A set-theoretic approach to the Copy Theory of Movement

Carlos Muñoz Pérez
University of Buenos Aires

0. Abstract

In this paper, I consider some technical issues of Nunes' (1995, 2004) Copy Theory of Movement regarding (i) deletion of uninterpretable features, (ii) nondistinctiveness of copies and (iii) semantic interpretation of lower copies. In order to solve these problems, I propose a novel mechanism of chain formation based on relations of inclusion (e.g., subsets and supersets) between syntactic objects. This approach, I argue, captures in a very simple and minimalist fashion the "sameness" relation that holds between the links of a nontrivial chain without losing Nunes' empirical coverage regarding pronunciation of lower copies and multiple copy realization.

Keywords: syntactic movement, copy theory of movement, nondistinctiveness, set theory, late insertion.

1. Introduction: Copy Theory in a nutshell

Currently, the implementation of the Copy Theory of Movement (CT) proposed by Nunes (1995, 2004) is one of the most prominent views about the displacement of constituents in the minimalist framework. Unlike its predecessor, the trace based-theory, CT maintains that the same syntactic object (SO) may be present several times in the structure of a sentence: each instance of this SO is a copy that has been merged in order to check some uninterpretable feature on it or on a functional head.

For example, take a passive sentence as *John was kissed*. This kind of construction is normally assumed to have a derivational step like (1a) in which the SO *John* has ϕ-features (ϕ) that are required by T to check its own uninterpretable ϕ-features ($u\phi$). In order to satisfy this requirement, a copy of *John* is generated (1b) and merged in the [Spec, T] position (1c). Once this local Spec-Head relation has been obtained, T checks its ϕ-features.

As a reflex of this relation, the DP *John* checks its uninterpretable Case features (*uC*) with T by receiving nominative Case. It is said that both copies of *John* form a discontinuous representation CH which is referred as (nontrivial) chain: CH = (John, John).

(1) a. Matrix structure: $[_{TP}\ T_{u\phi}\ [_{VP}\ \text{kissed John}_{\phi,uC}]]$.

b. Copy: $[_{DP}\ \text{John}_{\phi,uC}]$.

c. Merge: $[_{TP}\ \text{John}_{\phi,C}\ [_{T'}\ \text{was}_{\phi}\ [_{VP}\ \text{kissed John}_{\phi,uC}]]]$.

What should be noticed about CT in the first place is that when a movement chain like CH = (John, John) is generated, one could expect that the phonological representation for a sentence like the one derived in (1) would be something like (2b), where the word *John* is pronounced twice. Nevertheless, this kind of pattern is, almost always, banned. There are other logical possibilities, like pronouncing only the most embedded copy (2c) or pronouncing none of the copies (2d), but these options also lead to ungrammatical representations most of the times. By far, the most extended pattern in human language is the one where a single copy is pronounced, the one which is in the higher structural position, the head of the movement chain (2a).

(2) a. John was kissed ~~John~~.

b. *John was kissed John.

c. *~~John~~ was kissed John.

d. *~~John~~ was kissed ~~John~~.

According to Nunes (2004), any version of the CT should explain why it is that only the head of the chain receives phonetic representation. His system derives this property by assuming that movement chains should meet some requirements in order to be interpreted at PF. He follows ideas from Kayne (1994), who claims that several syntactic properties (e.g., the so called axioms of the X-bar Theory) are not primitively syntactic, but inherited from the linear nature of PF representations. In particular, Kayne assumes that any linear ordering has three defining characteristics.

(3) *Properties of linear orderings (adapted from Kayne 1994:4)*

Given a linear ordering L containing three elements x, y, z, L is

a. Transitive: if x precedes y and y precedes z, then x precedes z.

b. Total: for all distinct x and y, either x precedes y or y precedes x.

c. Antisymmetric: if x precedes y, then y can't precede x.

Inasmuch as the only syntactic structural relation that meets these three requirements is asymmetric c-command, Kayne proposes that there exists a systematic and constant correlation between asymmetric c-command and a PF linear ordering. Such correspondence is specified in the so called *Linear Correspondence Axiom* (LCA).

(4) *Linear Correspondence Axiom (from Nunes 2004:15)*

Let X, Y be nonterminals and x, y terminals such that X dominates x and Y dominates y. Then if X asymmetrically c-commands Y, x precedes y.

If we assume the CT, then we should expect that copies obey the same kind of rules and grammatical conditions than any other SO. So, if PF linearizes a syntactic representation as (1c) according to the LCA in (4), then (the higher copy of) *John* would asymmetrically c-command *was* and, at the same time, *was* would asymmetrically c-command (the lower copy of) *John.* Given that the system interprets both copies of *John* as the same SO, this would lead to a paradoxical linearization instruction: *John* would precede *was* and *was* would precede *John.* Crucially, this kind of instruction is problematic with one of the basic defining properties of linear orderings: antisymmetry (3c). With the purpose of avoiding these scenarios, Nunes claims that it is necessary to erase some of the links in a nontrivial chain by using the mechanism in (5).

(5) *Chain Reduction (from Nunes 2004: 27)*

Delete the minimal number of constituents of a nontrivial chain CH that suffices for CH to be mapped into a linear order in accordance with the LCA.

Note that the LCA may only linearize a representation in which there's only one version of a SO, so (5) would apply until there remains just one link in every nontrivial chain.

Now, how does the system choose which link of the chain is going to be pronounced (and is going to survive *Chain Reduction*)? The old trace-based theory of movement stipulated that the head is the only SO in the movement chain with phonetic representation. Nunes (1995, 2004) derives the same result as a consequence of assuming that there is a post-syntactic operation that eliminates the remaining uninterpretable features from the syntactic representation before it reaches PF.

(6) *FF-Elimination (from Nunes 2004: 31)*

Given the sequence of pairs $\sigma = \langle (F, P)_1, (F, P)_2, \ldots, (F, P)_n \rangle$ such that σ is the output of Linearize, F is a set of formal features, and P is a set of phonological features, delete the minimal number of features of each set of formal features in order for σ to satisfy Full Interpretation at PF.

The highest link in a movement chain is always the one that is engaged in more checking relations. So, in order to reduce the number of applications of *FF-Elimination*, the optimal candidate to survive *Chain Reduction* is the head of the movement chain. Thereby, (almost always) the pronounced link in a chain is the head.

2. Three basic problems with this theory

This version of the Copy Theory of Movement presents at least three technical issues: (i) it seems that there is not a principled way for the unpronounced links of a movement chain to check their uninterpretable features; (ii) the only mechanism proposed at the moment in order to explain the "sameness" relation between the links in a movement chain violates *Inclusiveness*; (iii) even if the original advantage of Copy Theory was to explain in an elegant –yet simple– way some interpretative patterns (e.g., reconstruction), it requires the postulation of an extra mechanism (*i.e., Trace Conversion*) in order to predict the correct semantics for several constructions involving displacement of constituents.

Let us consider these problems one by one.

2.1. CHECKING/DELETION OF UNINTERPRETABLE FEATURES.

The operations *Chain Reduction* and *FF-Elimination* are applied in the PF side of the grammar only. Thus, an additional mechanism should be proposed to delete the uninterpretable features that, otherwise, may reach LF. This mechanism consists of two definitions.

(7) *Feature Uniformity Condition (from Nunes 2004: 70).*
Given a chain $CH = (\alpha_1, \ldots, \alpha_n)$, every α_i $(1 \leq i \leq n)$ must have the same set of features visible at LF.

(8) *Chain Uniformization (from Nunes 2004: 71).*
Delete the minimal number of features of a nontrivial chain CH in order for CH to satisfy the Feature Uniformity Condition.

The *Feature Uniformity Condition* of (7) establishes that all the links in a chain must possess the same features at LF. The operation *Chain Uniformization* of (8) implements (7). Aside from the ad-hoc nature of these definitions, it is necessary to point out that the mere existence of post-syntactic operations that act as "filters for uninterpretable features" before the syntactic representation reaches the interfaces implies that syntax may generate uninterpretable structures. This conception of (narrow) syntax goes against several minimalist assumptions, including the *Strong Minimalist Thesis* of (9), since it is predicted that (narrow) syntax cannot generate representations directly interpretable at the performance systems.

(9) *Strong Minimalist Thesis*
Syntax is an optimal solution to interface conditions.

Two main alternatives are considered in the literature in order to deal with the problem of uninterpretable features in the unpronounced copies. The first one is based on the assumption that the features of a movement chain are a unit (i.e., when a feature is checked in one copy, it's also checked in the remaining copies), a proposal made by Chomsky (1995). This idea, however, eliminates all possible asymmetries between the links of a chain (except, of course, for c-command). If we assume this, there would be no obvious way in which PF may know which of the copies must receive phonetic representation. Therefore, the pronunciation of the higher link in the movement chain would need to be stipulated.

The second alternative is due to Bošković (2007), which is also the approach taken by Nunes (2010). It is based on the idea that an element only may check its uninterpretable features if it probes for them. So, for example, Let us suppose that a DP needs to check some uninterpretable feature *u*X (10a), so a copy of this DP is generated and merged in the [Spec, X] position (10b). Once there, the DP probes X^0 and checks *u*X (10c). Then, the higher DP probes for the *u*X of its lower version, thus the original copy of the DP finally checks its uninterpretable features (10d).

(10) a. *A DP needs to check an uninterpretable feature uX*

$[_{XP} \quad X^0 \quad [\ldots \quad DP_{uX} \quad]]$

b. *A copy of the DP is merged in [Spec,X]*

$[_{XP} \quad DP_{uX} \quad [_{X'} \; X^0 \quad [\ldots \quad DP_{uX} \quad]]]$

c. *The DP probes X*

$[_{XP} \quad DP_{X} \quad [_{X'} \; X^0 \quad [\ldots \quad DP_{uX} \quad]]]$

AGREE

d. *DP_X probes DP_{uX} and checks its uninterpretable features*

$[_{XP} \quad DP_{X} \quad [_{X'} \; X_C \quad [\ldots \quad DP_{C} \quad]]]$

AGREE

However, for this proposal to work it would be necessary to stipulate that the syntactic representation must be transferred to PF in the precise moment when only the first copy had checked its features (e.g., in the derivational step illustrated in (10c)) and that the rest of the chain should do so in the covert component. This condition is necessary in order to maintain the feature interpretability-based asymmetry between copies in the movement chain that guarantees the application of *Chain Reduction* for the lower links only. Because of this strong assumption, adopting this system would be just as stipulative as postulating that the higher copy is the one that should always be pronounced.

2.2. NONDISTINCTIVENESS OF COPIES

As said before, the fact that some copies in a nontrivial chain should be erased follows from a PF requirement: when there is more than one instance of the same SO, it is impossible to apply the LCA in order to linearize the syntactic representation. This very idea assumes that there is a "sameness" relation between the links in a movement chain (i.e., they are all the same element in some relevant way). The problem is how to define this "sameness" relation. To make the point clear, Let us look again the syntactic representation of a sentence as *John was kissed.*

(11) $[_{\text{TP}}$ John$_{\varphi,C}$ was $[_{\text{vP}}$ kissed John$_{\varphi,uC}]]$.

This sentence has two copies of the SO *John,* but the lower one is erased by the LCA requirements already explained. So, it is assumed that both copies of *John* are the same SO. Nevertheless, these SO differ in a very important way: the higher version of *John* does not have uninterpretable features, whereas the lower one does. This is an important issue, because it is not possible to state that the relation of "sameness" between copies is one of strict identity. Nunes (2004:164-165) claims that the relation of "sameness" involved in the formation of chains is one of *Nondistinctiveness*. This notion allows identifying as "the same" two SO that may differ in the value of their features. Nondistinctiveness is an essential concept for Nunes' version of the CT not only because it motivates the application of Chain Reduction, but because it is considered one of the necessary conditions for *Chain Formation*.

(12) *Conditions on Form Chain (from Nunes 2004: 91)*

Two constituents α and β can form the nontrivial chain CH = (α; β) if

a. α is nondistinct from β;
b. α c-commands β;
c. there is at least one feature F of α such that F enters into a checking relation with a sublabel of the head of the projection with which α merges and for any such feature F of α, the corresponding feature F of β is accessible to the computational system; and
d. there is no constituent γ such that γ has a feature F' that is of the same type as the feature F of α, and γ is closer to α than β is.

As (12b) states, Nondistinctiveness is not only the property that makes all the links in a nontrivial chain "the same" for the LCA: it may also be understood as the notion that designates all those properties that a group of SO in a local relation should share in order to be considered a nontrivial chain.

The main problem posed by nondistinctiveness is how to implement such notion in a principled manner. There seems to be no way in which nondistinctiveness may be derived from the formal features of the copies, since Chain Reduction ignores differences between copies regarding Case, for example. So, it is necessary to assume a special mechanism in order to introduce nondistinctiveness in the syntactic system. For this, there are two alternatives in the relevant literature. The first one has to do with the classic idea of marking with an index all the elements in a chain, either because of a mechanism inherent to the operation *Select* (Chomsky 1995) or to the operation *Copy* (Bošković and Nunes 2007). However, this idea is problematic since introducing marks that are not present in the numeration is not allowed because of the *Inclusiveness Condition*.

(13) *Inclusiveness Condition (Chomsky 1995: 228)*

Any structure formed by the computation is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation, apart from rearrangements of lexical properties (in particular, no indices, bar levels in the sense of X-bar theory, etc.).

Let us remember that (13) is one of the original reasons in order to abandon the trace-based theory of movement: this theory is based on the idea that every application of *Move-α* introduces two non-lexical elements, a trace and its index, so it violates the Inclusiveness Condition. If CT is in the need to introduce indexes in order to justify Chain Formation, then it also violates the Inclusiveness Condition, so we lose one of the conceptual motivations to adopt CT.

Another possibility to derive nondistinctiveness is proposed by Nunes (2004: 165, note 15). It is based on the idea of keeping track of the history of application of the operation *Select* in a derivation. Thus, if a new token of a syntactic object appears in the structure in a derivational step D_{i+1}, it is possible to determine if it is a copy if the Numeration had not changed from the derivational step D_i. But, of course, it is necessary to remark that the actual implementation of this mechanism has not been carried out. It is possible to conjecture that this is because of the complexity that this kind of computation would imply.

2.3. Interpretation of lower copies

Despite the advantages that CT presents in order to provide an elegant explanation of the reconstruction phenomena (as demonstrated by Chomsky 1993), it is impossible to disregard the very important role that the trace-based theory of movement played in order to understand the semantics of syntactic displacement. Being treated as variables at LF, traces offered an elegant justification for the semantics of A'-movement: a SO A'-moves in order to generate a quantifier-variable dependency.

Conversely, it is not clear how CT alone may be used to explain the fact that lower copies of A'-movement seem to receive the interpretation of a variable. Thus, for example, the meaning of (14) and (15) is unexpected in terms of the CT.

(14) *Wh-Movement*

Which boy Mary visited ~~which boy~~?

'Which is the boy x, such that Mary visited the boy x'.

(15) *Quantifier Raising*

~~every boy~~ A girl talked to every boy.

'For every boy x, there is a girl who talked to the boy x'.

In order to predict the correct interpretation of sentences of this type, Fox (2003) proposes a mechanism called *Trace Conversion*.

(16) *Trace Conversion (from Fox 2003: 67)*

a. Variable Insertion: (Det) Pred → (Det [Pred Ly(y=x)]

b. Determiner Replacement: (Det) [Pred Ly(y=x)] → the [Pred Ly(y=x)]

The rules of (16a) and (16b) have the objective of forming a definite description with anaphoric value from the lower link of the movement chain, something similar to the kind of variable classically associated with traces.

Now, there's an obvious problem when we want to include *Trace Conversion* as part of CT: it is an extremely powerful rule (a transformation, in exactly the same sense of Chomsky 1957) that is not based on syntactic operations and that modifies, by stipulation, structural and lexical aspects of the semantic representation.

3. A set-theoretic approach to chain formation

The main objective of this paper is to advance a new conception of *Chain*. I hope to demonstrate that the problems presented so far (and some new others, I believe) can be elegantly solved if we think of chains as based on relations of inclusion between sets. In order to accept this idea, it is necessary to adopt three assumptions. The first assumption I am going to make is that SO are sets of features. This is a very usual idea: *Merge* is an operation which combines lexical items, sets of features, into phrases, sets of sets. Thus, all syntactic objects are sets with linguistic features as minimal elements.

Also, I am denying the existence of uninterpretable features. Here, I am not making an additional assumption, but I am abandoning one. Crucially, what I am going to disregard is, in Preminger's (2011: 18-19) terms, an approach to syntactic operations which implements its obligatoriness on the existence of *derivational time-bombs*, features that need to be eliminated through syntactic mechanisms in the course of the derivation in order to obtain a well-formed final representation. There is no empirical loss in abandoning this approach to syntactic operations: as observed by López (2007: 24), there is no evidence for the idea that some feature, if not checked, leads to a crash of the derivation at the interfaces. So, for a more concrete proposal, I am going to assume that *Agree* is an operation that is not motivated by the necessity of checking uninterpretable features, but for the need of valuing some feature F of a functional head H[1]. As in Chomsky's (2001) system, I am also going to assume that Agree (i) may trigger the movement (i.e., Copy + Merge) of the goal into the [Spec, H] position, and that (ii) H may assign a feature (e.g., Case) to the Goal if H is a non-defective head.

Finally, I am assuming that syntax operates with abstract morphemes, and that their phonological features are supplied lately at PF. For concreteness, I will adopt some

[1] In order to formalize the same idea, Preminger proposes an alternative operation to *Agree*, which he calls *Find(f)*.

(i) *Find(f) (from Preminger 2011: 94)*
Given an unvalued feature *f* on a head H^0, find an XP bearing a valued instance of *f*, and assign that value to H^0.

proposals of the *Distributed Morphology* framework (Halle & Marantz 1993, *inter alia*) and assume that the operation that introduces phonological exponents into syntactic terminals is called *Vocabulary Insertion,* a mechanism ruled by Halle's (1997) *Subset Principle*.

(17) *Subset Principle (from Halle 1997: 128)*

The phonological exponent of a Vocabulary item is inserted into a morpheme in the terminal string if the item matches all or a subset of the grammatical features specified in the terminal morpheme. Insertion does not take place if the Vocabulary item contains features not present in the morpheme. Where several Vocabulary items meet the conditions for insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen.

By taking these three assumptions it is possible to understand a movement chain as an inclusion relationship between cantorian subsets and supersets[2]. For example, Let us say that we have the SO α corresponding to the trivial chain CH = ($\alpha_{\{...\}}$), in which {…} represents the set of features associated with α. Then, α is probed by a higher head in the syntactic structure (let us say Y) and a copy of it is merged in [Spec,Y]. Let us suppose that Y is a non-defective head and that, by virtue of Agree, Y assigns some feature to α (the Y feature), so the chain becomes CH = ($\alpha_{\{Y, ...\}}$, $\alpha_{\{...\}}$). As may be noticed, there is a natural relation between the two elements in the chain: every member of the set $\alpha_{\{...\}}$ is also a member of the set $\alpha_{\{Y, ...\}}$. So, we can say that $\alpha_{\{Y, ...\}}$ is a superset of $\alpha_{\{...\}}$ ($\alpha_{\{Y, ...\}} \supseteq \alpha_{\{...\}}$), or, conversely, that $\alpha_{\{...\}}$ is a subset of $\alpha_{\{Y, ...\}}$ ($\alpha_{\{...\}} \subseteq \alpha_{\{Y, ...\}}$). Interestingly, this inclusion relation would persist for every new copy of α: it would be a superset of each of the lower copies[3].

---

[2] A set S_1 is a subset of set S_2 ($S_1 \subseteq S_2$) if every member of S_1 is also a member of S_2. The notion of *subset* is different from the notion of *Proper Subset*: a set S_1 is a proper subset of set S_2 ($S_1 \subset S_2$) if all elements of S_1 are elements of S_2, but not all elements of S_2 are elements of S_1.

[3] Even if the head attracting α is defective (i.e., it doesn't assign any feature to α), the new copy is going to be a superset of α, since any set S is a subset of S.

A more concrete example is found in (13). For the sake of the argument, Let us assume that Y is a Case assigner and X is some category of the left periphery which assigns a discourse-oriented feature (or, in Speas and Tenny (2003) terms, a Pragmatic-role). Let us suppose that a DP is merged as a complement of Z, and is then probed by the higher non-defective head Y. A copy of this DP is generated and merged in [Spec, Y]. Because of the Agree relation, the DP receives a Case feature Y. Next, the DP is once again probed, this time by the left peripheral category X. Since X is a non-defective head, a new copy of the DP is merged at [Spec, X] and receives a discourse-oriented feature X. The relevant representation is the one in (18).

(18) $[_{XP}\ \mathbf{DP_{\{X, Y, \ldots\}}}\ [_{X'}\ X\ [_{YP}\ \mathbf{DP_{\{Y, \ldots\}}}\ [_{Y'}\ Y\ [_{ZP}\ Z\ \mathbf{DP_{\{\ldots\}}}]]]]]$

As said before, the three copies of the DP are related in an inclusion relation: $DP_{\{\ldots\}}$ is a subset of $DP_{\{Y, \ldots\}}$ ($DP_{\{\ldots\}} \subseteq DP_{\{Y, \ldots\}}$) and $DP_{\{Y, \ldots\}}$ is a subset of $DP_{\{X, Y, \ldots\}}$ ($DP_{\{Y, \ldots\}} \subseteq DP_{\{X, Y, \ldots\}}$)[4]. It is possible to capitalize this systematic relation in order to redefine the conditions on the formation of chains.

(19) *Conditions on Chain Formation*

Two constituents α and β form a chain if

a. α is a superset of β;

b. α c-commands β;

c. there is no constituent γ between α and β such that (i) β is a subset of γ and (ii) γ is not a subset of α.

The conditions stated in (19) are very similar to those in (12). The main difference between them is that in this second case we are replacing the notion of nondistinctiveness with a relation of inclusion between sets. This change is well-motivated on minimalist grounds. If we take for granted that SOs are sets, then we are compelled to understand any relation between SOs as a relation between sets. Normally, the first kind of relation that comes to

[4] There is a third subset relation which is irrelevant at the moment: the $DP_{\{\ldots\}}$ is a subset of $DP_{\{X, Y, \ldots\}}$ ($DP_{\{Y, \ldots\}} \subseteq DP_{\{X, Y, \ldots\}}$).

mind in order to underlay “sameness” is strict identity, maybe because identity is usually considered a primitive type of relation. In this line of thought, if we wanted to use identity to define “sameness” we would have to say that for any SO α that conforms the set S_α, and for any SO β that conforms the set S_β, α and β would be the same SO if there is strict identity between S_α and S_β ($S_\alpha = S_\beta$). However, this characterization of “sameness” is too strong and fails to capture the inherent changes that new copies suffer when they engage new syntactic relations, as Nunes (1995: 87) originally observed. Here, we have two options: (i) postulate some new and more complex kind of “sameness”, or (ii) seek for a simpler relationship between sets that may underlay chain formation. Nondistinctiveness is a consequence of option (i); (19b) is a consequence of option (ii). If we take into consideration set theory, identity is a notion derived from the more primitive inclusion relation: two sets S_α and S_β are identical if S_α is a subset of S_β and, at the same time, S_β is a subset of S_α. (if $A \subseteq B$ and $B \subseteq A$, then $A = B$). So, (19b) is based on a relation between sets that is simpler than identity: inclusion.

Another way to think about this proposal is to consider what properties should share a group of SOs in a local configuration to form a nontrivial chain. Nondistinctiveness is a notion that needs to introduce indexes in order to indicate what SOs would be members of the same chain. So, an extrinsic index *i*, for instance, is introduced in the derivation to signal all those elements that will conform the nontrivial chain CH = (X^i, …, Z^i). As said, this violates the Inclusiveness Condition of (13). Instead, the set-theoretic approach to chain formation advanced in (19) makes possible to understand that the minimal properties that a group of SOs need to share to form a chain are those that characterize the tail (the lower link) of the chain. In other words, since the tail is a subset of all the remaining copies in the chain, it may be understood as something like the “common denominator” of the chain.

Coming back to (19), if we apply these conditions to the representation in (18) they predict the existence of the chains $CH_1 = (DP_{\{X, Y, \ldots\}}, DP_{\{Y, \ldots\}})$ and $CH_2 = (DP_{\{Y, \ldots\}}, DP_{\{\ldots\}})$, which may be understood as a single chain $CH = (DP_{\{X, Y, \ldots\}}, DP_{\{Y, \ldots\}}, DP_{\{\ldots\}})$. Nunes (2004: 17) remarks that for any version of the copy theory it is necessary to address two main questions: (i) why is that a movement chain, in general, cannot have all of its links phonetically realized?, and (ii) why is it that only the higher copy of the moment chain, in general, receives phonetic representation? Both questions may be synthesized in one: why is it the case that from all the possible phonetic representations for (18) in (20) only (20a) follows the general pattern in human language?

(20)
a. $[_{XP}\ DP_{\{X, Y, \ldots\}} \ldots X \ldots [_{YP}\ \cancel{DP_{\{Y, \ldots\}}} \ldots Y \ldots [_{ZP}\ Z \ldots \cancel{DP_{\{\ldots\}}} \ldots]]]$
b. $[_{XP}\ DP_{\{X, Y, \ldots\}} \ldots X \ldots [_{YP}\ DP_{\{Y, \ldots\}} \ldots Y \ldots [_{ZP}\ Z \ldots DP_{\{\ldots\}} \ldots]]]$
c. $[_{XP}\ \cancel{DP_{\{X, Y, \ldots\}}} \ldots X \ldots [_{YP}\ \cancel{DP_{\{Y, \ldots\}}} \ldots Y \ldots [_{ZP}\ Z \ldots \cancel{DP_{\{\ldots\}}} \ldots]]]$
d. $[_{XP}\ DP_{\{X, Y, \ldots\}} \ldots X \ldots [_{YP}\ \cancel{DP_{\{Y, \ldots\}}} \ldots Y \ldots [_{ZP}\ Z \ldots DP_{\{\ldots\}} \ldots]]]$
e. $[_{XP}\ DP_{\{X, Y, \ldots\}} \ldots X \ldots [_{YP}\ DP_{\{Y, \ldots\}} \ldots Y \ldots [_{ZP}\ Z \ldots \cancel{DP_{\{\ldots\}}} \ldots]]]$
f. $[_{XP}\ \cancel{DP_{\{X, Y, \ldots\}}} \ldots X \ldots [_{YP}\ DP_{\{Y, \ldots\}} \ldots Y \ldots [_{ZP}\ Z \ldots \cancel{DP_{\{\ldots\}}} \ldots]]]$
g. $[_{XP}\ \cancel{DP_{\{X, Y, \ldots\}}} \ldots X \ldots [_{YP}\ \cancel{DP_{\{Y, \ldots\}}} \ldots Y \ldots [_{ZP}\ Z \ldots DP_{\{\ldots\}} \ldots]]]$
h. $[_{XP}\ \cancel{DP_{\{X, Y, \ldots\}}} \ldots X \ldots [_{YP}\ DP_{\{Y, \ldots\}} \ldots Y \ldots [_{ZP}\ Z \ldots DP_{\{\ldots\}} \ldots]]]$

Trace-based theory of movement stipulated the pattern in (20a): the head of the movement chain receives phonetic representation and the remaining links are null traces. Nunes' version of the CT derives the pattern in (20a) from (i) nondistinctiveness, (ii) the LCA and (iii) economy considerations on the deletion of uninterpretable features. Here, I am not assuming nondistinctiveness nor uninterpretable features, so it remains to be explained why (20a) is the most common pattern of phonetic realization of chains in human language.

As stated previously, the system that I am trying to sketch assumes late insertion of phonological matrices in the syntactic representation. That is: syntax operates with abstract morphemes and functional categories without phonological information. The phonological exponents of the syntactic terminals are provided at PF. This is a very important difference with regard to Nunes' (1995, 2004) system, which assumes that the lexical items used in the syntactic derivation carry their phonological features. Thereby, non-pronunciation of copies is for Nunes a matter of copy deletion. For us, from a late insertion point of view, non-pronunciation of copies is non-insertion of phonological matrices at PF.

The most direct consequence of this difference between both systems is reflected in terms of economy: in Nunes' system, the more economical PF representation is the one where no deletion of copies (i.e., Chain Reduction) applies, so the optimal representation for this kind of theory is, regardless of the LCA, something like (20b). Likewise, in a late insertion-based model, the optimal PF representation would be the one where *Vocabulary Insertion* does not apply at all (20c), an output that should be ruled out since at least one copy needs to be pronounced in order to satisfy the conditions of recoverability of information[5]. Therefore, this economy consideration on the application of Vocabulary Insertion rules out all the representations in (20) that phonetically realize more than one copy.

For now, Let us suppose that this answers one of Nunes' questions (i.e., why is it that only the higher copy of the moment chain receives phonetic representation?). However, it is important to take into account that the final form of this system will not rely completely on economy considerations to answer this concern. In particular, and as it is going to be seen later, this system predicts that chains may phonetically realize only one of its links, so when we have multiple copies of an element we have multiple PF chains too.

Moving on, the manner to answer Nunes' second question (i.e., why is it that only the higher copy of the moment chain receives phonetic representation?) also involves the functioning of *Vocabulary Insertion*. Late insertion models explored in the tradition of Distributed Morphology (e.g., Halle & Marantz 1993, Embick & Noyer 2004) assume that there is *competition* between those vocabulary items that meet the conditions for being introduced in a syntactic terminal: since only one phonological exponent may occupy a node, the one which possess more features in common with that node would be selected. According to the *Subset Principle* of (17), the vocabulary item that wins the competition may be *underspecified* regarding the position that it is going to occupy, but it cannot have more features than the ones associated with that syntactic terminal. Given these

[5] A more specific proposal may be found in Saab (2009: 425), who proposes a *Condition on the Pronunciation of Nontrivial Chains* that establishes that at least one link in a nontrivial chain should receive phonetic representation at PF. This condition, he discusses, may be derived from *Recoverability Condition on Deletions* (Chomsky 1965).

assumptions, this system explains in a very natural way the existence of syncretism: a fully-specified terminal node may be occupied by phonological exponent that does not need to be fully specified, thus a single vocabulary item may appear in several syntactic contexts.

Note that the logic of the Subset Principle only makes sense if we assume that syntactic terminal nodes are fully specified with features. This is a crucial assumption if we accept the chain formation system that I am presenting here, because it implies that the higher link in a nontrivial chain will always be the optimal candidate regarding the application of *Vocabulary Insertion*. Think of what would happen if *Vocabulary Insertion* applies systematically on the lower link of movement chains. If that were the case, the phonological exponents would compete for being inserted in a very poorly specified position. It is possible to make two predictions regarding this hypothetic scenario: (i) the presence of syncretic elements in linguistic representations would be the general case, with all the issues that this would imply (e.g., systematic structural and lexical ambiguity, vacuity of several syntactic mechanisms, etc.); (ii) there would be a lot of cases in which it would be impossible to introduce any phonological exponent in the syntactic terminals, since there would be no sufficiently underspecified exponent for several of them, and consequently there would be a lot of chains without phonetic representation[6]. Thus, it may be assumed that the reason to pronounce the higher copy in a movement chain is to avoid this kind of situation: the higher copies are always the ones with the higher specification in terms of features, so they become the best candidates in order to disambiguate the lexical competition when Vocabulary Insertion applies[7]. For concreteness, I will assume that, as a

[6] See footnote 5 for a brief discussion on a possibility like this.

[7] The present theory results very attractive from a broader point of view than the syntactic one, a perspective that emphasizes the importance of understanding language as a phenomenon of human cognition. As Williams (2004: 364) claims, there exists a very huge similarity between the principles that determine the competition among phonological representations (e.g., Subset Principle, Panini Principle (XXXX), Elsewhere Principle (XXXX)) and the gricean maxim that establishes that the more informative forms of expression are the ones that should be used: both mechanisms, as the one in (21), are based in a criterion of "choose the more complete option". The possibility of explaining syntactic, phonological and pragmatic mechanisms based on the same

general rule, the copy that receives phonetic representation is the one that contains all the features of the remaining copies, according with the following principle:

(21) *Pronounce the Superset*[8]

Given a Chain CH, pronounce the link which is the superset of all the remaining links in CH, unless this causes a violation of some PF condition.

If we apply (21) to the structure in (18), it predicts that the pattern (20a), the one where $DP_{\{X, Y\}}$ receives phonetic representation, is the correct one. Thereby, it has been shown that the present approach to chain formation successfully answers the second question posed by Nunes (2004).

It still remains to be explained why the lower copies in an A'-movement chain do not have exactly the same semantic interpretation as the higher copies, but they seem to be headed by a definite determiner instead.

(22) *Wh-Movement*

Which boy Mary visited ~~which boy~~?

*'Which is the boy x, such that Mary visited **the** boy x'.*

(23) *Quantifier Raising*

~~every boy~~ A girl talked to every boy.

*'For every boy x, there is a girl who talked to **the** boy x'.*

Under the system that has been advanced, it is possible to sketch an analysis of this phenomenon without resorting to complex LF operations as *Trace Conversion*. The basic

---

general principle opens a wide range of investigative opportunities for a framework that assumes the Strong Minimalist Thesis.

[8] Another way to understand this principle is as the counterpart of the competition between vocabulary items: copies in a movement chain compete for phonological representation; the one with more features is chosen for application of Vocabulary Insertion.

idea is that a wh-determiner as *which* originally has the "elementary" meaning of a definite determiner, and its quantificational value comes from a feature assignment from the interrogative complementizer. In other words, in a system that "enriches" the SOs according to their Agree dependencies, the semantics of displaced interrogative particles may be understood compositionally. Thus, a wh-determiner as *which* that selects a NP as complement forms a SO that is interpreted as a definite expression at LF, but a higher copy of this SO may receive a feature Q that enables a quantificational reading.

(24) $[_{CP}$ **$[_{DP}$ Which boy$]_{\{Q, D, \ldots\}}$** $[_{C'}$ C_Q $[_{TP}$ Mary $[_{T'}$ T $[_{VP}$ visited **$[_{DP}$ Which boy$]_{\{D, \ldots\}}$** $]]]$
'Which is the boy x, such that Mary visited the boy x'.

The present approach to chain formation explains elegantly this aspect of the semantics of syntactic movement. Further research will tell if this system may be extended for the better understanding of some other troubling phenomena concerning interpretation of chains at LF.

4. Phonetic realization of lower and multiple copies

As the reader may have noticed, the main difference between this system and the one proposed by Nunes (1995, 2004) lies in their assumptions. Nunes' system assumes nondistinctiveness and uninterpretable features, both notions that the chain formation system advanced in this paper does not adopt. Now, an additional theoretical concern is in order: it is necessary to demonstrate that there is no empirical loss in this theoretical choice.

One of the greatest virtues of Nunes' system is that it explains in a very elegant way two kinds of data that the trace-based theory of movement cannot: (i) cases where the link that receives phonetic representation is not the head of the movement chain, and (ii) cases where there is more than one link of the same chain phonetically realized. The first type of phenomenon may be exemplified with the patterns of multiple wh-fronting in Serbo-Croatian. In this language, if a sentence has two wh-phrases, both should move to the left periphery (25), This is the general case, unless (i) both wh-phrases were exactly the same interrogative pronoun and (ii) they were in adjacent positions (26a).

(25) *Serbo-Croatian (from Bošković 2002: 355)*

a. Ko šta kupuje?

Who what buys

b. *Ko kupuje šta?

who buys what

'Who buys what?'

(26) *Serbo-Croatian (from Bošković 1999, 2002, apud Nunes 2004: 35)*

a. *Šta šta uslovljava?

What what conditions

b. Šta uslovljava šta?

What conditions what

'What conditions what?'

c. Šta neprestano šta uslovljava?

what constantly what conditions

d. *?Šta neprestano uslovljava šta?

what constantly conditions what

'What constantly conditions what?'

According to Bošković (2002), this pattern may be easily explained if we assume that (26a) and (26b) share the same syntactic representation, but (26a) violates a morphological restriction that bans the adjacency of identical wh-elements in the PF representation. (26b), on the other hand, respects this condition by pronouncing the lower copy of the interrogative pronoun *šta*.

(27) [šta ~~šta~~^{i} uslovljava šta^{i}]

Nunes (2004) discusses several cases of pronunciation of low copies and concludes that every time that this happens it is because there is a PF restriction that would be violated if the head of the chain receives phonological representation. So, the pronunciation of lower copies may be conceived as a strategy in order to avoid the derivation to crash at PF.

This analysis may be directly implemented in the chain formation system presented in the previous section. The only difference regarding Nunes' system is the technical implementation of the rescuing mechanism: in a system that explains non-pronunciation as copy erasure (Chain Reduction) it is necessary to postulate an obligatory deletion on the copies that will infringe a PF condition if pronounced (i.e., if the pronunciation of SO X causes a PF violation, then apply Chain Reduction on it); on the other hand, in a system that explains non-pronunciation as non-insertion of phonological exponents it is necessary to ban the application of Vocabulary Insertion on certain syntactic terminals if their phonetic realization will cause a crash on the derivation (i.e., if the pronunciation of SO X causes a PF violation, then Vocabulary Insertion should not apply to the terminal node corresponding to X). In order to formalize and generalize the latter mechanism, I propose the notion of *Non-Insertion Domain*.

(28) *Non-Insertion Domain (NID)*

A NID is a structural segment in which it is impossible to apply *Vocabulary Insertion* because of some language specific prosodic or phonological constraint.

Note that "NID" is a covert term that allows explaining a very dynamical and idiosyncratic type of constraint on Vocabulary Insertion[9]. In this particular case, since the restriction on the pronunciation of adjacent interrogative pronouns in Serbo-Croatian seems to be restricted to homophones, it follows that the morpho-phonological context that triggers the NID arises during the process of insertion of phonological matrices. Thus, given two instances of the interrogative pronoun *šta* in the left periphery, one corresponding to the subject chain (*šta*$_{SUBJ}$) and other to the object chain (*šta*$_{OBJ}$), a NID arises right at the moment after introducing the phonological features of *šta*$_{SUBJ}$ and before introducing those of *šta*$_{OBJ}$

[9] The postulation of the NID notion may seem a bit gimmicky, since Nunes, for example, doesn't postulate any cover term in order to designate all the instances of "PF reparation by pronunciation of lower copies". Nevertheless, the reason to introduce this concept is that I believe it may be helpful to explain some other types of "silences" as null subjects (Muñoz Perez in preparation).

as a restriction on the iterative insertion of the same phonological matrix. Since the higher copy of the object chain could not be pronounced, Vocabulary Insertion applies on the tail.

(29) [Šta$_{SUBJ}$ [$_{NID}$ ~~šta$_{OBJ}$~~] uslovljava šta$_{OBJ}$]

Thereby, it has been shown that the set-theoretic approach to chain formation is compatible with Bošković's (2002) analysis of multiple wh-fronting phenomena and, more generally, with the idea defended by Nunes (2004) that the lower copies in a nontrivial chain may be pronounced if the higher one causes the derivation to crash due a PF violation.

The second kind of phenomenon that the trace-based theory of movement fails to explain and Nunes' version of CT manages to account for is pronunciation of multiple copies. This is a well-known phenomenon in certain dialects of German, for example: the sentence in (30) shows a displaced interrogative pronoun *wen* that is pronounced twice.

(30) *German (from McDaniel 1986, apud Nunes 2004: 38)*

Wen	glaubt	Hans	*wen*	Jakob	gesehen	hat?
Whom	thinks	Hans	whom	Jakob	seen	has

'Who does Hans think Jakob saw?'

Since the second *wen* appears in a position where it is supposed to be a copy of the higher *wen* generated by cyclic movement, it is tempting to take both as phonetic realizations of two copies of the same SO. If this is really the case, Nunes (2004) observes that this kind of phenomenon shows two main properties: (i) the pronunciation of low copies is restricted to intermediate ones, so the first merged and original lower copy is never pronounced; and (ii) the phenomenon is restricted to simplex morphological items (i.e., there is no multiple realization of full phrases).

In order to explain this phenomenon we could postulate two kinds of account: (i) assuming that in certain cases it is possible to realize more than one link of a nontrivial chain; and (ii) assuming that there are two (or more) PF chains and each one of them follows the general pattern of pronunciation (i.e., each one realizes its higher copy). The option explored by Nunes (2004) is the first one. Now, I will show how the present system derives an analysis of multiple copy realization in line with the second option.

Let us remember that Nunes motivates the application of Chain Reduction by the PF requirement of linearizing the syntactic structure according to the LCA: since the presence of more than one copy violates the antisymmetry condition, all but one link in a nontrivial chain should be erased. Now, Nunes assumes together with Chomsky (1995) that the LCA does not apply word-internally (i.e., the LCA is blind to the internal structure of syntactic terminals), so if a link of a nontrivial chain is phonetically realized as part of a bigger word, it would be invisible to the LCA and, as a consequence, to Chain Reduction. So, in Nunes' system it is possible to pronounce more than one copy of the same nontrivial chain if one of them is reanalyzed at PF as part of a word. Saab (2009) postulates this idea as a theoretical generalization.

(31) *Nunes Generalization (from Saab 2009: 407)*
The phonetic realization of multiple copies is possible only if at least one copy has been reanalyzed morphologically.

The technical implementation of this idea in Nunes' system goes as follows. It is assumed that successive-cyclic movement of interrogative pronouns may proceed by adjunction to intermediate complementizers, so a sisterhood relation is obtained between subordinate C^0 and the wh-pronoun *wen* [10]. Thus, for the case of the German example in (30), the relevant

[10] This assumption is, I think, unnecessary. If we take some aspects of Matushansky's (2006) system, it is possible to derive the sisterhood relation between the wh-word and the subordinate C as a case of head movement in Matushansky's terms. Since the SO that moves to the specifier position of the subordinate complementizer is a morphological unit, an X^0, it is possible to postulate an application of *morphological merger* (Halle & Marantz 1993; or *Lowering* in the more precise terminology of Embick & Noyer 2001) in order to obtain the representation in (32). I will maintain the adjunction approach to cyclic movement for the rest of the paper, but for purely expository reasons.

syntactic structure is represented in (32). Here it may be observed that one of the copies of *wen* and subordinate C^0 are sister nodes.

(32) [$_{CP}$ Wen^i … [$_{CP}$ **[$_{C°}$ wen^i C^0]** [$_{TP}$ … wen^i …]]]

As said, Nunes (2004) suggests that there is a morphological reanalysis that allows to interpret the constituent [$_{C°}$ wen C^0] as a morphological word. This idea is implemented through the operation *Fusion*.

(33) *Fusion (from Halle & Marantz 1993: 116)*
Fusion takes two terminal nodes that are sisters under a single category node and fuses them into a single terminal node.

Thus, Fusion takes the interrogative pronoun *wen* and the subordinate complementizer and fuses them in a single terminal node # wen + C #, where the sign # indicates the boundaries of this new single element.

(34) [$_{CP}$ Wen^i … [$_{CP}$ #**wen^i + C^0**# [$_{TP}$ … wen^i …]]]

Since the LCA cannot access the segment delimited by #, Chain Reduction does not apply to the copy of *wen* that's inside it. So, multiple copies of the interrogative pronoun *wen* appear in the phonological representation of the sentence. Note that the properties of this phenomenon noted by Nunes are both elegantly explained: realization of multiple copies is restricted to simplex morphological elements because it depends on the application of the morphological operation Fusion; and it is limited to intermediate copies because the syntactic configuration that's necessary to apply Fusion, sisterhood, relies on a previous moment operation.

A similar analysis may be implemented in a set-theoretic approach to chain formation. Let us sketch it step by step[11]. In the first place, the original version of the interrogative pronoun *wen* is merged in its thematic position, selected by a V. As a pronoun, it carries a D feature and φ-features and, due to an Agree relation with v, also an accusative feature (*Acc*).

(35) $[_{vP} \ldots [_{VP}\ V\ \mathbf{wen}_{\{Acc,\ D,\ \varphi,\}}]]$

The derivation continues until the CP layer, where a copy of *wen* is adjoined to the subordinate C as a cyclic stop in its way to the specifier position of the matrix C. Since the subordinate C is a defective head, there's no feature assignation to *wen.*

(36) $[_{CP}\ [_{C°}\ C^0\ \ \mathbf{wen}_{\{Acc,\ D,\ \varphi,\}}]\ [TP \ldots [_{vP} \ldots [_{VP}\ V\ \mathbf{wen}_{\{Acc,\ D,\ \varphi,\}}]]]]$

Finally, the syntactic derivation ends when a copy of *wen* is attracted to the specifier position of the matrix interrogative C. Since this time C is a non-defective head, it assigns a Q (quantificational) feature to *wen*, which provides its interrogative meaning to the pronoun. The resulting representation that reaches the morphological component is (37).

(37) $[_{CP}\ \mathbf{wen}_{\{Q,\ Acc,\ D,\ \varphi,\}}\ [_{C'}\ C \ldots [_{CP}\ [_{C°}\ C^0\ \ \mathbf{wen}_{\{Acc,\ D,\ \varphi,\}}]\ [TP \ldots [_{VP}\ V\ \mathbf{wen}_{\{Acc,\ D,\ \varphi,\}}]]]]]$

Following Nunes (2004), I will assume that the phonetic realization of the chain $CH = (wen_{\{Q,\ Acc,\ D,\ \varphi,\}}, wen_{\{Acc,\ D,\ \varphi,\}}, wen_{\{Acc,\ D,\ \varphi,\}})$ is affected by a morphological reanalysis, consisting on the application of Fusion. However, for purely operative reasons I will adopt a slightly different definition of this operation.

(38) *Fusion (from Embick 2010: 78)*

$[_X\ \alpha]^\frown[_Y\ \beta] \rightarrow [_{X/Y}\ \alpha,\beta]$

where α and β are features of X and Y.

[11] I represent only the relevant steps and operations in order to explain the analysis. Thus, it is possible to postulate a more elaborate derivation with more intermediate copies of *wen*.

Here, Fusion is defined as some kind of union between sets. So, if we follow Nunes and apply Fusion to the intermediate copy of the interrogative pronoun *wen* and C, we will combine the sets {Acc, D, φ}, corresponding to *wen,* and {C}, which is at least the set of features corresponding to C, so we will obtain the set {C, Acc, D, φ}. Thus, the resulting representation after applying Fusion is (39)

(39) $[_{CP}$ **wen**$_{\{Q, Acc, D, \varphi,\}}$ $[_{C'}$ C …$[_{CP}$ **#C + wen#**$_{\{C, Acc, D, \varphi,\}}$ [TP … $[_{VP}$ V **wen**$_{\{Acc, D, \varphi,\}}$]]]]]

According to the conditions on chain formation stated in (19), the representation in (39) has two chains: a trivial chain CH_1 = (wen$_{\{Q, Acc, D, \varphi,\}}$) and a nontrivial chain CH_2 =(#C + wen#$_{\{C, Acc, D, \varphi,\}}$, wen$_{\{Acc, D, \varphi,\}}$). So, if the application of the principle in (21) proceeds normally, then the syntactic pattern predicted by this system is the expected one.

As said, the main difference between Nunes' system and the present in regard to explaining multiple copy pronunciation is a technical one: in the approach taken by Nunes, phonetic realization of multiple copies of the same SO involves a chain that didn't require to delete some of its links, because the LCA could not accede to them; in my approach, morphological reanalysis of an intermediate copy implies the generation of two chains at PF (or, in a more descriptive term, the fracture of a chain into two minor chains), being the higher copies of both chains pronounced.

Note that in the set-theoretic approach, even when more than one copy of the same SO is pronounced, only one copy per chain is phonetically realized. Earlier, I answered one of Nunes' questions about the phonetic realization of chains (i.e. why is it that only the higher copy of the movement chain receives phonetic representation?) based on the idea that Vocabulary Insertion is applied economically (i.e., applying Vocabulary Insertion once is cheaper than applying it twice or more times). At that moment I claimed that this explanation, although possible, would not be of particular interest for this system. Now I will explain why I said that. Since this system never realizes phonetically two links of the same chain, an account for that property based on a criterion of economy is too weak. I think that a better explanation may be stated in terms of properties of Vocabulary Insertion. Taking this into account, I will assume that Vocabulary Insertion is an operation that applies on (trivial or nontrivial) chains. In other words, we may define Vocabulary Insertion as a complex mechanism that (i) takes a chain as input, (ii) selects an optimal link in terms of its features according to (21), and (iii) inserts phonological exponents in its terminal nodes.

5. Conclusion

I have tried to sketch a new proposal about how chain formation may be conceived in a minimalist framework. The main idea is very simple: if SOs are sets, then a chain, a collection of SOs related by some common property, should be defined by a set-theoretic relation. I claimed that the relevant relation is *inclusion*, so I have tried to show how the notions of subset and superset may be used in order to predict the general pattern of pronunciation of nontrivial chains. Finally, I argued that there is no empirical loss in assuming this kind of system for the Copy Theory of Movement.

References

Bošković, Ž. 2002. "On multiple wh-fronting". *Linguistic Inquiry* 33, 351–383.

Bošković, Ž. 2007. "On the Locality and Motivation of Move and Agree: An Even More Minimal Theory". *Linguistic Inquiry* 38: 589-644.

Chomsky, N. 1993. "A Minimalist Program for Linguistic Theory". Kenneth Hale & Samuel J. Keyser (eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bronberger.* Cambridge, MA: MIT Press.

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

Chomsky, N. 2001. "Derivation by Phase". Michael Kenstowicz (ed.), *Ken Hale. A Life in Language*. Cambridge, Mass., 1-52.

Embick, D. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Embick, D. & Noyer, R. 2001. "Movement Operations after Syntax". *Linguistic Inquiry* 32 (4): 555-595.

Embick, D. & Noyer, R. 2004. "Distributed Morphology and the Syntax–Morphology Interface". Gillian Ramchand & Reiss, Charles (eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford: OUP. 289-324.

Fox, D. 2003. "On Logical Form". In R. Hendrick (ed), *Minimalist Syntax,* Oxford: Blackwell Publishers.

Halle, M. 1997. "Distributed Morphology: Impoverishment and Fission". *MIT Working Papers in Linguistics* 30: 425-449.

Halle, M. & Marantz, A. 1994. "Some Key Features of Distributed Morphology". Andrew Carnie, Heidi. Harley & Tony Bures (eds.), *Papers on Phonology and Morphology*. *MITWPL* 21: 275-288.

Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.

López, L. 2007. Locality and the architecture of syntactic dependencies. New York: Palgrave Mcmillian.

Matushansky, O. 2006. "Head Movement in Linguistic Theory". *Linguistic Inquiry* 37(1): 69-109.

McDaniel, D. 1986. *Conditions on wh-chains*. Doctoral dissertation, City University of New York.

Nunes, J. 1995. *The copy theory of movement and linearization of chains in the Minimalist Program*. Doctoral dissertation, University of Maryland, College Park.

Nunes, J. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press

Preminger, O. (2011). *Agreement as a fallible operation*. Doctoral dissertation, MIT.

Saab, A. (2009). *Hacia una Teoría de la Identidad Parcial en la Elipsis.* Doctoral dissertation, University of Buenos Aires.

Speas, P. & Tenny, C. 2003. "Configurational Properties of Point of View Roles". Anna Maria DiSciullo (ed.), *Asymmetry in Grammar*. Amsterdam: John Benjamins.

Williams, E. (2004). "Dumping Lexicalism". Gillian Ramchand & Reiss, Charles (eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford: OUP. 353-382.