptual Basis of Cyclicity. *MIT Working Papers in Linguistics, Papers on Minimalist Syntax*, 27: 269-291.
Zadeh, L. 1987. *Fuzzy Sets and Applications: Selected Papers*. New York: Wiley.

Maximiliano Guimarães
University of Maryland
Department of Linguistics
1401 Marie Mount Hall
College Park, MD, 20742-7505
maxgm@wam.umd.edu