

## An inclusion-based mechanism of chain formation and some of its consequences

Carlos Muñoz Pérez  
cmunozperez@filo.uba.ar

November 11, 2014

Copy Theory (Chomsky 1993) is based on the idea that the displacement property of Language is an epiphenomenon based on how non-trivial chains are phonetically realized. Thereby, key properties of syntactic movement are supposed to be derived from the mechanisms that allow chain formation and interpretation at the interfaces. Non-distinctiveness, the idea that all the links in a chain should be “the same” in some relevant way, is a fundamental component of Copy Theory. However, at the moment there are no theories about what non-distinctiveness is; it is usually assumed to be an axiomatic property that needs to be implemented by introducing special descriptive technology (e.g., two elements are “the same” if they share the same index). In this presentation it will be proposed that non-distinctiveness is based on an inclusion relation between the features of the syntactic objects forming a chain: higher copies in a chain receive additional features throughout the syntactic derivation, which is why they are supersets of their lower counterparts. This “sameness” criterion predicts, at least, two different kinds of phenomena: cases where two copies are pronounced differently and cases where two elements that are not copies are interpreted as members of the same chain. Both cases will be illustrated respectively with examples from non-identical copy realization in Dutch and null subjects in Spanish.

### 1. Three general problems introduced by Copy Theory

Under *Trace Theory* (Fiengo 1977) and assuming the GB framework (Chomsky 1981, 1986), a sentence like (1) would receive an analysis in line with (2).

(1) John said that John was kissed.

(2)  $[_{TP} \text{John}_i [_{T'} T [_{VP} t_i [_{V'} \text{said} [_{CP} \text{that} [_{TP} \text{John}_j [_{T'} \text{was} [_{VP} \text{kissed } t_j]]]]]]]]]$ .

Thus, under a classic definition of *chain* (“a reflection of a ‘history of movement’, consisting of the positions through which an element has moved from the A-position it occupied at D-Structure”; Chomsky 1986), it is possible to assert that there are two (non-trivial) chains in this sentence.

(3) a.  $\text{CH}_1 = (\text{John}_i, t_i)$   
b.  $\text{CH}_2 = (\text{John}_j, t_j)$

Now, consider the same sentence under the so-called *Copy Theory of Movement* (Chomsky 1993). Remember: under Copy Theory, movement does not exist as an operation (i.e., nothing really moves), but it is an epiphenomenon of how the PF component phonetically realizes chains consisting of *identical* copies.

(4)  $[_{TP} \text{John}_1 [_{T'} T [_{VP} \text{John}_2 [_{V'} \text{said} [_{CP} \text{that} [_{TP} \text{John}_3 [_{T'} \text{was} [_{VP} \text{kissed } \text{John}_4]]]]]]]]]$ .

According to (4), we have four copies of *John*. As linguists, we know that *John*<sub>1</sub> and *John*<sub>2</sub> are related by a movement dependency (i.e., *John*<sub>1</sub> is a copy of *John*<sub>2</sub>), whereas *John*<sub>3</sub> is not transformationally related with *John*<sub>1</sub> and *John*<sub>2</sub>, but it is with *John*<sub>4</sub>. Thus, we may suppose that *John*<sub>1</sub> and *John*<sub>2</sub> form a chain CH<sub>1</sub> while *John*<sub>3</sub> and *John*<sub>4</sub> form a chain CH<sub>2</sub>.

(5) a.  $\text{CH}_1 = (\text{John}_1, \text{John}_2)$   
b.  $\text{CH}_2 = (\text{John}_3, \text{John}_4)$

And since the lower copy of each chain is supposed to be silent<sup>1</sup>, like a *trace*, we may also suppose that *John*<sub>2</sub> and *John*<sub>4</sub> are phonologically deleted.

(6) a.  $\text{CH}_1 = (\text{John}_1, \text{John}_2)$   
b.  $\text{CH}_2 = (\text{John}_3, \text{John}_4)$

But, as you can imagine, things are never that easy. This sketchy and inductive introduction to Copy Theory has left without answer three very important problems that any implementation of Copy Theory should try to solve. Let’s call them *The Challenges of Copy Theory*<sup>2</sup>.

- (7) *Challenges of Copy Theory*
- a. *Non-Distinctiveness Challenge*  
What kind of principle allows identifying as “the same” two instances of the same lexical item?
  - b. *Uniqueness Challenge*  
Why in the general case a chain does not surface with more than one link phonetically realized?
  - c. *Rank Challenge*  
Why lower copies cannot be realized in the general case?

---

<sup>1</sup> For concreteness, we may assume the following rule:

- (i) *Speak Up* (Bobaljik 1995: 350)  
Pronounce the topmost/leftmost copy of each element.

<sup>2</sup> I am grateful to Basima Mahmood for the idea of using the word “challenges” instead of “problems”.

Nunes (1995, 2004) proposes a version of Copy Theory that attempts to answer the challenges in (7). His proposal, however, is not exempt of problems. Based on Nunes' seminal proposals, I will offer novel answers for these three issues. Finally, I will also try to show that my proposal offers some empirical advantages.

## 2. Nunes' answers

### 2.1. THE NON-DISTINCTIVENESS CHALLENGE

What kind of syntactic principle allows Language to identify as “the same” *John*<sub>1</sub> and *John*<sub>2</sub> and as distinct *John*<sub>2</sub> and *John*<sub>3</sub> in a representation like (4)? In other words, what kind of criterion does grammar use to form chains?

We know that it cannot be a phonological criterion of identity or a lexical criterion of identity.

According to Nunes (1995), it is not a criterion related with the featural checking of each syntactic object either. Consider the representation in (8), where the (uninterpretable) Case features of each copy of *John* are sketched.

(8) [ ... John<sub>C</sub> ... John<sub>uC</sub> ... John<sub>C</sub> ... John<sub>uC</sub> ]

The first and the second copies are considered “the same” regardless the fact that their Case features are not identical. However, the same difference is observed between the second and the third copies, and they are considered distinct syntactic objects.

Since there seem to be no principled way to derive the “sameness” relation between copies, Chomsky (1995) and Nunes (1995, 2004) introduce indexes to indicate when two (or more) syntactic objects are non-distinct.

I will follow Chomsky's (1995, 227) proposal that two lexical items *l* and *l'* selected from a numeration should be marked as distinct for the computational system if they are accessed by distinct applications of Select. As for phrasal objects, I assume that their labels encode the relevant piece of information regarding distinctiveness; the DP [the man], for instance, is to be represented in bare phrase structure terms as  $K = \{\text{the}^i, \{\text{the}^i, \text{man}^k\}\}$ . Given that Copy just replicates the targeted material and does not alter distinctiveness markings, it is possible to determine, at any point in a given derivation, whether two terms with identical sets of features are to be interpreted as copies or distinct constituents. (Nunes 2004: **XX**)

For convenience, let's say that Nunes' system is based on a definition of non-distinctiveness like the following.

(9) *Non-Distinctiveness*  
Two syntactic objects are non-distinct when they share the same index.

Thus, non-distinctiveness remains as an axiomatic property of chains in Nunes' proposal.

### 2.2. THE UNIQUENESS CHALLENGE

We may summarize this problem by observing the pattern in (10). Why is (10b) unacceptable?

(10) a. John was kissed ~~John~~.  
b. \*John was kissed John.

According to Nunes, the fact that only one copy is pronounced follows from the *Linear Correspondence Axiom* (LCA) of Kayne (1994). LCA imposes an antisymmetry requirement on syntactic structure and linear orderings:

(11) *Antisymmetry*  
If X precedes Y, then Y cannot precede X.

Since two (or more) copies of a constituent are “the same” (in a sense to be discussed later), a PF representation like (10b) would not comply with the antisymmetry condition. Thus, to linearize these kind of representations it is necessary to reduce all movement chains to only one link. The mechanism in charge to do so is called *Chain Reduction*.

(11) *Chain Reduction* (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.

### 2.3. THE RANK CHALLENGE

Again, the problem may be summarized in a pair. Why is (12b) unacceptable?

(12) a. John was kissed ~~John~~.  
b. \*~~John~~ was kissed John.

An obvious observation first: in order to explain why the higher copy is usually pronounced, we need to assume that there is some kind of **asymmetry** between the copies in a chain. It is very simple: if all copies were exactly the same, there would be no way to explain why one of them, the higher copy in the chain, is systematically selected to be pronounced.

In this sense, despite the fact that there is a “sameness” relation between the copies in a non-trivial chain, **their formal features may be different**. This is clear in (13), where the lower copy of John bears an uninterpretable Case feature (uC), whereas the higher copy has already checked that feature with the T head.

(13) [TP John<sub>C</sub> [T' was [v<sub>P</sub> kissed John<sub>uC</sub>]]]

This representation is not legible at PF since, by definition, uninterpretable features (uFF) cause the crash of the derivation at the interfaces. Thus, it is necessary to assume that some kind of operation is applied to (13) in order to eliminate all the uFF in the syntactic representation. This operation is defined in (14).

(14) *FF-Elimination (Nunes 2004: 31)*

Given the sequence of pairs  $\sigma = \langle (F, P)_1, (F, P)_2, \dots, (F, P)_n \rangle$  such that  $\sigma$  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  $\sigma$  to satisfy Full Interpretation at PF.

Assuming (14) it is possible to explain the contrast between (12a) and (12b). If, as in (12a), the lower copy of *John* is deleted, then it is not necessary to apply FF-Elimination since Chain Reduction has already deleted the problematic uC feature; on the contrary, if the higher copy of *John* is deleted, as in (12b), then it will be necessary to apply FF-Elimination on the uC of the lower copy to make the representation legible at PF. Schematically:

(15) a. [TP John<sub>C</sub> [T' was [v<sub>P</sub> kissed John<sub>uC</sub>]]] ONLY REQUIRES CHAIN REDUCTION  
 b. [TP ~~John~~<sub>C</sub> [T' was [v<sub>P</sub> kissed John<sub>uC</sub>]]] ALSO REQUIRES FF-ELIMINATION

Given that the derivation in (15a) is shorter and involves fewer operations than (15b), the later is blocked by a (global) economy consideration.

If derivational economy is assumed (as in most versions of the Minimalist Program), grammatical operations only may take place if they prevent the syntactic derivation from crashing. The classic principle implementing this idea is the Last Resort condition:

(16) *Last Resort (Collins 1997: 9)*

An operation OP involving  $\alpha$  may apply only if some property of  $\alpha$  is satisfied.

Uninterpretable features are a particular implementation of this idea: they are properties of syntactic objects that need to be satisfied to make the syntactic representation legible at the interfaces. This is, at least, the way they work in Chomsky's (1995) system

(17) *Uninterpretable feature checking in Chomsky's (1995) system*

a. [TP T [v<sub>P</sub> DP<sub>uF</sub> [v' ...]]] UNINTERPRETABLE REPRESENTATION  
 b. [TP DP<sub>uF</sub> [T' T [v<sub>P</sub> DP<sub>uF</sub> [v' ...]]]] INTERPRETABLE REPRESENTATION

The problem with a representation as (17b) is that both copies are exactly the same regarding their features, so no asymmetry persists to explain why the higher copy should be pronounced.

Nunes's system, on the other hand, maintains certain asymmetry between the copies in a chain: only new copies check their uFF, while the lower ones preserve them. However, this raises a problem for the very postulation of uninterpretable features. Consider the following derivation.

(18) *Uninterpretable feature checking in Nunes' (1995, 2004) system*

a. [TP T [v<sub>P</sub> DP<sub>uF</sub> [v' ...]]] UNINTERPRETABLE REPRESENTATION  
 b. [TP DP<sub>uF</sub> [T' T [v<sub>P</sub> DP<sub>uF</sub> [v' ...]]]] UNINTERPRETABLE REPRESENTATION

If the application of movement operations is vacuous regarding the interpretability of the original occurrence of the DP, then it cannot be maintained that the operation depicted in (18) complies with the Last Resort condition. And, as said, since uFF are part of the technical implementation of Last Resort, their postulation in this system becomes gratuitous.

As a précis, the form of the argument is the following:

- There must be some asymmetry between copies to explain which one is pronounced.
- Since they are "copies", such an asymmetry should proceed from the syntactic engagement of each copy with the main structure they were merged with.
- If this asymmetry is implemented in terms of uFF, we have a problem: uFF need to be deleted by syntactic mechanisms to comply with the Last Resort Condition
- But, if the uFF are deleted, there is no asymmetry between copies.

Personally, I think there is some tension between a principled implementation of Copy Theory and the use of uFF to motivate movement.

### 3. An inclusion-based approach to chain formation and two of its predictions

The system I am going to sketch here is based on four main assumptions regarding the nature of syntactic computation.

(19) *Assumption A*

Syntax is about sets.

Following an important tradition in generative grammar, I assume that lexical items are sets of features and phrases are sets of lexical items. This is a very usual idea.

(20) *Assumption B*

Late insertion: phonological information is introduced at PF.

The idea is that syntactic categories are purely abstract, having no phonological content (Halle & Marantz 1993). The phonological matrixes are introduced post-syntactically by an operation called *Vocabulary Insertion* (e.g., Embick & Noyer 2001, 2004). Late Insertion is a hypothesis independently motivated by syncretic (e.g., Bonet 1991) and allomorphic (e.g., Embick 2010) patterns.

Thus, a first big difference with Nunes' system arises. In Nunes' system, the phonological features of syntactic terminals are present throughout the entire syntactic derivation, so the lower links in a chain are *deleted* at PF. In the present case, syntactic terminals are phonologically "empty", so PF is in charge of "deciding" which links in a chain are going to receive phonological representation.

- (21) *Assumption C*  
There are no uFF in the moving elements.

As said, uFF are, at least, problematic to code some asymmetry between the links in a chain. For the moment, let's assume that movement operations are triggered by a functional head  $X^0$  attracting an YP to its specifier position and not by a specific requirement of  $YP^3$ .

- (22) *Assumption D*  
If an XP enters in an *Agree* relation with a (non-defective<sup>4</sup>) head  $H^0$ ,  $H^0$  assigns a feature [H] to XP.

So, for example, if a DP agrees with an inflected T head, the DP will be assigned a T-feature that is going to be interpreted as nominative Case by the system. Thus, every syntactic object that enters in an Agree relation with a non-defective head is going to receive an additional feature.

Let's suppose an scenario where a non defective *Probe*  $Y^0$  with an EPP-feature enters in an Agree relation with a *Goal*  $\alpha_{\{\dots\}}$ , where  $\{\dots\}$  is used to indicate the set of features of  $\alpha$ .

- (23)  $[_{YP} Y^0_{EPP} [_{XP} \dots \alpha_{\{\dots\}}]] \rightarrow [_{YP} \alpha_{\{Y, \dots\}} [_{Y'} Y^0_{EPP} [_{XP} \dots \alpha_{\{\dots\}}]]]$

There is a relation of Inclusion between both copies of  $\alpha$ : the set  $\{\dots\}$  is a subset of  $\{Y, \dots\}$  ( $\alpha_{\{Y, \dots\}} \subseteq \alpha_{\{\dots\}}$ ). This inclusion relation between copies is going to arise systematically every time a new copy is generated (and enters in an Agree relation with a non-defective functional head).

<sup>3</sup> Of course, abandoning uFF is not that simple. However, I am not going to talk today about the problems related with this assumption.

<sup>4</sup> I am not sure about this "non-defectiveness" condition. See, for example, the case in (62).

Since the inclusion relation between copies is systematic, it is possible to use to define the conditions on chain formation:

- (24) *Conditions on Chain Formation*<sup>5</sup>  
Two constituents  $\alpha$  and  $\beta$  form a chain iff
- $\alpha$  is a superset of  $\beta$ ;
  - $\alpha$  c-commands  $\beta$ ;
  - There is no syntactic object  $\gamma$  between  $\alpha$  y  $\beta$  such as (i)  $\beta$  is a superset of  $\gamma$  and (ii)  $\gamma$  is not a superset of  $\alpha$ .

The conditions on (24) make no use of inductive and maybe more complex definitions of "sameness".

- (25) *Kinds of "sameness"*
- |    |                      |  |                          |
|----|----------------------|--|--------------------------|
| a. | Strict identity:     | $(CH = (A, B), \text{ iff } A = B)$        | It is not going to work! |
| b. | Non-Distinctiveness: | $(CH = (A^1, B^1))$                        | Nunes (2004)             |
| c. | Inclusion:           | $(CH = (A, B) \text{ iff } B \subseteq A)$ |                          |

This is very elegant solution from a minimalist point of view to the Non-Distinctiveness Challenge posed by Copy Theory: inclusion is a basic relation between sets, even more than strict identity since the later may be defined in terms of the former.

- (26) *Strict Identity*  
A set A is strictly identical to a set B ( $A = B$ ) iff (i) A is a subset of B ( $A \subseteq B$ ), and (ii) B is a subset of A ( $B \subseteq A$ ).

Thus, even if a principled explanation for the "sameness" types in (25a) and (25b) is proposed, the option involving inclusion would be preferable due its simplicity.

Once the Non-Distinctiveness Challenge has been surpassed, we should focus on the Uniqueness Challenge and the Rank Challenge. For expository purposes, we are going to attempt to explain the following pattern.

- (27)
- John was kissed ~~John~~
  - \*~~John~~ was kissed John
  - \*John was kissed John
  - \*~~John~~ was kissed ~~John~~

<sup>5</sup> I believe it is useful to think about these conditions as some kind of "common factor" principle of chain formation: the thing that all elements in a chain have in common is containing the features of the original externally-merged lexical item.

In a Late Insertion system, the phonological realization of the syntactic terminals depends on the application of the Vocabulary Insertion operation. Thus, an economy criterion on its application may explain the unacceptability of (27c), answering the question posed by the Uniqueness Challenge<sup>6</sup>. But we still need to explain the rest of the pattern:

- (28) a. John was kissed ~~John~~  
 b. \*~~John~~ was kissed John  
 d. \*~~John~~ was kissed ~~John~~

A simple way to derive such a pattern is by appealing to general conditions on recoverability of deletion, “which we may assume to exist though they are not understood in detail” (Chomsky 1977: 86). An operative definition of these conditions (based in Chomsky’s 1964 original definition<sup>7</sup>) is proposed in (29):

- (29) *Condition on Recoverability of Deletion*  
 An element may be deleted (i.e., not pronounced) if it is totally determined by a structurally related syntactic constituent.

Thus, it is explained why it is necessary to pronounce at least one copy (which derives the unacceptability of (28d) and it is predicted that the higher copy in a chain (i.e., the link in the chain with more features) is the one that should receive phonological representation<sup>8</sup>. This theoretical result is captured in the following principle:

- (30) *Pronounce the Superset*  
 Given a Chain CH, apply Vocabulary Insertion to the link which is the superset of all the remaining links.

The definition in (30) is nothing more than a Copy Theory-oriented version of the well-known general principle “choose the more specific form”, which underlies several competition-based linguistic mechanisms.

<sup>6</sup> There are, at least, two other ways to surpass the Uniqueness Challenge. The first one is by assuming that linear orderings need to comply with an antisymmetry requirement, just as Nunes (1995, 2004) did. The other one is, I think, theoretically more appealing: it consists on assuming that Vocabulary Insertion is an operation that takes (trivial and non-trivial) chains as inputs. This second option, however, requires further elaboration.

<sup>7</sup> “A transformation can delete an element only if this element is the designated representative of a category, or if the structural condition that defines this transformation states that the deleted element is structurally identical to another element of the transformed string” (Chomsky 1964: 41).

<sup>8</sup> In a Late Insertion system, this result is necessary given the logic of the *Subset Principle* (Halle 1997). The Subset Principle presupposes that “the terminal nodes that are the sites for insertion are fully specified; that is to say, they contain a full complement of syntactico-semantic features” (Embick & Noyer 2004: 299).

One of the main claims of this system is that syntactic “sameness” does not involve a strict type of identity, but a “flexible” one. Particularly, there are two important predictions that a system of chain formation based on a “flexible” criterion of identity makes:

- (31) *Predictions we are going to explore*  
 a. Transformationally related syntactic objects (i.e., copies) may pertain to different chains at PF if some morpho-phonological operation disrupts the inclusion relation between their features.  
 b. Non-transformationally related syntactic objects (i.e., non-copies) may pertain to the same chain if there is an inclusion relation between their features.

#### 4. Multiple copy phenomena and non-identical doubling in Dutch

One of the most important sources of evidence for Copy Theory are the so-called cases of multiple copy realization: in several languages it is possible to pronounce both the head of the chain and some other link:

- (32) *German (McDaniel 1986)*  
**Wem** glaubt Hans **wem** Jakob gesehen hat?  
 Who thinks Hans who Jakob seen has  
 ‘Who does Hans think Jakob saw?’

- (33) *Romani (McDaniel 1986)*  
**Kas** misline **kas** o Demiri dikhlâ?  
 Who think who Demir saw  
 ‘Who do you think Demir saw?’

- (34) *Dutch (Drenthe) (Barbiers et al. 2010)*  
**Wie** denk je **wie** ik gezien heb?  
 Who think you who I seen have  
 ‘Who do you think I have seen?’

Nunes (2004) argues that this type of pattern is explained if it is assumed that an intermediate copy may be morphologically reanalyzed as part of some functional head (e.g., a complementizer) through an application of *Fusion*.

- (35) [<sub>CP</sub> Wen<sup>i</sup> ... [<sub>CP</sub> [<sub>C</sub> wen<sup>i</sup> C<sup>0</sup>] [<sub>TP</sub> ... wen<sup>i</sup> ... ]]]

- (36) *Fusion (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.

Therefore, assuming Chomsky's (1995) proposal that the LCA does not apply word-internally, the phonetic realization of this kind of copy is explained in terms of economy: since they are not going to violate the antisymmetry condition, it is not necessary to erase them.

This analysis makes two very strong predictions regarding the nature of the multiple copy phenomena: (i) only cyclic movement-based copies in "intermediate positions" may produce the kind of configuration that is necessary to apply Fusion, so it is predicted that there is no multiple copy pronunciation involving the tail a chain (see (37) vs (38)); and (ii) since Fusion is a morphological operation, it is predicted that there is no multiple copy pronunciation of full phrases (38).

- (37) *German (Fanselow and Mahajan 1995)*  
 Wen denkst Du wen sie meint wen Harald liebt?  
 who think you who she believes who Harald loves  
 'Who do you think that she believes that Harald loves?'

- (38) *German (Nunes 2004)*  
 \*Wen glaubt Hans wen Jakob wen gesehen hat?  
 whom thinks Hans whom Jakob whom seen has  
 'Who does Hans think Jakob saw?'

The same kind of analysis (and its consequences) may be implemented in the inclusion-based system. The following is the syntactic representation I am assuming for (32)<sup>9</sup>.

- (39)  $[_{CP} \text{wen}_{\{Q, \text{Acc}, D, \phi, \}} [_{C'} C_Q \dots [_{CP} \text{wen}_{\{\text{Acc}, D, \phi, \}} \dots [_{TP} \dots [_{VP} V \text{wen}_{\{\text{Acc}, D, \phi, \}}]]]]]$

For convenience, let's assume the following definition of Fusion:

- (40) *Fusion (Embick 2010: 78)*  
 $[_x \alpha] \frown [_y \beta] \rightarrow [_{x/y} \alpha, \beta]$   
 where  $\alpha$  and  $\beta$  are features of X and Y.

When subordinate C and *wen* fuse, the result is a syntactic terminal with a set of features containing both the features of *wen* and C:

- (41)  $[_{CP} \text{wen}_{\{Q, \text{Acc}, D, \phi, \}} [_{C'} C_Q \dots [_{CP} \#C + \text{wen}_{\{C, \text{Acc}, D, \phi, \}} [_{TP} \dots [_{VP} V \text{wen}_{\{\text{Acc}, D, \phi, \}}]]]]]$

According to the Conditions on Chain Formation, there are two chains in (41). And, as normal, the head of each chain is pronounced.

<sup>9</sup> I am assuming that the symmetric c-command relation between the subordinate C and *wen* is obtained by *m-merger* (Matushansky 2006).

- (42) a.  $CH_1 = (\text{wen}_{\{Q, \text{Acc}, D, \phi, \}})$   
 b.  $CH_2 = (\#C + \text{wen}_{\{C, \text{Acc}, D, \phi, \}}, \text{wen}_{\{\text{Acc}, D, \phi, \}})$

A more complex kind of pattern is the one attested by cases of non-identical doubling in Dutch. The following data is from Barbiers et al (2010).

- (42) *Neuter and non-neuter wh-pronouns (Overijssel)*  
 Wat denk je wie ik gezien heb?  
 What think you who I seen have  
 'Who do you think I saw?'

- (43) *Non-neuter and (non-neuter) relative pronouns (North-Holland)*  
 Wie denk je die ik gezien heb?  
 Who think you REL.PRON I seen have  
 'Who do you think I saw?'

- (44) *Neuter and (non-neuter) relative pronouns (Overijssel)*  
 Wat denk je die ik gezien heb?  
 What think you REL.PRON I seen have  
 'Who do you think I saw?'

These are the only possible doublings in Dutch. Any other logical combination is unacceptable.

- (45) \*Wie denk je wat ik gezien heb?  
 Who think you what I seen have  
 'Who do you think I saw?'

- (46) \*Die denk je wie ik gezien heb?  
 REL.PRON think you who I seen have  
 'Who do you think I saw?'

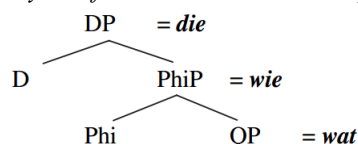
- (47) \*Die denk je wat ik gezien heb?  
 REL.PRON think you what I seen have  
 'Who do you think I saw?'

Thus, the only possible orders are the ones sketched in (48).

- (48) a. wat (neuter pronoun) ... wie (non-neuter pronoun)  
 b. wie (non-neuter pronoun) ... die (non-neuter relative pronoun)  
 c. wat (neuter pronoun) ... die (non-neuter relative pronoun)

To explain these facts, Barbiers et al. (2010) assume that pronouns are phrasal (i.e., pronouns are not spell-outs of terminals but spell-outs of phrases). They propose, then, that *die*, *wie* and *wat* are realizations of different layers of a DP projection:

(49) *Layers of the non-neuter relative pronoun 'die'*



To provide an explanation, they provide an analysis of these pronouns according to the following featural composition.

- (50)
- a. **wat** = indefinite numeral
  - b. **wie** = *wat* +  $\phi$ -features (gender)
  - c. **die** = *wie* + definiteness

And since the patterns in (48) always show the most defective element at the left position, they proposed the following theoretical generalization:

- (51) In a syntactic movement chain, a higher chain link is not more specified than a lower chain link.

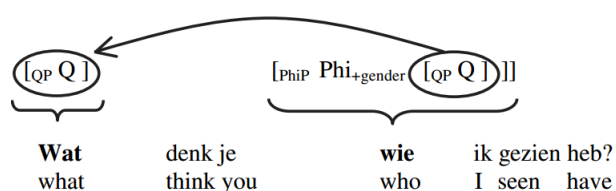
(Which, as you may notice, is exactly the opposite idea to what I am proposing).

Their analysis involves an operation they call *partial copying*, an operation that according to them follows from Copy Theory:

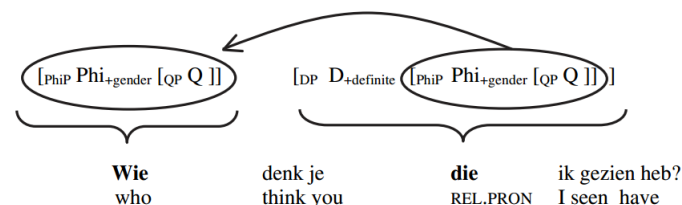
This theory allows the syntax to copy a constituent  $\alpha$  and remerge  $\alpha$  higher in the structure. This will give standard movement. What syntax should also be allowed to do is to partially copy  $\alpha$ . This is what happens with sub-extraction: trivially, an object must be able to move out of the VP, stranding the rest of the VP. (Barbiers et al. 2010: 4).

Thus, they propose an analysis based on partial copying: non-identical doubling of wh-pronouns follows from sub-extracting a low layer from a more specified pronoun.

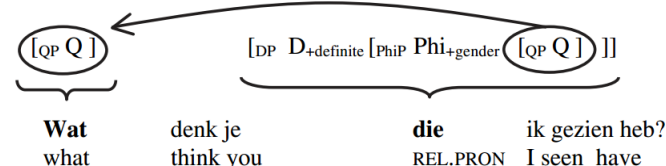
(52) *Neuter and non-neuter pronouns (cf. (42))*



(53) *Non-neuter and (non-neuter) relative pronouns (cf. (43))*



(54) *Neuter and (non-neuter) relative pronoun (cf. (44))*

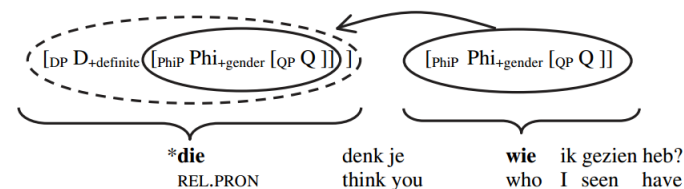


This analysis predicts successfully the unacceptability of cases as (45), (46) and (46).

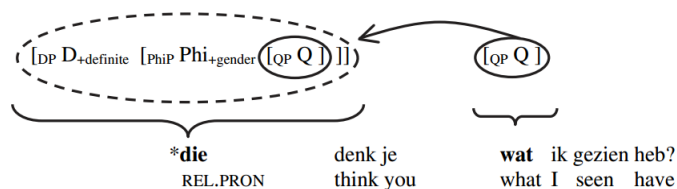
(55) *Non-neuter and neuter pronouns (cf. (45))*



(56) *(Non-neuter) relative and non-neuter pronouns (cf. (46))*



(57) (Non-neuter) relative and neuter pronouns (cf. (47))



Elegant as it is, I see three problems with this analysis. First, it does not explain why doubling is restricted to embedded contexts. In other words, under the assumptions of Barbiers et al. (2010) there is no obvious reason to rule out (58).

(58) \***Wat** heb ik **wie** gezien?  
 What have I who seen  
 'Who did I saw?'

Second, it does not explain why doubling is restricted to morphologically simple units and never is attested with full phrases.

Third, as far as I understood, they seem to assume that sub-extraction (i.e. partial copying) implies doubling ("if after partial copying only the higher chain link is spelled out, this inevitably creates a recoverability problem. Hence, partial copy entails doubling"; Barbiers et al 2010: 4). In order to rule out this erroneous prediction, they seem to require assuming some kind of distinction between "sub-extraction from complex words" and "sub-extraction from phrases", which posits a new set of problems (e.g., excorporation, lexical integrity).

The analysis I propose solves the second problem straightforwardly: the phenomenon is restricted to morphologically simple units because it is based on a morphological operation, namely *Impoverishment* (Bonet 1991), a feature-deleting operation that precedes Vocabulary Insertion. Thus, I assume that an Impoverishment operation applies on the pronoun, an element  $P^{min/max}$ , in the specifier position of matrix C targeting (i) definiteness features (D) or (ii)  $D + \phi$ -features.

The following is a sketch of the syntactic structure where the non-neuter relative pronoun *die* moves to the specifier of  $C_Q$  and receives a Q-feature.

(59)  $[_{CP} P_{\{D,\phi,N,Q\}} [_{C'} C_Q [_{CP} P_{\{D,\phi,N\}} [_{C'} C [_{TP} \dots P_{\{D,\phi,N\}} ] ] ] ] ]$

This representation is sent to PF, where the impoverishment operation applies on the higher copy. As said, it may delete the (i) D-feature or (ii) both D and  $\phi$ -features.

(60) a.  $P_{\{D,\phi,N,Q\}} \rightarrow P_{\{\phi,N,Q\}}$   
 b.  $P_{\{D,\phi,N,Q\}} \rightarrow P_{\{N,Q\}}$

(61) a.  $[_{CP} P_{\{\phi,N,Q\}} [_{C'} C_Q [_{CP} P_{\{D,\phi,N\}} [_{C'} C [_{TP} \dots P_{\{D,\phi,N\}} ] ] ] ] ]$   
 b.  $[_{CP} P_{\{N,Q\}} [_{C'} C_Q [_{CP} P_{\{D,\phi,N\}} [_{C'} C [_{TP} \dots P_{\{D,\phi,N\}} ] ] ] ] ]$

Since the higher copy of P is not a superset of the following copy in neither case, it will form a different chain (i.e., it is going to be pronounced). And following the featural characterization in (48), (61a) is going to be spelled-out as (43) and (61b) as (44).

It remains to explain what triggers the Impoverishment operation. It is not clear to me yet what kind of condition is involved, but it may be related with some kind of agreement between the moving pronoun and the embedded C. For now, let's say that the *wh*-pronoun agrees with the embedded C on its way to  $C_Q$ , and some feature is assigned to P (a C-feature)

(62)  $[_{CP} P_{\{D,\phi,C\}} [_{C'} C [_{TP} \dots P_{\{D,\phi\}} ] ] ]$

Later, when P moves to  $C_Q$ , the impoverishment operation applies triggered by the C-feature.

(63) *Impoverishment of D-features*  
 $\{D, \phi, C, Q\} \rightarrow \{\phi, Q\} / \text{--- } C_Q$

This would also explain why doubling is restricted to embedded contexts. However, a more principled motivation for (63) is still required.

## 5. A non-insertion approach to null subjects in Spanish

As is well known, Spanish has *null subjects*, silent gaps that may be interpreted referentially (and not impersonally, for example).

(64) (El niño) corre rápido.  
 The kid runs fast  
 'The kid runs fast'.

(65) \*(The kid) runs fast.

This property is usually related with the rich verbal inflection of the language.

(66) *Taraldsen's Generalization*

There is a connection between the possibility of referential, definite silent pronominal subjects of finite clauses and the notional "richness" of the verbal agreement paradigm.



(67) *Theory I*  
In null subject languages, verbal inflection is pronominal and may function as a subject.

(68) *Theory II*  
Rich verbal inflection licenses and identifies an empty category (namely *pro*).

(69) *Theory III*  
Verbal inflection establishes some kind of “sameness” relation with a pronoun, which causes the deletion of the later.

This is the most “popular” theory. It was originally proposed by traditional grammarians (e.g., Fernández Ramírez 1951), but it has been recently updated by several scholars (e.g., Alexiadou & Anagnostopoulou 1998, Ordoñez 1997, Ordoñez & Treviño 1999).

- (70) a. Corre rápido.  
Runs fast  
b. Corr<sub>[PRONOUN -e]</sub> rápido.
- (71) a. El niño corre rápido.  
The kid runs fast  
b. [El niño]<sub>i</sub> corr<sub>[PRONOUN -e]<sub>i</sub></sub> rápido.

(72) *Assumption on the interpretability of inflectional features*  
In null subject languages, the features of the T head are interpretable (i.e., they are not assigned to T by any agreement or government relation).

(73) *Tense identity on verbs* (Murguía 2004: 86)  
 \*Juan ha leído mucho y María en el futuro ~~habrá leído mucho~~ también.  
 'Juan has read a lot and María in the future ~~will read a lot~~ too'.

- (74) *Gender identity on nouns* (Saab 2009: 493)  
 \*Juan visitó a su tía y Pedro visitó al ~~tío~~ suyo.  
*'Juan visited his aunt and Pedro visited his ~~uncle~~.'*
- (75) *Gender identity on adjectives* (Saab 2009: 544)  
 Juan es petis-o y María también ~~es petis-a~~.  
*Juan is short-MASC and María ~~is short-FEM~~ too*
- (76) *Inflection identity on verbs* (Saab 2009: 234)  
 Juan fue al cine y nosotros también ~~fuimos al cine~~.  
*Juan went.3SG to the cinema and we too ~~[went.1PL to the cinema]~~*

An important aspect of a theory that assumes the existence of silent categories is the explanation of its distribution.

- |      |            |                                  |                 |                      |          |                       |                 |          |              |  |
|------|------------|----------------------------------|-----------------|----------------------|----------|-----------------------|-----------------|----------|--------------|--|
| (77) | <i>pro</i> | corre<br>runs                    | rápido.<br>fast |                      |          |                       |                 |          |              |  |
| (78) | A:         | Este<br>This                     | libro<br>book   | es<br>is             | de<br>of | Juan.<br>Juan         |                 |          |              |  |
|      |            | <i>'This is Juan's book'</i>     |                 |                      |          |                       |                 |          |              |  |
|      | B:         | *Y<br>And                        | este<br>this    | cuaderno<br>notebook |          | también es<br>also is | [ <sub>PP</sub> | de<br>of | <i>pro</i> ] |  |
|      |            | <i>'And also this notebook'.</i> |                 |                      |          |                       |                 |          |              |  |

(79) *Identification of pro according to Chomsky's (2001) system*

$$[\text{TP } \textit{pro}_{\text{u}\varphi} [\text{T}' \text{ T}_{\varphi} \dots]] \xrightarrow[\text{Agree}]{} [\text{TP } \textit{pro}_{\varphi} [\text{T}' \text{ T}_{\varphi} \dots]]$$

Thus, it is necessary to assume (72) in order to implement this kind of analysis. And as we saw, there are reasons to believe that (72) is wrong.

### 5.3. BRIEF DISCUSSION OF THEORY III

The third option is assuming that null subjects are ordinary pronouns that, by some PF reason, are deleted (e.g., Perlmutter 1969). A priori, this kind of theory allows explaining the ellipsis-related pattern attested in (73) to (76).

- (80) a.  $[_{DP} \text{El niño}] \text{ corre rápido.}$   
 b.  $[_{TP} \text{SD}_{\varphi} [_{T'} \text{T}_{\varphi} \dots]]$

- (81) a.  $\emptyset \text{ corre rápido.}$   
 b.  $[_{TP} \text{D}_{\varphi} [_{T'} \text{T}_{\varphi} \dots]]$

The problem for these theories is how to explain the silent nature of null subjects. Let's see if we can derive it from the inclusion-based implementation of Copy Theory.

### 5.4. SHUSHING SUBJECTS

There are two important antecedents for my analysis: Roberts (2010) and Saab (2009). Despite the fact that I am using elements of both proposals, this particular implementation makes a different and really important prediction (I think).

Let's begin with some assumptions. First, we need to explain the parametric variation regarding null-subjects. Following Roberts & Holmberg (2010), I will assume that presence of a definiteness D-feature in the T head is responsible for the licensing of null subjects.

- (82) *The Null Subject Parameter (Roberts & Holmberg 2010: 14)*  
 Does T bear a D-feature?

Second, I assume that null-subjects are weak pronouns D with  $\varphi$ -features. They are merged in the [Spec,v] positions, where they receive a  $\theta$ -role.

- (83)  $[_{TP} \text{T}_{\{T, D\}} [_{vP} \text{D}_{\{D, \varphi\}} \dots]]$

As in a standard derivation, T needs to value its  $\varphi$ -features. Thus, there is an Agree relation between T and D. Since T is non-defective, it is going to assign a feature to D, the T-feature (namely, nominative Case).

- (84)  $[_{ST} \text{T}_{\{T, D, \varphi\}} [_{SV} \text{D}_{\{D, \varphi, T\}} \dots]]$

Given our definitions, T and D form a non-trivial chain  $CH = (T, D)$ . However, since both elements are identical, there seem to be no way to decide which link is going to be deleted.

- (85) *Possible realizations for the chain  $CH = (T, D)$*   
 a.  $CH = (T, \emptyset)$   
 b.  $CH = (\emptyset, D)$

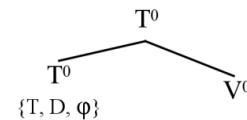
The alternative in (85b), however, violates some PF principles. As is well known, in Spanish it is not possible spelling-out the verbal root without any inflection.

- (86) a. canto, cantábamos, cantar, cantarás  
 sing.1SG, sing-PAST-1PL, sing.INF, sing-FUT-2.SG  
 b. \*cant

The restriction in (86) may be explained by assuming any of two independently proposed PF conditions:

- (87) *Excorporation Filter (Baker 1988: 73)*  
 A trace can never be non-exhaustively dominated by a zero-level category (i.e., there are no traces inside words).  
 (88) *Subword Deletion Corollary (adapted from Saab 2009: 375)*  
 No Subword can be subject to non-insertion if the Morphological Word that contains it is not subject to non-insertion (I-assigned).

- (89) *Structure of the T node due V-to-T movement*



(So, it seems that PF conditions may override the principles of chain realization, as Franks (1998), Boskovic (2002) and Nunes (2004) previously observed).

As a consequence, the Vocabulary Insertion is applied on the T head while the pronoun D is not pronounced.

As you may notice, this analysis does not require assuming an EPP-feature on T in order to derive the silent nature of D. This is the most important difference regarding the proposals in Roberts (2010) and Saab (2009): for their deletion mechanisms to work, the weak pronoun D should occupy the [Spec,T] position.

- (88)  $[_{TP} \emptyset [_{T'} \text{T}_{EPP} \dots]]$

However, as Barbosa (2009) observes, assuming that T has an EPP-feature in null-subject languages makes a wrong prediction regarding their lack of *that-trace effect*.

- (89) \*Who did John say that bought a computer?

(90) Quién dijo Juan que compró una computadora?  
 Who said Juan that bought a computer

As Rizzi (1982) and Jaeggli (1984) suggested, it seems that this contrast follow from the fact that Spanish and other null subject languages allow free inversion:

(91) Compró Juan la computadora.  
 Bought Juan the computer.

Since Spanish allow post-verbal subjects, it follows that subject extraction in cases as (90) may be from post-verbal positions, explaining the parametric difference with English (89) and some other non-pro-drop languages.

(92) [<sub>CP</sub> Quién [<sub>C</sub> C [<sub>TP</sub> dijo [<sub>VP</sub> Juan ... [<sub>CP</sub> que [<sub>TP</sub> compró [<sub>VP</sub> ~~quién~~ ... ]]]]]]]]

Thus, with Barbosa (and others) I contend that there is no EPP-feature driven subject movement to pre-verbal position in the Romance consistent Null Subject Languages (NSL).

There are some very nice predictions from this assumption. But I am going to talk to you about them some other day.

## 6. To conclude...

- We reviewed the problems of Nunes' (1995, 2004) Copy Theory regarding the Three Challenges.
- It was offered a principled explanation for the Three Challenges grounded on an inclusion-based mechanism of chain formation.
- It was shown how this theory might deal with multiple copy phenomena.
- It was proposed a deletion analysis of null-subjects based on the inclusion based-theory.

PS: sorry! No references this time.