

Recognizing copies: On the definition of Non-Distinctiveness¹

Carlos Muñoz Pérez

University of Buenos Aires & CONICET

cmunozperez@filo.uba.ar

Abstract

The standard approach to Non-Distinctiveness, i.e., the “sameness” relation between constituents forming a chain under Copy Theory, involves an indexing mechanism that marks as non-distinct the syntactic objects created by the Copy operation. I argue that Non-Distinctiveness is better explained as an inclusion relation between the features of constituents in a phrase marker. A representational mechanism of chain formation based on this definition is shown to offer analytical and conceptual advantages with respect to wh-copying, non-identical wh-doubling and anti-reconstruction phenomena.

Keywords

Copy theory; Non-Distinctiveness; chain formation; wh-copying; partial copying; late merger

1 Introduction

According to *Copy Theory* (Chomsky 1993), the displacement property of human language is an epiphenomenon of how the interfaces deal with collections of non-distinct constituents in a phrase marker. Thus, the movement dependency underlying a passive sentence as (1a) involves (at least) two occurrences of the constituent *Cosmo* (1b). The intuition is that both occurrences of *Cosmo* form some kind of unit, i.e., a *chain* CH = {Cosmo, Cosmo}, in virtue of being “the same” element. Chains are assumed to have distinctive properties at the interfaces (e.g., only one chain member is spelled out at PF, the elements of a chain share the same θ-roles at LF).

- (1) a. Cosmo was arrested.
b. [_{TP} Cosmo [_T was [_{VP} arrested ~~Cosmo~~]])]

Since *Non-Distinctiveness* is the key property linking together the elements in a chain, it should be regarded as one of the core concepts associated with movement dependencies under Copy Theory. However, offering a principled definition of Non-Distinctiveness allowing to distinguish between real

¹ For comments and discussion, I am grateful to Andreas Blümel, Andrés Saab, Fernando Carranza, Guadalupe Herrera, Hedde Zeijlstra, Jairo Nunes, Jovana Gajić, Laura Stigliano, Marika Lekakou, Matías Verdecchia, Pablo Zdrojewski, Romina Trebisacce, Viola Schmitt, and three anonymous reviewers. I would especially like to thank to Dennis Ott, Radek Šimík and the audience of the Workshop ‘What Drives Syntactic Computation’ at DGfS 37. This work was supported by a doctoral fellowship granted by CONICET.

copies (i.e., transformationally related elements) and unrelated tokens of a single lexical item has proven to be exceptionally difficult. Most part of the literature on Copy Theory (e.g., Chomsky 1995; Nunes 1995; 2004; 2011; Bošković & Nunes 2007; among others) adopts an approach to Non-Distinctiveness that is analogous to the “coindex via movement” mechanism assumed to be part of the *Move- α* operation in the GB framework (e.g., Lasnik & Uriagereka 1988).² It follows a proposal made by Chomsky (1995) according to which every syntactic object in a *Numeration* should be marked as distinct by the operation *Select*. Thus, in a representation like (2a), every element in the derivation carries an index distinguishing it from the remaining lexical items. When a copy of a constituent is generated, as in (2b), the Copy operation replicates all its properties, including its index.³ Once this new copy is merged with the main structure, as in (2c), the representation contains two constituents carrying the same index.

- (2) a. $K = [_{TP} \text{was}^i [_{VP} \text{arrested}^j \text{Cosmo}^k]]$
- b. $K = [_{TP} \text{was}^i [_{VP} \text{arrested}^j \text{Cosmo}^k]]$
 $L = \text{Cosmo}^k$
- c. $K = [_{TP} \text{Cosmo}^k [_{T} \text{was}^i [_{VP} \text{arrested}^j \text{Cosmo}^k]]]$

As the index allows recognizing both occurrences of *Cosmo^k* as non-distinct, they are trivially predicted to form the chain $CH = \{\text{Cosmo}^k, \text{Cosmo}^k\}$. For ease of presentation, I will call this type of marking device *Indexical Sameness*, or *Indexical-S* for short. The mechanism can be informally defined as follows:

(3) Indexical-S

Two constituents α and β are non-distinct if and only if they are assigned the same index/marking through an application of the Copy operation (or any other derivational procedure).

There are two main theoretical problems with this approach to Non-Distinctiveness. First, it violates the *Inclusiveness Condition*.

² Although, see Martin & Uriagereka (2014) for an exception.

³ There are alternative ways of introducing these indexes. For example, Nunes (1995) assumes that all lexical items in a syntactic representation are inherently distinct from each other unless two (or more) of them are specified as being copies. For him, the Copy operation can be informally defined in the following terms: “(i) if a term T has no index, Copy targets T , assigns an unused index i to it and creates a copy of the indexed term; or (ii) if a term T was already indexed by a previous copying operation, Copy simply creates a copy of the indexed term” (Nunes 1995: 86). This approach is totally equivalent to the one presented in the main text and suffers the same issues.

(4) Inclusiveness Condition

Any structure formed by the computation is constituted of elements already present in the lexical items. No new objects are added in the course of computation apart from rearrangements of lexical properties.

Since indexes (or any kind of markings) are not inherent properties of any lexical item, the condition in (4) bans them.

While many useful proposals in the literature seem to depart from Inclusiveness, non-obeying (4) is particularly significant for the case at hand. Satisfaction of Inclusiveness is supposed to be one of the key advantages of Copy Theory over *Trace Theory*.⁴ If Copy Theory requires introducing indexes to deal with Non-Distinctiveness, one of the conceptual benefits of adopting it vanishes. As Neeleman & van de Koot (2010: 332) put it, “the copy theory by itself does not resolve the tension between Inclusiveness and the displacement property of natural language”, at least if Indexical-S is assumed.

Second, Indexical-S involves no real theory of Non-Distinctiveness; it is just a marking mechanism, an inductive device to get the right chains without further complications. In contrast, a true theory of Non-Distinctiveness should be able to explain on independent grounds (i) what kind of elements count as non-distinct for grammar and (ii) what kind of criteria are taken into consideration in such a calculus.

To put it in different terms, once Copy Theory is assumed, Non-Distinctiveness should be regarded as a theoretical problem that is similar to defining the identity conditions on ellipsis, a classic topic in linguistic theory since, at least, Ross (1969). The resemblance between these two types of phenomena is evident. Both require some kind of parallelism between a phonetically realized antecedent and a silent gap; and in fact, it has been proposed that ellipsis and copy deletion are indeed two varieties of the same general phenomenon (e.g., Chomsky 1995; Donati 2003; Saab 2008). If this parallelism requirement is an active domain of inquiry for the theory of ellipsis, then it should also be so for Copy Theory. Therefore, there would be no reason to aprioristically accept and maintain an unprincipled definition of Non-Distinctiveness such as Indexical-S if an empirically adequate alternative is offered. The present paper attempts to offer such an alternative.

The structure of the paper is as follows. In section 2, I introduce the premises that allow defining Non-Distinctiveness as an *inclusion* relation between the morphosyntactic values of two constituents in a local configuration. Additionally, I argue that chain formation is based on a representational computation that takes place at the interfaces. In section 3, I compare Indexical-S in (3) and the inclusion-based approach regarding three empirical domains that have received accounts within Copy Theory: wh-copying, non-identical wh-doubling, and anti-reconstruction effects. I show that many properties of these phenomena follow as straightforward consequences of the system advanced in section 2. Section 4 summarizes the conclusions.

⁴ See Nunes (1995; 2004) for discussion.

2 Towards a principled definition of Non-Distinctiveness

I pursue a theory of Non-Distinctiveness that relies on certain assumptions on grammatical features. The premise in (5) serves as a starting point for discussing them.

- (5) Syntactic objects are abstract sets of features without any phonological content.

This is the (generalized) *Late Insertion* hypothesis advanced by proponents of *Distributed Morphology* (Halle & Marantz 1993). According to it, syntactic terminals consist only of grammatical features without any phonological information. Phonological matrixes are introduced into syntactic representations during the mapping to the phonological component due to an operation called *Vocabulary Insertion*.

Notice that the assumption in (5) treats features as the most basic unit of syntactic computation, thus some definitions are in order. I follow Gazdar et al. (1985) and Adger & Svenonius (2011) in taking grammatical features to be ordered pairs formed by an attribute and a corresponding value.

- (6) Valued feature (Adger & Svenonius 2011: 38)

- a. A valued feature is an ordered pair $\langle \text{Att}, \text{Val} \rangle$ where
- b. Att is drawn from the set of attributes, $\{\text{A}, \text{B}, \text{C}, \text{D}, \text{E}, \dots\}$
- c. and Val is drawn from the set of values, $\{\text{a}, \text{b}, \dots\}$

The set of attributes contains classes of features (e.g., Number or Gender), while the set of values contains morphosyntactic properties pertaining to these classes (e.g., singular, plural; or feminine, neuter). If a given lexical item expresses any particular property in a language (e.g., plural), it means it inherently has the corresponding attribute (e.g., Number). Therefore, there are no privative syntactic features under these assumptions; every distinctive behavior between two tokens of a constituent is due to opposing values in certain features.

Following Adger (2010), a feature is taken to be *unvalued* if it has the empty set \emptyset instead of a value.

- (7) Unvalued feature

- a. An unvalued feature is an ordered pair $\langle \text{Att}, \emptyset \rangle$ where
- b. Att is drawn from the set of attributes, $\{\text{A}, \text{B}, \text{C}, \text{D}, \text{E}, \dots\}$
- c. and \emptyset needs to be replaced with an element from the set of values, $\{\text{a}, \text{b}, \dots\}$

For simplicity, syntactic features that do not participate in processes based on valuation will be either replaced with ellipses (...) or represented as values {Val}. For instance, a categorial V-feature may be represented as {V}, and not as an ordered pair $\langle \text{Cat}, \text{V} \rangle$.

As stated in (7c), an unvalued feature $\langle \text{Att}, \emptyset \rangle$ requires replacing the empty set \emptyset with a value. I follow Chomsky (2000; 2001) in adopting the Agree system for feature valuation. This implies accepting the *Activity Condition* in (8).

(8) Activity Condition (Chomsky 2001)

A Goal G is accessible for Agree if G has at least one unvalued feature.

The usual instance of activity/inactivity involves φ -agreement and Case assignment. A DP carrying an unvalued Case feature $\langle \text{Case}, \emptyset \rangle$ is active for φ -agreement, so it may be a Goal for a Probe requiring φ -features. As a consequence of agreement, the Probe values the Case feature of the DP, turning it inactive for further φ -related operations. I take that these mechanisms also hold for left-peripheral features.

(9) The Activity Condition applies for both A and A'-dependencies.

From now on, I will use the Greek letters κ and ω to designate activity-features for A and A'-dependencies, respectively. For concreteness, κ is simply an abbreviation for classic abstract Case, while ω is an attribute that allows a constituent carrying a left-peripheral value (e.g., Wh) to be targeted by a Probe in the C-domain.

According to (9), a wh-pronoun like *who* in (10) requires entering into the derivation with two unvalued features, a Case feature $\langle \kappa, \emptyset \rangle$, allowing it to enter in an Agree relation with matrix T, and an ω -feature $\langle \omega, \emptyset \rangle$, allowing it to move to the specifier position of the interrogative complementizer.

(10) Who seems to be happy?

The derivation of (10) involves four occurrences of *who* valuing features in different positions in the structure according to (11): *who*⁴ is externally merged in a small clause SC carrying both unvalued features $\langle \kappa, \emptyset \rangle$ and $\langle \omega, \emptyset \rangle$; *who*³ is a copy of *who*⁴ generated through successive cyclic movement; *who*² is a copy of *who*³ that A-moved to matrix Spec,T and received nominative Case; and finally, *who*¹ is a copy of *who*² that A'-moved to Spec,C valuing its ω -feature.⁵

- (11) a. $[\text{CP} \text{Who}^1 [\text{TP} \text{who}^2 \text{seems} [\text{TP} \text{who}^3 \text{to be} [\text{SC} \text{who}^4 \text{happy}]]]]$?
 b. $\text{Who}^1_{\{\langle \kappa, \text{NOM} \rangle, \langle \omega, \text{Q} \rangle, \dots\}} \dots \text{who}^2_{\{\langle \kappa, \text{NOM} \rangle, \langle \omega, \emptyset \rangle, \dots\}} \dots \text{who}^3_{\{\langle \kappa, \emptyset \rangle, \langle \omega, \emptyset \rangle, \dots\}} \dots$
 $\text{who}^4_{\{\langle \kappa, \emptyset \rangle, \langle \omega, \emptyset \rangle, \dots\}}$

⁵ Numerical indexes on copies are introduced for expository purposes only.

A principled definition of Non-Distinctiveness must determine the type of linking principle that binds together the copies of *who* in order to form the chain $\text{CH} = \{\text{who}^1, \text{who}^2, \text{who}^3, \text{who}^4\}$. In doing so, it should not introduce markings nor indexes into the representation; the relevant relation must be based on the properties of the elements forming the chain. Therefore, Non-Distinctiveness must follow from some kind of association between the features of the copies of *who*. An inspection of the representation in (11b) reveals that the attributes of the features of the four copies remain constant, while there are some differences regarding their values. Suppose that V^α is the set of morphosyntactic values of a constituent α , and that the sets V^1, V^2, V^3 and V^4 correspond to the values of $\text{who}^1, \text{who}^2, \text{who}^3$ and who^4 , respectively. According to (11b): the sets V^4 and V^3 are identical (cf. (12a)); V^3 is a proper subset of V^2 (cf. (12b)); and V^2 is a proper subset of V^1 (cf. (12c)).

- (12) a. $V^4 = V^3$ (i.e., $\{\dots\} = \{\dots\}$)
- b. $V^3 \subset V^2$ (i.e., $\{\dots\} \subset \{\text{NOM}, \dots\}$)
- c. $V^2 \subset V^1$ (i.e., $\{\text{NOM}, \dots\} \subset \{\text{NOM}, \text{Q}, \dots\}$)

The relations of identity between sets in (12a), and proper inclusion in (12b) and (12c) may be unified as a single type of relation: (improper) inclusion. That is because (i) for every set A identical to a set B, A is a subset of B (i.e., if $A = B$, then $A \subseteq B$), and (ii) for every set A that is a proper subset of a set B, A is a subset of B (i.e., if $A \subset B$, then $A \subseteq B$). In other words, the values of the copies of *who* in (11) are related through inclusion, i.e., the values of who^3 contain the values of who^4 (cf. (13a)); the values of who^2 contain the values of who^3 (cf. (13b)); and the values of who^1 contain the values of who^2 (cf. (13c)).

- (13) a. $V^4 \subseteq V^3$ (i.e., $\{\dots\} \subseteq \{\dots\}$)
- b. $V^3 \subseteq V^2$ (i.e., $\{\dots\} \subseteq \{\text{NOM}, \dots\}$)
- c. $V^2 \subseteq V^1$ (i.e., $\{\text{NOM}, \dots\} \subseteq \{\text{NOM}, \text{Q}, \dots\}$)

Given my assumptions, such an inclusion relation will arise systematically for every new copy of a constituent. Hence, it may be exploited to define Non-Distinctiveness. Call this definition *Inclusion-based Sameness*, or *Inclusion-S*, for short.

(14) **Inclusion-S**

A constituent β is non-distinct from a constituent α if for every value of β there is an identical value in α .

The definition in (14) is a more formal version of a very intuitive idea: if the features of α contain the morphosyntactic information encoded in the features of β , then β is indistinguishable from (part

of) α . Therefore, Inclusion-S involves an asymmetric comparison between two constituents, in which one of them may be underspecified with respect to the other.

Notice that (14) does not introduce any specification of the structural conditions that two constituents must comply to be evaluated as non-distinct. This is undesirable on both empirical and conceptual grounds. In principle, there are two requirements that seem almost ineludible for any two elements forming a movement dependency: (i) they must be in a c-command relation, and (ii) they must be local. By adopting a locality constraint based on *Relativized Minimality* (Rizzi 1990), the conditions in (15) follow.⁶

- (15) Two constituents α and β are part of the same chain if
- a. α c-commands β ,
 - b. β is non-distinct from α (by Inclusion-S),
 - c. there is no δ between α and β such as (i) β is non-distinct from δ , or (ii) δ is non-distinct from α .

These conditions define *chain links*, i.e., they allow recognizing two consecutive members of a movement dependency. Chains of more than two members are obtained by transitivity. That is, if the constituents α and β comply with the conditions in (15), and β and γ also comply with (15), then α , β and γ should be part of the same chain.⁷

A demonstration of the functioning of (15) is in order. Consider once again the passive sentence in (1), repeated for convenience in (16). In this example, both copies of *Cosmo* must form a single chain. As shown in (16c), *Cosmo*¹ has a valued Case feature $\langle \kappa, \text{NOM} \rangle$ while *Cosmo*² carries its unvalued counterpart $\langle \kappa, \emptyset \rangle$.

- (16) a. *Cosmo* was arrested.
- b. $[_{\text{TP}} \text{Cosmo}^1 [_{\text{T}} \text{was} [_{\text{VP}} \text{arrested } \text{Cosmo}^2]]]$.
 - c. $[_{\text{TP}} \text{Cosmo}^1_{\{\langle \kappa, \text{NOM} \rangle, \dots\}} \dots [_{\text{VP}} \dots \text{Cosmo}^2_{\{\langle \kappa, \emptyset \rangle, \dots\}}]]$

⁶ Defining chains in terms of representational conditions as in (15) does not imply a theoretical commitment. An equally valid implementation of Inclusion-S is as an operation based on *Minimal Search* (e.g., inspect the c-command domain of α for a constituent β that is non-distinct from α . Abort the search if a constituent δ is found, such as the set of values V^δ of δ is a proper superset of the set of values V^α of α). There are no meaningful differences between both approaches.

⁷ This aspect of the system makes it *compatible* with Chomsky's (2008) proposal that Non-Distinctiveness is computed within a single phase. Details aside, suppose the constituents α and γ are in distinct but adjacent phases, so the conditions in (15) cannot apply to them at the same time. In this scenario, α and γ may still be part of the same chain under certain conditions. That is, if (i) there is a constituent β in the edge between both phases (i.e., in a position accessible to both α and γ), and (ii) β forms a chain link with both α and γ , then the chain $\text{CH} = \{\alpha, \beta, \gamma\}$ is obtained by transitivity. However, it still remains to be shown on empirical grounds whether a phase-based constraint on Non-Distinctiveness is necessary.

First, *Cosmo*¹ c-commands *Cosmo*², so they comply with (15a). Second, (15b) requires these constituents to be non-distinct according to Inclusion-S. The set of values of *Cosmo*², i.e., {...}, is a subset of the set of values of *Cosmo*¹, i.e., {NOM, ...}, so this requirement is also satisfied. Finally, there is no intervener δ between *Cosmo*¹ and *Cosmo*² in the sense of (15c). Given that the three conditions in (15) are satisfied, the copies of *Cosmo* in (16) form the chain CH = {*Cosmo*¹, *Cosmo*²}, which is the wanted result.

The next example is the active sentence in (17). It has three occurrences of the constituent *Cosmo*: *Cosmo*³ receives accusative Case in its base position,⁸ *Cosmo*² occupies a thematic position and lacks Case, while *Cosmo*¹ is a copy of *Cosmo*² generated through A-movement that receives nominative Case in Spec,T.

- (17) a. Cosmo arrested Cosmo.

- b. $[\text{TP} \text{Cosmo}^1 [\text{T} [\text{VP} \text{Cosmo}^2 [\text{v} \text{arrested Cosmo}^3]]]]$
- c. $[\text{TP} \text{Cosmo}^1_{\{\kappa, \text{NOM}\}, \dots} \dots [\text{VP} \text{Cosmo}^2_{\{\kappa, \emptyset\}, \dots} \dots [\text{v} \dots \text{Cosmo}^3_{\{\kappa, \text{ACC}\}, \dots}]]]$

Here, (i) *Cosmo*¹ c-commands *Cosmo*², (ii) the set of values of *Cosmo*² is a subset of the set of values of *Cosmo*¹ (i.e., {...} \subseteq {NOM, ...}), and (iii) there are no interveners between them, thus *Cosmo*¹ and *Cosmo*² form the chain CH₁ = {*Cosmo*¹, *Cosmo*²}. *Cosmo*³ cannot be considered non-distinct from the elements in CH₁ since it carries a value ACC that is absent in both *Cosmo*¹ and *Cosmo*² (e.g.,

⁸ According to early proposals in the minimalist framework (e.g., Chomsky 1993), an internal argument DP must move over the external argument to receive accusative Case.

- (i) $[\text{TP} \text{T} [\text{VP} \text{IA}_{\kappa, \text{ACC}} [\text{VP} \text{EA}_{\kappa, \emptyset} [\text{v'} \text{v} [\text{VP} \text{V} \text{IA}_{\kappa, \emptyset}]]]]]$

This configuration posits problems for both the Agree system and Inclusion-S. First, the structure in (i) yields a scenario of *defective intervention* (Chomsky 2000), in which the external argument cannot be probed by T because of the presence of an (inactive) intervening DP. Second, this structure leads Inclusion-S to incorrectly form a chain containing occurrences of both the external argument and the internal argument.

There are, at least, two ways of avoiding these issues while maintaining the assumption that the internal argument moves to receive accusative Case. The first alternative is following Harley (1995; 2009), Koizumi (1995), Lasnik (1999), López (2012), among others, in assuming that accusative Case is assigned in a projection below the head that introduces the external argument (cf. (ii)). Alternatively, it may be assumed that the internal argument moves to Spec,v to receive accusative Case before the external argument is introduced (cf. (iii)).

- (ii) $[\text{TP} \text{T} [\text{VoiceP} \text{EA}_{\kappa, \emptyset} [\text{Voice} \text{Voice} [\text{VP} \text{IA}_{\kappa, \text{ACC}} [\text{v'} \text{v} [\text{VP} \text{V} \text{IA}_{\kappa, \emptyset}]]]]]]$
- (iii) $[\text{TP} \text{T} [\text{VP} \text{EA}_{\kappa, \emptyset} [\text{VP} \text{IA}_{\kappa, \text{ACC}} [\text{v'} \text{v} [\text{VP} \text{V} \text{AI}_{\kappa, \emptyset}]]]]]$

$\{\text{ACC}, \dots\} \not\subseteq \{\dots\}$) Therefore, *Cosmo*³ forms a trivial chain of its own $\text{CH}_2 = \{\text{Cosmo}^3\}$. Again, this is the wanted result.

A more complex case is posed by the sentence in (18a). It contains four occurrences of the constituent *Cosmo* (cf. (18b)). In the embedded clause, *Cosmo*⁴ is externally merged in the internal argument position, and *Cosmo*³ is generated from it through A-movement; in the matrix clause, *Cosmo*² occupies the external argument position, while *Cosmo*¹ is generated from *Cosmo*² through A-movement.

- (18) a. Cosmo said that Cosmo was arrested.
b. $[\text{TP} \text{ Cosmo}^1 [\text{T} [\text{VP} \text{ Cosmo}^2 [\text{V} \text{ said that } [\text{TP} \text{ Cosmo}^3 [\text{T} \text{ was } [\text{VP} \text{ arrested } \text{Cosmo}^4]]]]]]]$.
c. $[\text{TP} \text{ Cosmo}^1_{\{\kappa, \text{NOM}\}, \dots} \dots [\text{VP} \text{ Cosmo}^2_{\{\kappa, \emptyset\}, \dots} \dots [\text{TP} \text{ Cosmo}^3_{\{\kappa, \text{NOM}\}, \dots} \dots [\text{VP} \text{ ... } \text{Cosmo}^4_{\{\kappa, \emptyset\}, \dots}]]]]]$

This sentence contains the chains $\text{CH}_1 = \{\text{Cosmo}^1, \text{Cosmo}^2\}$ and $\text{CH}_2 = \{\text{Cosmo}^3, \text{Cosmo}^4\}$, both formed in a similar way to the one in (16). Notice that Inclusion-S would make some erroneous predictions in this case if not combined with a locality constraint like (15c). For example, the set of values of *Cosmo*³ is a subset of the set of values of *Cosmo*¹ (i.e. $\{\text{NOM}, \dots\} \subseteq \{\text{NOM}, \dots\}$), so the chain ${}^*\text{CH} = \{\text{Cosmo}^1, \text{Cosmo}^3\}$ would be, in principle, incorrectly predicted. This unwanted result is avoided by assuming that the calculus of Inclusion-S obeys Relativized Minimality. In other words, *Cosmo*³ cannot be considered non-distinct from *Cosmo*¹ because there is an element being non-distinct from *Cosmo*¹ in the way, i.e., *Cosmo*². The same kind of intervention prevents forming the chain ${}^*\text{CH} = \{\text{Cosmo}^2, \text{Cosmo}^4\}$: even though the values of *Cosmo*² contain the values of *Cosmo*⁴ (i.e., $\{\dots\} \subseteq \{\dots\}$), *Cosmo*⁴ cannot be considered non-distinct from *Cosmo*² since *Cosmo*³ acts as an intervener.

The last example has two A'-dependencies, one in the main clause and the other in the embedded clause, which involve six occurrences of the wh-pronoun *who*: *who*³ and *who*⁶ carry both unvalued Case $\langle \kappa, \emptyset \rangle$ and left-peripheral $\langle \omega, \emptyset \rangle$ features; *who*² and *who*⁵ are copies generated through A-movement; *who*¹ and *who*⁴ are copies generated through A'-movement.

- (19) a. Who said who was arrested?
b. $[\text{CP} \text{ Who}^1 [\text{C} \text{ C} [\text{TP} \text{ who}^2 [\text{T} \text{ T} [\text{VP} \text{ who}^3 [\text{VP} \text{ said } [\text{CP} \text{ who}^4 [\text{C} \text{ C} [\text{TP} \text{ who}^5 [\text{T} \text{ was } [\text{VP} \text{ arrested } \text{who}^6]]]]]]]]]]]$.
c. $[\text{CP} \text{ Who}^1_{\{\kappa, \text{NOM}\}, \langle \omega, \text{Q} \rangle, \dots} \dots [\text{TP} \text{ who}^2_{\{\kappa, \text{NOM}\}, \langle \omega, \emptyset \rangle, \dots} \dots [\text{VP} \text{ who}^3_{\{\kappa, \emptyset\}, \langle \omega, \emptyset \rangle, \dots} \dots [\text{CP} \text{ who}^4_{\{\kappa, \text{NOM}\}, \langle \omega, \text{Q} \rangle, \dots} \dots [\text{TP} \text{ who}^5_{\{\kappa, \text{NOM}\}, \langle \omega, \emptyset \rangle, \dots} \dots [\text{VP} \text{ ... } \text{who}^6_{\{\kappa, \emptyset\}, \langle \omega, \emptyset \rangle, \dots}]]]]]]]$

This sentence contains two chains $\text{CH}_1 = \{\text{who}^1, \text{who}^2, \text{who}^3\}$ and $\text{CH}_2 = \{\text{who}^4, \text{who}^5, \text{who}^6\}$. As already pointed out, the conditions in (15) calculate chain links of two elements, so chains of more than two members must be formed by transitivity. In the case at hand, *who*¹ cannot form a chain

directly with *who*³ since *who*² intervenes between them. However, given that *who*¹ and *who*² comply with the conditions in (15), and *who*² and *who*³ do as well, the chain CH₁ is formed. The same applies for the Non-Distinctiveness relation between *who*⁴ and *who*⁶: it is mediated by *who*⁵.

There is still one important aspect of this system that should be discussed before exploring further empirical consequences of adopting it. Inclusion-S (cf. (14)) and its associated set of conditions (cf. (15)) rely on a *representational* characterization of chains.

- (20) Representational characterization of chains (Rizzi 1986: 66)

Chains are read off from S-structures (and/or other syntactic levels), hence chain formation is a mechanism independent from “move α”, and in principle chains do not necessarily reflect derivational properties.

Adopting this characterization has three main consequences in a minimalist and copy-based theoretical setting. First, given that (20) states that narrow syntactic operations (e.g., Copy, Merge) and chain formation apply independently at distinct computational cycles, it becomes necessary advancing an algorithm of chain recognition that makes no use of narrow syntactic devices but exploits representational properties of phrase markers (e.g., features, geometrical relations between nodes). I take that Inclusion-S in (14) together with the conditions in (15) offer such an algorithm.

Second, given that chains are supposed to be computed over a representation, and there are no levels of representation other than the interface levels (Chomsky 1993), it follows that chains must be computed at the interfaces. Furthermore, since there is no direct link between PF and LF, then chains must be calculated independently and in parallel at both interfaces. According to this, notions such as *Non-Distinctiveness* or *chain* should be regarded as exclusive of the grammatical processes that take place at PF and LF.⁹ This result allows capturing Chomsky's (2001) observation that no narrow syntactic mechanism seems to employ chains.

Third, if chains are read off from syntactic representations, then there is no need to define them as linguistic objects existing separately from a phrase marker. That is, chains are nothing more than an abstract relation holding between some nodes in a syntactic structure, a relation that ultimately denotes a set CH. Therefore, the conditions in (15) must be understood as an intensional definition for this set.

3 Indexical-S and Inclusion-S are not equivalent

As already discussed, the Inclusion-S system computes chains over both interface representations independently and in parallel; such a calculus is based only on information encoded in the phrase marker. Since the Copy operation determines only indirectly the form of chains, there may be “mismatches” on how narrow syntax, PF and LF process movement dependencies. For example, it could be the case that narrow syntax generates a set of copies that is not recognized as a chain in

⁹ See Brody (1998a; b) for related ideas.

one of the interfaces. Conversely, it could also happen that two non-transformationally related constituents comply with Inclusion-S at the interfaces and, therefore, form a chain.

Scenarios like these are not expected under Indexical-S. This marking mechanism establishes a univocal connection between the Copy operation and chain formation during the derivational procedure itself by assigning indexes to the copies. In other words, transformational procedures and chains are inexorably isomorphic under Indexical-S.

This section argues that the “mismatches” predicted by the Inclusion-S system do occur. For conciseness, the discussion focuses on three phenomena that have been accounted for within Copy Theory: (i) Nunes’ (2004) treatment of *wh-copying*, (ii) Barbiers et al.’s (2010) analysis of *non-identical wh-doubling* in Dutch, and (iii) Takahashi & Hulsey’s (2009) extension of Lebeaux’s (1988) ideas on *anti-reconstruction*. These three proposals adopt (implicitly or explicitly) Indexical-S. It is shown that each proposal can be either empirically or conceptually improved by adopting the Inclusion-S system.

3.1 Uniqueness and its apparent exceptions

As mentioned in the introduction, Copy Theory explains the displacement property of language by assuming that only one member of a chain receives pronunciation. Call this general property of chains *Uniqueness*.¹⁰

(21) Uniqueness

Given a chain CH, only one member of CH is pronounced.

In a Late Insertion model as the one adopted here, Uniqueness may be regarded as the result of a natural tension between (i) economy-related considerations on the application of Vocabulary Insertion and (ii) the general conditions governing the recoverability of information. That is, the most economical way of pronouncing a chain is applying Vocabulary Insertion to only one of its members.¹¹

Even though Uniqueness states a crucial property of movement dependencies under Copy Theory, it is usually regarded as a false generalization. This is because some constructions exhibit more than one overt copy. One of these cases is wh-copying, a phenomenon that has been attested in German, Hindi, Romani, and other languages.¹² Sentences involving wh-copying as (22) contain more than one overt occurrence of the same wh-pronoun, despite the fact that they seem to have the same meaning as a regular long distance wh-question (cf. (23)).

¹⁰ The other important generalization about non-trivial chains is that, in most cases, the head of the chain must be the element that receives pronunciation. For discussion of this generalization and its exceptions in terms of a Late Insertion model, see Muñoz Pérez (2016; 2017).

¹¹ See Kennedy (2002) and Landau (2006) for related proposals.

¹² The present discussion focuses on German data. For examples from other languages, see Fanselow (2006) and references therein.

- (22) German (McDaniel 1986: 183)

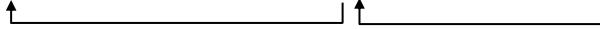
Wen glaubt Hans wen Jakob gesehen hat?
who thinks Hans who Jakob seen has
'Who does Hans think Jakob saw?'

- (23) German

Wen glaubt Hans dass Jakob gesehen hat?
who thinks Hans that Jakob seen has
'Who does Hans think Jakob saw?'

Given the semantic similarity between these sentences, it has become standard to assume that they have the same syntactic structure. In particular, the wh-copying pattern is typically analyzed as involving the overt realization of a copy of the wh-pronoun that has been generated through successive cyclic movement and occupies the specifier position of an embedded complementizer. The relevant representation for the sentence in (22) is sketched in (24).

- (24) [_{CP} Wenⁱ C_{INT} glaubt Hans [_{CP} wenⁱ C_{DECL} Jakob [_{VP} wenⁱ gesehen] hat]]



Under Indexical-S (cf. (3)), the three copies of the wh-pronoun *wen* must necessarily form a single chain CH = {wenⁱ, wenⁱ, wenⁱ}. Given that two members of this chain receive pronunciation, wh-copying must be analyzed as an exception to the Uniqueness property.

Nunes (2004) offers a general account of cases involving pronunciation of more than one copy per chain. His system incorporates the following three main assumptions.

- (25) Nunes' (2004) assumptions

- Chain Reduction (i.e., the operation deleting chain members at PF) is costly.
- Chain Reduction applies until the structure is linearizable according to the *Linear Correspondence Axiom* (LCA) of Kayne (1994). A structure is non-linearizable if the LCA computes two or more non-distinct constituents.
- The LCA "cannot see" inside words.

According to Nunes, whenever there is a case of multiple copy pronunciation it is because one of the copies has been morphologically reanalyzed as part of a bigger word through an application of *Fusion* (Halle and Marantz 1993),¹³ an operation that combines two terminal nodes into one. For

¹³ As stated by Halle & Marantz (1993), two terminals must be in a sisterhood relation to undergo Fusion, which is not the case for a complementizer and its specifier. To solve this, Nunes assumes that successive cyclic movement of wh-pronouns can proceed via adjunction. See Nunes (1998) for relevant discussion.

convenience, I adopt Embick's (2010) definition of this operation, which explicitly states that the features of two syntactic terminals merge into a single set.

- (26) Fusion (Embick 2010: 78)

$$[_x \alpha] \wedge [_y \beta] \rightarrow [_{x/y} \alpha, \beta]$$

where α and β are features of X and Y.

Regarding a structure like (24), Nunes proposes that the intermediate copy of *wen* and the embedded complementizer C_{DECL} undergo Fusion and form a single terminal node [*wen* + C_{DECL}]. Given that Chain Reduction applies to comply with the LCA (cf. (25b)), and the LCA “does not see” the internal structure of words (cf. (25c)), Chain Reduction is not required to delete the copy of *wen* inside [*wen* + C_{DECL}]. Therefore, only the lowest copy of *wen* undergoes Chain Reduction, and the doubling pattern is obtained.

The reanalysis-based account of wh-copying allows deriving two defining properties of the construction. First, given that (by assumption) this morphological reanalysis always affects an embedded complementizer, it follows that a pronoun in its base position cannot be spelled-out in wh-copying constructions. This prediction is borne out. Consider the sentences in (27) and (28). The unacceptability of (28) is due to the presence of an overt occurrence of *wen* within the VP of the embedded clause.

- (27) German (Fanselow & Mahajan 2000: 219)

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?'								

- (28) German (Nunes 2004: 39)

* Wen	glaubt	Hans	wen	Jakob	wen	gesehen	hat?
who	thinks	Hans	who	Jakob	who	seen	has
'Who does Hans think Jakob saw?'							

Second, given that the morphological reanalysis is based on an application of Fusion, a PF operation targeting terminal nodes, it follows that there cannot be cases of multiple copy pronunciation involving full wh-phrases. This prediction also seems to be true. In (29), for example, the wh-phrase *welchen Mann* ‘which man’ cannot be repeated.

An alternative solution is independently advanced by Matushansky (2006), who proposes a morphological operation *m-merger* that creates a head-adjunction kind of configuration from a spec-head relation. For concreteness and simplicity, I adopt the latter option.

- (29) German (Fanselow & Mahajan 2000: 220)

* Welchen Mann	glaubst	Du	welchen Mann	sie	liebt?
which man	believe	you	which man	she	loves
‘Which man do you believe that she loves?’					

The account of wh-copying based on morphological reanalysis can be straightforwardly implemented under the Inclusion-S system with two conceptual advantages: (i) there is no need to treat the phenomenon as an exception to the Uniqueness principle in (21), and (ii) the explanation does not rely on any specific theory of linearization (e.g., the LCA). Consider once again the representation in (24), repeated for convenience in (30a), this time including the featural content of the occurrences of *wen* (cf. (30b)). Basically, *wen*³ receives accusative Case in-situ, *wen*² is a copy of *wen*³ generated through successive cyclic movement, and *wen*¹ values its ω -feature.

- (30) a. $[\text{CP} \text{ Wen}^1 \text{ C}_{\text{INT}} \text{ glaubt Hans} [\text{CP} \text{ wen}^2 \text{ C}_{\text{DECL}} \text{ Jakob} [\text{VP} \text{ wen}^3 \text{ gesehen}] \text{ hat}]]$
b. $[\text{CP} \text{ Wen}^1_{\{\kappa, \text{ACC}\}, \{\omega, \text{Q}\}, \dots} \dots [\text{CP} \text{ wen}^2_{\{\kappa, \text{ACC}\}, \{\omega, \emptyset\}, \dots} \dots [\text{VP} \text{ wen}^3_{\{\kappa, \text{ACC}\}, \{\omega, \emptyset\}, \dots} \dots]]]$

Narrow syntax generates this phrase marker and delivers it to the interfaces for interpretation. At LF, Inclusion-S generates the chain $\text{CH}_{\text{LF}} = \{\text{wen}^1, \text{wen}^2, \text{wen}^3\}$ since (i) *wen*¹ c-commands *wen*², and *wen*² c-commands *wen*³, (ii) the set of values of *wen*¹ contains the values of *wen*² (i.e., $\{\text{ACC}, \dots\} \subseteq \{\text{ACC}, \text{Q}, \dots\}$), and the values of *wen*² contain the values of *wen*³ (i.e., $\{\text{ACC}, \dots\} \subseteq \{\text{ACC}, \dots\}$), and (iii) there are no potential interveners. The chain CH_{LF} determines that this movement dependency is semantically interpreted as regular long distance wh-movement.

Nevertheless, something different happens at PF. There, *wen*² and the embedded complementizer are reanalyzed as a single terminal node through an application of Fusion. According to (26), Fusion combines the features of *wen*² and its neighbor complementizer C_{DECL} into a single syntactic terminal $[\text{wen}^2 + \text{C}_{\text{DECL}}]$. Suppose that C_{DECL} has, at least, a categorial value $\{C\}$. Therefore, the new node carries the Case and left-peripheral features corresponding to *wen*² and the value corresponding to C_{DECL} .¹⁴

- (31) PF representation after Fusion

$$[\text{CP} \text{ Wen}^1_{\{\kappa, \text{ACC}\}, \{\omega, \text{Q}\}, \dots} \dots [\text{CP} [\text{wen}^2 + \text{C}_{\text{DECL}}]_{\{\kappa, \text{ACC}\}, \{\omega, \emptyset\}, C, \dots} \dots [\text{VP} \text{ wen}^3_{\{\kappa, \text{ACC}\}, \{\omega, \emptyset\}, \dots} \dots]]]$$

¹⁴ While I follow Nunes’ implementation based on Fusion for explicitness, it should be noticed that Inclusion-S does not strictly require combining *wen*² and C_{DECL} into a single terminal node to derive the wh-copying pattern. For instance, the morphological reanalysis of *wen*² and C_{DECL} in (31) could generate a structure in which these constituents are distinct terminals under a single node that carries the values $\{\text{ACC}, C, \dots\}$. As far as I can tell, both alternatives are equivalent. For discussion on whether morphological reanalysis in wh-copying constructions relies on Fusion see Kandybowicz (2007) and Saab (2008).

According to Inclusion-S, the copies of *wen* in this representation should form two distinct chains at PF. Since the values of $[wen^2 + C_{DECL}]$ contain the values of *wen*³ (i.e., $\{ACC, \dots\} \subseteq \{ACC, C, \dots\}$), these elements form the chain $CH_{PF2} = \{[wen^2 + C_{DECL}], wen^3\}$. However, the set of values of *wen*¹ does not contain the values of $[wen^2 + C_{DECL}]$ (i.e., $\{ACC, C, \dots\} \not\subseteq \{ACC, Q, \dots\}$), therefore *wen*¹ forms a chain of its own $CH_{PF1} = \{wen^1\}$. Given that *wen*¹ and $[wen^2 + C_{DECL}]$ are heads of distinct chains at PF, both elements receive pronunciation independently, in accordance with Uniqueness in (21).

The account of wh-copying based on Inclusion-S offers an additional empirical advantage over Nunes' (2004) assumptions in (25).¹⁵ Consider a case where a head Y carrying a feature β moves to a head X carrying a feature α . The resulting structure contains two copies of Y, i.e., Y^2 in the original base position and Y^1 adjoined to X.

$$(32) \quad [_{XP} [_{X^0} Y^1_{\{\beta\}} X_{\{\alpha\}}] [_{YP} Y^2_{\{\beta\}} \dots]]$$

In a standard case of head movement, Y^2 should not be pronounced. This follows in both proposals from forming the chain $CH = \{Y^1, Y^2\}$ and pronouncing only the head of the chain, as usual.¹⁶

Suppose that Fusion applies to Y^1 and X after head movement. This would form a node $[Y^1 + X]$ that carries both features α and β .

$$(33) \quad [_{XP} [Y^1 + X]_{\{\alpha, \beta\}} [_{YP} Y^2_{\{\beta\}} \dots]]$$

According to Nunes' system, this sequence of operations should yield the pronunciation of both members of the chain $CH = \{Y^1, Y^2\}$. That is, if Y^1 undergoes Fusion, it becomes inaccessible to the LCA. Therefore, the LCA only “sees” the other member of the chain, i.e., Y^2 . At this point, there is no need to apply Chain Reduction, as the structure is already linearizable.

Under Inclusion-S, Y^2 should remain silent. This is due to the fact that $[Y^1 + X]$ and Y^2 form the chain $CH_{PF} = \{[Y^1 + X], Y^2\}$ since the set of values of Y^2 is a subset of the set of values of $[Y^1 + X]$ (i.e., $\{\beta\} \subseteq \{\alpha, \beta\}$), and, as usual, only the head of the chain receives pronunciation.

This second pattern is the one attested in the literature for every scenario in which head movement feeds Fusion. For instance, Julien (2002) sketches an analysis on these lines for fused markers of polarity and tense in Bambara.

¹⁵ I am grateful to Jonathan Bobaljik (p.c.) for this observation.

¹⁶ I follow Chomsky (1986) in assuming that X has two *segments* in an adjunction configuration as (32). For X to dominate Y^1 , every segment of X must dominate Y^1 . Since this is not the case, the first node in the structure dominating Y^1 is XP. Therefore, Y^1 c-commands its copy Y^2 , so they can form a chain.

- (34) Bambara (Kastenholz 1989, as cited in Julien 2002: 307)

a.	dúnan	yé	dɔ̄lɔ̄	mìn	
	guest	AFF.PAST	millet.beer	drink	
‘The guest drank millet beer.’					
b.	í	mùso	má	jègè	fèere
	my	wife	NEG.PAST	fish	buy
‘My wife did not buy fish.’					

If tense and polarity are distinct heads, a transformational derivation must have caused them to end up in a single syntactic terminal. This derivation is the one already sketched in (32) and (33), i.e., the Polarity head moves to Tense, and Fusion combines them. As (34) shows, such a derivation is supposed to proceed in exactly the same way as predicted by Inclusion-S, i.e., the lowest occurrence of the Polarity head must remain silent. As already discussed, this pattern does not follow from Nunes’ system since it predicts multiple copy pronunciation every time a moved element undergoes Fusion.

To sum up, it has been shown that an account of wh-copying based on morphological reanalysis does not require positing any additional assumptions under Inclusion-S. That is, applying Fusion on an intermediate copy entails the formation of more than one chain at PF. Moreover, unlike Nunes’ (2004) system, Inclusion-S predicts the correct pattern of chain pronunciation in cases of head movement feeding Fusion.

3.2 Partial Copying and chain formation

While there is certain consensus that the proper analysis of wh-copying constructions involves a movement dependency in narrow syntax, there is a much bigger controversy around the pattern exemplified in (35). The obvious difference between this pattern and the one discussed in the previous section is that in this case both wh-pronouns are not the same.

- (35) German (Fanselow & Mahajan 2000: 196)

Was	denkst	Du	wen	sie	gesehen	hat	
	what	think	you	who	she	seen	has
‘Who do you think that she has seen?’							

This phenomenon is known as *wh-scope marking* or *what-construction*. However, I adopt Barbiers et al.’s (2010) terminology and refer to it as non-identical wh-doubling.

There are two main types of analysis for this phenomenon. The first one postulates that there is a *direct dependency* between both wh-elements. According to McDaniel (1989), the wh-element *was ‘what’* is an expletive associated to the wh-pronoun *wen ‘who’*. At LF, the wh-pronoun replaces the

expletive, so the resulting semantic representation is identical to one involving regular long distance wh-movement.

According to the second type of analysis, the wh-element in the matrix clause is related to the whole embedded CP, so there is only an *indirect dependency* between both wh-elements. Dayal (1994; 2000) proposes that *was* is a wh-pronoun functioning as the object of the matrix verb, while the embedded CP is an adjunct. Since *was* is a clausal pronoun, it can refer to the whole embedded CP, which allows explaining the semantics of the construction.

Consider now the patterns of non-identical wh-doubling in Dutch varieties reported by Barbiers et al. (2010). These sentences show three different pronouns participating in the construction: (i) the neuter pronoun *wat* ‘what’, (ii) non-neuter pronoun *wie* ‘who’, and (iii) the relative pronoun *die*.

- (36) *Dutch, dialect from Overijssel* (Barbiers et al. 2010: 2)

Wat denk je wie ik gezien heb?
what think you who I seen have
'Who do you think I saw?'

- (37) *Dutch, dialect from North-Holland* (Barbiers et al. 2010: 2)

Wie denk je die ik gezien heb?
who think you REL.PRON I seen have
'Who do you think I saw?'

- (38) *Dutch, dialect from Overijssel* (Barbiers et al. 2010: 2)

Wat denk je die ik gezien heb?
what think you REL.PRON I seen have
'Who do you think I saw?'

The sentences in (36), (37) and (38) display the orders in which the pronouns can appear in these constructions, i.e., *wat* must precede *wie* or *die*, and *wie* must precede *die*. As shown in (39), any other order is ruled out.

- (39) a. *Wie denk je wat ik gezien heb?
 who think you what I seen have
b. *Die denk je wie ik gezien heb?
 REL.PRON think you who I seen have
c. *Die denk je wat ik gezien heb?
 REL.PRON think you what I seen have

What is particularly interesting about these data is that offering a unified account for them in terms of any of the theoretical alternatives in the literature seems quite difficult. In principle, positing that the left-peripheral wh-pronouns in (36), (37), and (38) are expletives, in line with McDaniel (1989), does not seem to constitute a satisfactory analysis. If both *was* ‘what’ and *wie* ‘who’ are expletives, the restriction on their distribution remains unexplained, i.e. why would *wat* co-appear with *wie* (cf. (36)) and *die* (cf. (38)) while *wie* can co-appear only with *die* (cf. (37))? On the other hand, it does not seem possible either to posit that a wh-pronoun like *wie* ‘who’ in (37) refers to a clause, in line with Dayal (1994, 2000).

Barbiers et al. (2010) advance an account of these puzzling patterns based on two main ingredients. First, they provide an analysis of the pronouns *wat*, *wie* and *die* according to which they are layers of a nominal structure. Since *wat* can appear in a number of syntactic contexts, the authors analyze it as a very impoverished pronominal form corresponding to the most embedded layer in the structure. It is basically an indefinite numeral that carries a quantificational Q-feature. The pronoun *wie* is assumed to contain the properties of *wat* (i.e., the Q-feature) plus φ-features. Finally, *die* contains the properties of *wie* (i.e., the Q-feature and φ-features) plus a definiteness D-feature. The structure is sketched in (40).

$$(40) \quad [_{\text{die}} \text{D} [_{\text{wie}} \varphi [_{\text{wat}} \text{Q}]]]$$

The second ingredient of the analysis is an operation that allows creating movement dependencies in which only a subpart of a constituent moves. Barbiers et al. (2010) call this operation *Partial Copying*. Consider as an example the derivation in (41). In (41a), the structure K contains an occurrence of the pronoun *die*; in (41b), Partial Copying targets the intermediate layer of this pronoun and creates a new copy of it, which corresponds to the pronoun *wie*; finally, *wie* merges into the structure K in (41c).

$$(41) \quad \begin{aligned} \text{a.} \quad & K = [_{\text{XP}} \text{X} \dots [_{\text{YP}} \dots [_{\text{die}} \text{D} [_{\text{wie}} \varphi [_{\text{wat}} \text{Q}]]] \dots]] \\ \text{b.} \quad & K = [_{\text{XP}} \text{X} \dots [_{\text{YP}} \dots [_{\text{die}} \text{D} [_{\text{wie}} \varphi [_{\text{wat}} \text{Q}]]] \dots]] \\ & L = [_{\text{wie}} \varphi [_{\text{wat}} \text{Q}]] \\ \text{c.} \quad & K = [_{\text{XP}} [_{\text{wie}} \varphi [_{\text{wat}} \text{Q}]] [_{\text{X}} \text{X} \dots [_{\text{YP}} \dots [_{\text{die}} \text{D} [_{\text{wie}} \varphi [_{\text{wat}} \text{Q}]]] \dots]]] \end{aligned}$$

For convenience, I reformulate the analysis in (40) and the Partial Copying operation in (41) in more traditional terms. That is, I propose that the pronouns *wat*, *wie* and *die* can be analyzed as sets of features as in (42). Activity features as $\langle \kappa, \emptyset \rangle$ or $\langle \omega, \emptyset \rangle$ are omitted for simplicity.

$$(42) \quad \begin{aligned} \text{a.} \quad & \text{wat} = \{\text{Q}, \dots\} \\ \text{b.} \quad & \text{wie} = \{\text{Q}, \varphi, \dots\} \\ \text{c.} \quad & \text{die} = \{\text{Q}, \varphi, \text{D}, \dots\} \end{aligned}$$

Accordingly, Partial Copying should be understood as an instance of the Copy operation targeting a proper subset of the features of a constituent, i.e., feature movement in the sense of Hiemstra (1986), Cheng (2000) and Sabel (2000).¹⁷ In these terms, a derivation like (41) involves (i) a structure K containing a node with the set of features $\{Q, \varphi, D, \dots\}$ corresponding to *die* as in (43a), (ii) copying from this node the subset of features $\{Q, \varphi, \dots\}$ as in (43b), and (iii) merging this newly generated set into the main structure K as in (43c).

- (43) a. $K = [_{XP} X \dots [_{YP} \dots \{Q, \varphi, D, \dots\} \dots]]$
- b. $K = [_{XP} X \dots [_{YP} \dots \{Q, \varphi, D, \dots\} \dots]]$
 $L = \{Q, \varphi, \dots\}$
- c. $[_{XP} \{Q, \varphi, \dots\} [_{X} X \dots [_{YP} \dots \{Q, \varphi, D, \dots\} \dots]]]$

Given that the sets of features $\{Q, \varphi, \dots\}$ and $\{Q, \varphi, D, \dots\}$ satisfy the Vocabulary Insertion rules of *wie* and *die*, respectively, the relevant nodes are spelled-out as sketched in (44).

- (44) $[_{XP} \text{wie} [_{X} X \dots [_{YP} \dots \text{die} \dots]]]$

Consider now the analysis of (36). The relevant movement dependency involves two steps. First, the wh-pronoun *wie* occupying the complement position of the embedded verb moves through successive cyclic movement to embedded Spec,C. Then, Partial Copying applies to the highest occurrence of *wie*, moving the set of features corresponding to the wh-pronoun *wat* to the left periphery of the sentence.

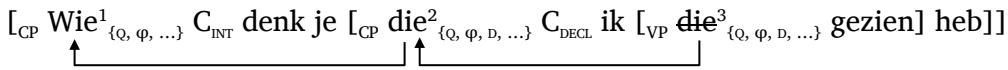
(45) Analysis of (36) in terms of Partial Copying

$[_{CP} \text{Wat}^1_{\{Q, \dots\}} C_{INT} \text{denk je} [_{CP} \text{wie}^2_{\{Q, \varphi, \dots\}} C_{DECL} \text{ik} [_{VP} \text{wie}^3_{\{Q, \varphi, \dots\}} \text{gezien}] \text{heb}]]$

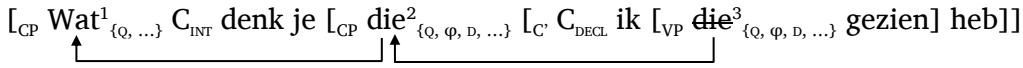
The same kind of derivation applies to the sentences in (37) and (38), as represented in (46) and (47), respectively. In (46), the wh-pronoun *wie* is generated by applying Partial Copying to the copy of *die* in the periphery of the embedded clause. In (47), Partial Copying generates the pronoun *wat* from the features of *die*. The only difference between these two constructions is that φ -features are not copied in the latter case.

¹⁷ Barbiers et al. (2010) argue that applying Partial Copying to the features of a constituent leads to a violation of *Lexical Integrity* (Lapointe 1980), and that this is the reason they adopt a layered structure for pronouns. I do not find this argument compelling. Assuming that pronouns are phrasal as in (40) is a way to allow syntax to manipulate bundles of features as if they were projections. No major advantage regarding Lexical Integrity is obtained by adopting this position.

- (46) Analysis of (37) in terms of Partial Copying



- #### (47) Analysis of (38) in terms of Partial Copying



This type of analysis allows deriving the restrictions on the distribution of wh-pronouns shown in (39). The unacceptable patterns involve a richer pronoun in a higher position that could not have been generated by copying features from the previous one. Since these representations cannot be derived by applying copy operations, they are predicted to be ungrammatical.

- (48) a. $*[_{\text{CP}} \text{wie}_{\{\text{Q}, \varphi, \dots\}} \dots [_{\text{CP}} \text{wat}_{\{\text{Q}, \dots\}} \dots]]$ (cf. (39a))
 b. $*[_{\text{CP}} \text{die}_{\{\text{Q}, \varphi, \text{D}, \dots\}} \dots [_{\text{CP}} \text{wie}_{\{\text{Q}, \varphi, \dots\}} \dots]]$ (cf. (39b))
 c. $*[_{\text{CP}} \text{die}_{\{\text{Q}, \varphi, \text{D}, \dots\}} \dots [_{\text{CP}} \text{wat}_{\{\text{Q}, \dots\}} \dots]]$ (cf. (39c))

While Barbiers et al. (2010) offer an elegant account of the distribution of wh-pronouns in non-identical wh-doubling constructions, their proposal has some flaws that result from their implicit adoption of Indexical-S. According to Indexical-S, partial copies should form chains with their original counterparts. Therefore, despite being distinct lexical items because of the features they carry, the wh-pronouns in (36), (37) and (38) should be regarded as “wh-elements belonging to a single chain that is established in overt syntax” (Barbiers et al. 2010: 25). The chains corresponding to these three sentences are sketched in (49), (50) and (51), respectively.

- (49) a. $[\text{CP} \text{ Wat}^i \dots [\text{CP} \text{ wie}^i \dots [\text{VP} \text{ wie}^i \dots]]]$ (cf. (36))
 b. CH = {watⁱ, wieⁱ, wieⁱ}

- (50) a. $[_{\text{CP}} \text{Wie}^i \dots [_{\text{CP}} \text{die}^i \dots [_{\text{VP}} \text{die}^i \dots]]]$ (cf. (37))
 b. CH = {wieⁱ, dieⁱ, dieⁱ}

- (51) a. $[\text{CP} \text{ Wat}^i \dots [\text{CP} \text{ die}^i \dots [\text{VP} \text{ die}^i \dots]]]$ (cf. (38))
 b. $\text{CH} = \{\text{wat}^i, \text{die}^i, \text{die}^i\}$

Under standard assumptions, these chains should behave exactly in the same way as any chain $CH = \{XP^i, XP^i, XP^i\}$ consisting of three occurrences of the same constituent. This prediction is not borne out. As discussed, chains are supposed to comply with the Uniqueness property in (21) while, on the contrary, non-identical wh-doubling involves pronouncing two wh-elements. To derive this pattern from a single chain, Barbiers et al. adopt Nunes' (2004) account of multiple copy pronunciation (cf. (25)). That is, they propose that the patterns of chain pronunciation in (36), (37)

and (38) follow from an intermediate wh-pronoun being reanalyzed as part of an embedded complementizer.

In principle, this solution seems attractive as Dutch also displays wh-copying patterns.

- (52) *Dutch, dialect from Drenthe* (Barbiers et al. 2010 : 2)

Wie denk je wie ik gezien heb?
who think you who I seen have
'Who do you think I have seen?'

However, there is an additional property distinguishing non-identical wh-doubling from a regular wh-movement dependency. Moving a wh-pronoun across negation is perfectly possible in Dutch (cf. (53)), while non-identical wh-doubling in the same context is unacceptable (cf. (54)). This asymmetry is unexpected if both sentences involve a non-trivial chain connecting a thematic position in the embedded clause with the specifier position of the matrix clause.

- (53) *Dutch* (Barbiers et al. 2010: 40)

Wie denk je niet dat zij uitgenodigd heeft?
who think you not that she invited has
'Who don't you think she has invited?'

- (54) *Dutch* (Barbiers et al. 2010: 40)

* Wat denk je niet wie zij uitgenodigd heeft?
what think you not who her invited has
'Who don't you think she has invited?'

Barbiers et al. do not offer an explanation for the contrast between (53) and (54). Instead, they point out that negation also creates an intervention effect in wh-copying constructions in both Dutch (cf. (55)) and German (cf. (56)). As discussed in the previous section, the standard assumption is that wh-copying is a phonological variant of regular long distance wh-movement. Therefore, negation is not supposed to produce any effect in these constructions.

- (55) *Dutch* (Barbiers et al. 2010: 40)

* Wie denk je niet wie zij uitgenodigd heeft?
who think you not who she invited has
'Who don't you think she has invited?'

- (56) German (Rett 2006: 359)

* Wen glaubst du nicht wen sie liebt?
who think you not who she loves
'Who don't you think she loves?'

Since there seems to be no unified analysis allowing to explain negative intervention effects in all these cases, the authors conclude that the contrast between (53) and (54) is not evidence enough to reject an account of non-identical wh-doubling according to which both overt wh-pronouns are members of the same chain.

It must be noticed, however, that many German speakers accept without problem sentences like (57), where wh-copying across negation is attested.

- (57) German (Pankau 2014: 17)

Wen glaubst du nicht wen sie gesehen hat?
who think you not who she seen has
'Who don't you think she has seen?'

The only way to account for the otherwise contradictory contrast between (56) and (57) is assuming that there are two alternative derivations allowing to generate wh-copying patterns, one that is sensitive to negative intervention and one that is not.

I propose that the derivation that is not sensitive to negative intervention is the one discussed in the previous section, i.e., these are regular wh-movement dependencies in which an intermediate copy is morphologically reanalyzed as part of an embedded complementizer at PF.

On the other hand, I argue that non-identical wh-doubling constructions and cases of wh-copying that are sensitive to negative intervention can be analyzed in a unified way by combining (i) a derivation based on Partial Copying and (ii) the Inclusion-S system. As already discussed, Partial Copying allows explaining in an elegant way the distribution of wh-pronouns in non-identical wh-doubling constructions. However, under Indexical-S, the operation does not derive straightforwardly (i) the fact that two members of the same chain receive pronunciation and (ii) the negative intervention effect. I contend that these properties find a principled explanation under Inclusion-S.

Consider again the structure in (45), which corresponds to the sentence in (36). This representation is generated in narrow syntax and delivered to the interfaces, where chain formation is calculated according to Inclusion-S in (14) and its associated conditions in (15). At PF, *wie*² and *wie*³ form a chain $CH_{PF2} = \{wie^2, wie^3\}$ as they carry the same features (i.e., $\{Q, \varphi, \dots\} \subseteq \{Q, \varphi, \dots\}$); however, since the features of *wie*² are not a subset of the features of *wat*¹ (i.e., $\{Q, \varphi, \dots\} \not\subseteq \{Q, \dots\}$), *wat*¹ forms a trivial chain of its own $CH_{PF1} = \{wat^1\}$. At LF, chain formation works in exactly the same way, i.e., *wie*² and *wie*³ form a chain $CH_{LF2} = \{wie^2, wie^3\}$ while *wat*¹ forms the trivial chain $CH_{LF1} = \{wat^1\}$. These results are summarized in (58).

(58) Chain formation at the interfaces according to Inclusion-S (cf. (36))

- a. $[\text{CP} \text{ wat}^1_{\{\text{Q}, \dots\}} \dots [\text{CP} \text{ wie}^2_{\{\text{Q}, \varphi, \dots\}} \dots [\text{VP} \text{ wie}^3_{\{\text{Q}, \varphi, \dots\}} \dots]]]$
- b. $\text{CH}_{\text{PH1}} = \{\text{wat}^1\}; \text{CH}_{\text{PF2}} = \{\text{wie}^2, \text{wie}^3\}$
- c. $\text{CH}_{\text{LF1}} = \{\text{wat}^1\}; \text{CH}_{\text{LF2}} = \{\text{wie}^2, \text{wie}^3\}$

In more explicit terms, Inclusion-S predicts that constituents that are transformationally related through Partial Copying must form distinct chains. That is, this system states that two constituents α and β are non-distinct if (i) α c-commands β and (ii) α contains the information encoded in β . However, Partial Copying systematically creates configurations where the features of the c-commanding element are contained in the lower constituent. Therefore, partial copies are always computed as distinct elements at the interfaces. This is attested once again in the derivations in (46) and (47), which correspond to the sentences in (37) and (38), respectively. In the former, the features of *die*² are not a subset of the features of *wie*¹ (i.e., $\{\text{Q}, \varphi, \text{D}, \dots\} \not\subseteq \{\text{Q}, \varphi, \dots\}$), so they form two separate chains as shown in (59). In the latter, the features of *die*² are not a subset of the features of *wat*¹ (i.e., $\{\text{Q}, \varphi, \text{D}, \dots\} \not\subseteq \{\text{Q}, \dots\}$), so they form two chains at each interface as shown in (60).

(59) Chain formation at the interfaces according to Inclusion-S (cf. (40))

- a. $[\text{CP} \text{ wie}^1_{\{\text{Q}, \varphi, \dots\}} \dots [\text{CP} \text{ die}^2_{\{\text{Q}, \varphi, \text{D}, \dots\}} \dots [\text{VP} \text{ die}^3_{\{\text{Q}, \varphi, \text{D}, \dots\}} \dots]]]$
- b. $\text{C}_{\text{PF1}} = \{\text{wie}^1\}; \text{C}_{\text{PF2}} = \{\text{die}^2, \text{die}^3\}$
- c. $\text{C}_{\text{LF1}} = \{\text{wie}^1\}; \text{C}_{\text{LF2}} = \{\text{die}^2, \text{die}^3\}$

(60) Chain formation at the interfaces according to Inclusion-S (cf. (41))

- a. $[\text{CP} \text{ wat}^1_{\{\text{Q}, \dots\}} \dots [\text{CP} \text{ die}^2_{\{\text{Q}, \varphi, \text{D}, \dots\}} \dots [\text{VP} \text{ die}^3_{\{\text{Q}, \varphi, \text{D}, \dots\}} \dots]]]$
- b. $\text{CH}_{\text{PF1}} = \{\text{wat}^1\}; \text{CH}_{\text{PF2}} = \{\text{die}^2, \text{die}^3\}$
- c. $\text{CH}_{\text{LF1}} = \{\text{wat}^1\}; \text{CH}_{\text{LF2}} = \{\text{die}^2, \text{die}^3\}$

Extending Partial Copying to capture wh-copying patterns is conceptually simple.¹⁸ Assume a feature $\{F\}$ that is common to *wat*, *wie* and *die*. Assume also that the Vocabulary Insertion rules for these pronouns do not refer to $\{F\}$, i.e. their corresponding syntactic terminals receive the same phonological exponent no matter $\{F\}$ is present or not. Now, consider the derivation in (61). First, the whole set of features $\{\text{Q}, \varphi, F, \dots\}$ corresponding to the wh-pronoun *wie* moves from the complement position of the embedded verb to embedded Spec,C through successive cyclic movement. Then, Partial Copying applies to this set neglecting its $\{F\}$ feature, so the set $\{\text{Q}, \varphi, \dots\}$ is generated and merged into matrix Spec,C. Since this newly formed set of features also corresponds to the phonological exponent *wie*, a third occurrence of this pronoun is generated in the structure.

¹⁸ See Hiemstra (1986) for a similar account of wh-copying patterns.

(61) Analysis of (52) in terms of Partial Copying

$\llbracket_{\text{CP}} \text{Wie}^1_{\{Q, \varphi, \dots\}} C_{\text{INT}} \text{ denk je } \llbracket_{\text{CP}} \text{wie}^2_{\{Q, \varphi, F, \dots\}} C_{\text{DECL}} \text{ ik } \llbracket_{\text{VP}} \text{wie}^3_{\{Q, \varphi, F, \dots\}} \text{gezien} \rrbracket \text{ heb} \rrbracket$

Once again, applying Partial Copying entails forming more than one chain at the interfaces, i.e., the features of *wie*² are not a subset of the features of *wie*¹ (i.e., $\{Q, \varphi, F, \dots\} \not\subset \{Q, \varphi, \dots\}$).

(62) Chain formation at the interfaces according to Inclusion-S (cf. (52))

- a. $\llbracket_{\text{CP}} \text{wie}^1_{\{Q, \varphi, \dots\}} \dots \llbracket_{\text{CP}} \text{wie}^2_{\{Q, \varphi, F, \dots\}} \dots \llbracket_{\text{VP}} \text{wie}^3_{\{Q, \varphi, F, \dots\}} \dots \rrbracket \rrbracket$
- b. $\text{CH}_{\text{PF}1} = \{\text{wie}^1\}; \text{CH}_{\text{PF}2} = \{\text{wie}^2, \text{wie}^3\}$
- c. $\text{CH}_{\text{LF}1} = \{\text{wie}^1\}; \text{CH}_{\text{LF}2} = \{\text{wie}^2, \text{wie}^3\}$

Given that $\text{CH}_{\text{PF}1}$ and $\text{CH}_{\text{PF}2}$ in (58), (59), (60) and (62) are separate chains at PF, both of them should receive pronunciation independently according to Uniqueness (cf. (21)). That is, there is no need to assume that a morphological reanalysis operation applies in these cases. Instead, Partial Copying entails doubling under Inclusion-S.

Consider now the LF chains $\text{CH}_{\text{LF}1}$ and $\text{CH}_{\text{LF}2}$ in (58), (59), (60) and (62). In each of these cases, the trivial chain $\text{CH}_{\text{LF}1}$ lacks a thematic interpretation, but its only member satisfies the formal requirements of the interrogative complementizer, i.e., it functions as an expletive. On the other hand, the chain $\text{CH}_{\text{LF}2}$ does receive a θ-role as one of its members occupies a thematic position; however, none of the wh-elements pertaining to $\text{CH}_{\text{LF}2}$ is in a spec-head configuration with C_{INT} . Since there is no overt syntactic relation between $\text{CH}_{\text{LF}2}$ and the interrogative C-domain, this chain must be interpreted by appealing to the mechanisms that allow licensing wh-in-situ.

As usually assumed (e.g., Lasnik & Saito 1992; Beck 1996; 2006; Pesetsky 2000; Kratzer & Shimoyama 2002; i.a.), in-situ wh-phrases in, for example, multiple wh-questions must be licensed by establishing a covert dependency with the interrogative C-domain.¹⁹ The German sentence in (63), for instance, is supposed to involve an abstract relation linking the wh-pronoun *wo* ‘where’ and the interrogative complementizer C_{INT} that is represented in (64).

(63) German (Beck 1996: 4)

Wen hat Luise wo gesehen?
 whom has Luise where seen
 'Where did Luise see whom?'

(64) Wen C_{INT} hat Luise **wo** gesehen?

$\underline{\hspace{2cm}}$

¹⁹ I remain agnostic whether this covert dependency involves LF-movement (Lasnik & Saito 1992), feature movement (Pesetsky 2000), or a focus-related computation (Beck 2006).

In a similar fashion, the head of each of the wh-chains $\text{CH}_{\text{LF}2}$ in (58), (59), (60) and (62) must be licensed by establishing a covert dependency with the interrogative complementizer in the matrix clause. In the representation in (65), for instance, this relation holds between C_{INT} and wie^2 .

- (65) $[\text{CP} \text{ Wat}^1 C_{\text{INT}} \text{ denk je } [\text{CP} \underline{\text{wie}^2} \text{ ik } \underline{\text{wie}^3} \text{ gezien heb}]]?$ (cf. (36))

As Beck (1996) points out, this type of dependency may be disrupted by an intervening negative element. As shown in (66), the presence of *niet* ‘not’ in between the interrogative complementizer C_{INT} and wie^2 prevents the licensing of the wh-chain $\text{CH}_{\text{LF}2} = \{\text{wie}^2, \text{wie}^3\}$.

- (66) $*[\text{SC} \text{ Wat}^1 C_{\text{INT}} \text{ denk je } \underline{\text{niet}} [\text{SC} \underline{\text{wie}^2} \text{ zij } \underline{\text{wie}^3} \text{ uitgenodigd heeft}]]$ (cf. (54))

The effect in (66) belongs to a natural class of phenomena together with instances of intervention triggered by negation and some other quantificational elements at LF. For instance, the multiple wh-question in (63) is acceptable as long as there is no negative element as *niemand* ‘nobody’ between *wo* ‘where’ and the left periphery of the sentence (cf. (67a)). Negation, however, does not disrupt overt movement, so *wo* can move over *niemand* as in (67b), yielding an acceptable result.

- (67) German (Beck 1996; 2006)

- a. *Wen hat niemand wo gesehen?
whom has nobody where seen
'Where did nobody see whom?'
- b. Wen hat wo niemand gesehen?
whom has where nobody seen
'Where did nobody see whom?'

Similar intervention patterns are attested in many languages. For instance, French allows moving a wh-pronoun to the left periphery (cf. (68a)) or interpreting it in-situ (cf. (68b)).

- (68) French (Bošković 2000)

- a. Qui as-tu vu?
whom have-you seen
- b. Tu as vu qui?
you have seen whom
'Whom have you seen?'

However, applying overt wh-movement seems to be the only available option if a negative element appears between the wh-pronoun and the left periphery.

- (69) French (Bošković 2000)

- a. Qui'est-ce que Jean ne mange pas?
What that Jean NEG eats not
- b. *Jean ne mange pas quoi?
Jean NEG eats not what

'What doesn't John eat?'

Discussing potential accounts of LF intervention effects goes beyond the aims of this paper. My purpose is simply to show that adopting Inclusion-S together with Partial Copying yields configurations in which one of the resulting chains must be licensed through covert dependencies, and that independent phenomena for which these covert dependencies are originally postulated display the same type of negative intervention.

In sum, the distribution patterns of wh-pronouns in non-identical wh-doubling constructions in Dutch is elegantly explained by appealing to Partial Copying, as proposed by Barbiers et al. (2010). However, if Indexical-S is adopted (i) additional assumptions are required to explain the doubling pattern, and (ii) the asymmetry between non-identical wh-doubling and regular wh-movement regarding negative intervention remains mysterious. Under Inclusion-S, on the contrary, both traits follow straightforwardly from partial copies forming separate chains at the interfaces.

3.3 There is no need for Late Merger

Reconstruction has been an important source of evidence for Copy Theory. Assuming that movement involves two (or more) occurrences of the same constituent allows explaining the unacceptability of sentences like (70) as violations of Condition C.

- (70) a. *Which argument that *Cosmo_i* is a genius did *he_i* believe?
b. * [_{DP} Which argument [_{CP} that *Cosmo₁* is a genius]] did *he₁* believe [_{DP} ~~which argument~~ [_{CP} ~~that Cosmo₁ is a genius]]?~~]

There are, however, some cases that do not follow straightforwardly from Copy Theory. For instance, if the complement CP *that Cosmo is a genius* in (70) were replaced by an adjunct CP also containing the R-expression *Cosmo*, the sentence would become acceptable.

- (71) [_{DP} Which argument [_{ADJ} that *Cosmo₁* made]] did *he₁* believe?

According to Lebeaux (1988), cases as (71) do not involve true violations of Condition C. I will refer to his approach as the *Lebeauxian Approach to Anti-Reconstruction* (LATAR).

(72) LATAR

Apparent violations of Condition C follow from the absence of the constituent containing the relevant R-expression in some members of the movement chain.

LATAR involves assuming that the adjunct containing the violating R-expression appears only in the overt member of the chain. Since the pronoun does not c-command the R-expression it binds, no violation of Condition C may arise.

(73) [_{DP} Which argument [_{ADJ} that *Cosmo*₁ made]] did *he*₁ believe [_{DP} ~~which argument~~]?

More recently, LATAR has been extended to capture phenomena involving A-movement (Takahashi & Hulsey 2009). For instance, the sentence in (74) would be expected to be unacceptable due to a Condition C violation under standard assumptions.

(74) [_{DP} The claim that *Cosmo*₁ was asleep] seems to *him*₁ to be correct.

Adapting Lebeaux's proposal, Takahashi & Hulsey (2009) argue that the complement NP of the determiner *the*, i.e., the constituent containing the R-expression *Cosmo*, appears only in the head of the chain.²⁰

(75) [_{DP} The [_{NP} claim that *Cosmo*₁ was asleep]] seems to *him*₁ to be [_{DP} ~~the~~] correct.

The same kind of analysis may be advanced for sentences such as (76). Here, the grammatical subject is interpreted as the logical subject of the predicate *intrusion* (i.e., the picture is an intrusion), so this DP must have been generated very low in the structure of the sentence. Notice that (i) the DP *the president* inside the subject can be coreferential with the pronoun *him*, so *him* should not be able to c-command *the president*, but (ii) the quantifier *every man* can bind the pronoun *his* inside the subject, so the quantifier must c-command this pronoun.

(76) [_{DP} *His*₁ picture of the president₂] seemed to every man₁ to be seen by him₂ to be an intrusion.

This deceptive contradiction may be accounted under LATAR. What is generated near the predicate *intrusion* is a bare possessive determiner that lacks an NP. This element moves through successive

²⁰ Convergent evidence that bare D-heads can occupy the tail position in a movement chain comes from extraction asymmetries in preposition-stranding contexts in English. See Stanton (2016) for relevant discussion.

cyclic A-movement and reaches a position where it can be bound by *every man*. Finally, its complement NP appears only in the head of the movement chain.²¹

- (77) [_{DP} His₁ picture of the president₂] seemed to [every man]₁ [_{DP} his₁] to be seen by him₂ [_{DP} his₁] to be [_{DP} his₁] an intrusion.

If this approach to *anti-reconstruction* is on the right track, then these movement dependencies involve chains like the ones depicted in (78), in which chain members are not isomorphic constituents.

- (78) a CH = {[_{DP} which argument that Cosmo made], [_{DP} which argument]} (cf. (71))
 b. CH = {[_{DP} the claim that Cosmo was asleep], [_{DP} the]} (cf. (74))
 c. CH = {[_{DP} his picture of the president], [_{DP} his], [_{DP} his], [_{DP} his]} (cf. (76))

Compare the way Indexical-S and Inclusion-S generate these chains. Under Indexical-S (cf. (3)), two (or more) constituents form a chain only if they receive the same index through the Copy operation. Therefore, to explain the differences between members of the same chain in (78) it would be necessary (i) generating two (or more) strictly identical copies with the same index, and then (ii) applying an additional operation on the higher copy to introduce the constituent containing the relevant R-expression. Consider the following sample derivation. A constituent αP is generated in a position where it is c-commanded by a pronoun; since αP does not contain an R-expression, Condition C is respected (cf. (79a)). Later in the derivation, αP moves to a position where it c-commands the pronoun; both copies of αP share the same index (cf. (79b)). As a third step, a βP containing an R-expression is inserted into αP as in (79c). At this point, the pronoun and the R-expression can be coreferential, and since both occurrences of αP share the same index, they form a chain.

- (79) a. [_{XP} X ... [_{YP} Pronoun₁ ... [_{ZP} ... [_{αP} α]ⁱ]]]]
 b. [_{XP} [_{αP} α]ⁱ [_{X'} X ... [_{YP} Pronoun₁ ... [_{ZP} ... [_{αP} α]ⁱ]]]]]
 c. [_{XP} [_{αP} α [_{βP} R-expression₁]]ⁱ [_{X'} X ... [_{YP} Pronoun₁ ... [_{ZP} ... [_{αP} α]ⁱ]]]]]

The derivational step in (79c) corresponds to the operations that are called *Late Merger* (Lebeaux 1988) and *Wholesale Late Merger* (Takahashi & Hulsey 2009). The difference between them is that Late Merger is supposed to be restricted to adjuncts (i.e., βP in (79c) must be an adjunct), while Wholesale Late Merger extends the empirical domain of the former to complements (i.e., βP in (79c) may be either an adjunct or a complement). In other words, both operations involve merging a constituent *countercyclically* inside a derived specifier. Thus, it may be concluded that implementing LATAR under Indexical-S implies abandoning strict cyclicity.

²¹ An alternative derivation may imply introducing the complement NP in any position above the pronoun *him* and then moving up the whole DP. There is no significant difference between both options.

As known, cyclicity is a theoretical desideratum in generative syntax since at least Chomsky (1965), and it is encoded in several derivational restrictions, such as the Extension Condition.

- (80) Extension Condition (Chomsky 1993)
 Syntactic operations must extend the tree at the root.

Therefore, if an extensionally equivalent and cyclicity-respecting implementation of LATAR is offered, it should be preferred on conceptual grounds.

The definition of Non-Distinctiveness based on Inclusion-S has two important traits that allow offering a cyclic implementation of LATAR. First, Inclusion-S does not require structural isomorphism between chain members; it only states a condition on their morphosyntactic values. Second, Inclusion-S does not require chain members to be related through the Copy operation.

Consider the cases involving anti-reconstruction in A-movement in (74) and (76). The sentence in (74) is repeated for convenience in (81) with a description of the features of the relevant constituents. A bare determiner $D^{\min/\max}$ is base-generated low down in the structure inside a small clause. This constituent does not carry a full set of valued φ -features as some of these are only inherently valued in the NP domain (e.g., Number, Gender). After T is merged, the Spec,T position is filled with a base-generated full-DP with a complete set of valued φ -features. This DP agrees with T and receives nominative Case.²²

- (81) $[_{DP} \text{The } [_{NP} \text{claim that } \text{Cosmo}_1 \text{ was asleep}]]_{\{<\kappa, \text{NOM}>, <\text{Num}, \text{SG}>, \dots\}}$ seems to him₁ to be correct $[_{DP} \text{the}]_{\{<\kappa, \emptyset>, <\text{Num}, \emptyset>, \dots\}}$.

The base-generated full-DP and the bare determiner comply with the conditions to form a chain according to Inclusion-S. That is, (i) they are in a c-command relation, (ii) the values of the full-DP contain the values of $D^{\min/\max}$ (i.e., $\{\dots\} \subseteq \{\text{NOM, SG, ...}\}$), and (iii) there are no potential interveners between them. Therefore, these two elements form the chain in (78b) despite the fact they are not transformationally related.

Consider now the sentence in (76), repeated for convenience in (82). Here, a bare possessive determiner merged low down in the structure undergoes successive cyclic A-movement to a position just below the quantifier *every man*. There, it gets bound by the quantifier. Later in the derivation, a full-DP is externally merged in the Spec,T position, receiving nominative Case.²³

²² While the configuration in (81) seems to require *Upward Agree* (Zijlstra 2012) between T and the subject DP, it could also be assumed that (i) the full-DP is merged in a position immediately below T, and (ii) it is attracted to Spec,T through traditional *Downward Agree*.

²³ There are reasons to believe that *his* is not the head of this DP, but the specifier of a null determiner, e.g., $[_{DP} \text{his } [_D \text{ every } [_{NP} \text{idea}]]]$. Translating this analysis to the sentence in (76) is fairly simple within Inclusion-S: a null $D^{\min/\max}$ and the DP *his picture of the president* form a chain in exactly the same way as in (82). However,

- (82) [_{DP} His₁ picture of the president₂]_{<κ,NOM>, <Num,SG>, ...} seemed to every man₁
 [_{DP} his₁]_{<κ,∅>, <Num,∅>, ...} to be seen by him₂ [_{DP} his₁]_{<κ,∅>, <Num,∅>, ...} to be
 [_{DP} his₁]_{<κ,∅>, <Num,∅>, ...} an intrusion.

According to Inclusion-S, the full-DP and the copies of the possessive determiner form the chain in (79c).

The sentence in (71), repeated for convenience in (83), involves an additional derivational step. Here, the DP *which argument* is externally merged in the complement position of the verb *believe*, where it receives accusative Case. Almost at the end of the derivation, a new DP *which argument that Cosmo made* is base-generated in the specifier position of the interrogative complementizer. Since this new DP carries an active ω -feature, it enters in an Agree relation with C_{INT} . Notice, however, that the DP still lacks a value for its κ -feature.

- (83) [_{CP} [_{DP} Which argument [_{ADJ} that Cosmo₁ made]]]_{<κ,∅>, <ω,Q>, ...} [_C C_{INT} did he believe [_{DP} which argument]_{{<κ,ACC>, <ω,∅>, ...}]]]]]}

To value its κ -feature, the higher DP probes the structure for a matching Goal. By hypothesis, it looks for an active element matching both κ and ω -features, so the DP *which argument* is the closest and only available candidate. Both DPs agree, and the κ -feature in the higher DP is valued, delivering the representation in (84) to the interfaces.

- (84) [_{CP} [_{DP} Which argument [_{ADJ} that Cosmo₁ made]]]_{<κ,ACC>, <ω,Q>, ...} [_C C_{INT} did he believe [_{DP} which argument]_{{<κ,ACC>, <ω,Q>, ...}]]]]]}

According to Inclusion-S, these two non-transformationally related DPs form the chain in (79a).

As seen, Inclusion-S allows generating non-isomorphic chains without assuming any countercyclical operation. Moreover, the principles restricting this implementation of LATAR are no different from the ones assumed by Takahashi & Hulsey (2009) for Wholesale Late Merger. These authors identify two types of constraint: (i) Agreement/Case, and (ii) semantic interpretability. Regarding the former, it has been already observed that a bare $D^{\min/\max}$ does not carry a complete set of φ -features. Therefore, this type of element is unable to value the φ -features of a Probe and receive Case. It follows, then, that a full-DP should be introduced in the derivation as late as in the specifier position of the relevant Case assigner. From this, prediction in (85) follows:

- (85) Only caseless chain members (i.e., “traces” of A-movement) may be $D^{\min/\max}$.

an account based on Wholesale Late Merger would have to assume a more complex derivation, e.g., that both the specifier *his* and the complement NP may be introduced countercyclically.

Regarding interpretability, Fox (2002) proposes that *Trace Conversion* applies at LF to the tail of an A'-movement dependency to obtain a valid operator-variable relation under Copy Theory.

(86) Trace Conversion (Fox 2002: 67)

- a. Variable Insertion: (Det) Pred → (Det [Pred $\lambda y(y=x)$])
- b. Determiner Replacement: (Det) [Pred $\lambda y(y=x)$] → the [Pred $\lambda y(y=x)$]

This rule transforms a wh-phrase into a definite description with anaphoric value.²⁴ The subpart in (86a) introduces a predicate $\langle e, t \rangle$ that functions as a variable and is interpreted compositionally with a complete nominal predicate, i.e., another $\langle e, t \rangle$ expression, through *Predicate Modification* (Heim & Kratzer 1998). Incomplete nominal predicates (i.e., nouns lacking some argument) are not $\langle e, t \rangle$ expressions, so they are not proper inputs for Trace Conversion. Therefore, the tail of an A'-movement dependency must always contain a noun with all its arguments. Consequently, the statement in (87) follows.

(87) Anti-reconstruction effects in A'-movement of DPs are restricted to non-arguments of nominal predicates.

Adopting Trace Conversion also allows ruling an unwanted consequence of base-generating chain members under Inclusion-S. Consider the structure in (88). If a DP as *which girl* is externally merged in a θ-position, and a different DP *which woman* is base-generated in Spec,C, they would be expected to form a chain.

(88) [CP [DP Which woman]_{<κ,ACC>, <ω,Q>, ...} [C_{INT} C_{INT} did Elaine meet [DP ~~which girl~~_{<κ,ACC>, <ω,Q>, ...}]]]

This unwanted result is ruled out as the output of applying Trace Conversion would yield the uninterpretable operator-variable dependency in (89). As the pair in (90) shows, lexical identity on the NP is a requisite for a definite expression to receive anaphoric interpretation.²⁵

²⁴ A reviewer observes that if Trace Conversion is an operation that truly replaces the wh-determiner for a definite determiner, then the inclusion relation between the features of the operator and the variable should not hold anymore at LF. There are two alternatives to deal with this issue. The first one is assuming that at some point the relation between the operator and its variable is purely anaphoric, i.e., Trace Conversion transforms a Non-Distinctiveness relation into a bound anaphora relation. The second option is taking Trace Conversion to be an interpretative rule that does not modify the phrase marker. For this alternative, see Fox (2003) and references therein.

²⁵ Base-generating constituents in Spec,C may also lead to predict patterns like the ones in (39), at least in an underlying representation. This unwanted result may be avoided if, as Fox (2002) proposes, low copies in A'-

- (89) *Which woman λx . Elaine met the girl x

- (90) a. The neighbor of [every comedian]₁ always takes advantage of [the comedian]₁.
 b. *The neighbor of [every comedian]₁ always takes advantage of [the postman]₁.

The unacceptability of (89) shows that there are semantic mechanisms imposing identity conditions on chain members. Importantly, these mechanisms do not seem to depend on any narrow syntactic device (e.g., the Copy operation). Presumably, some other independently motivated principles also introduce constraints on the properties of unpronounced chain members.²⁶

To sum up, Inclusion-S offers a straightforward way of capturing anti-reconstruction effects under LATAR. Moreover, it allows getting rid of countercyclical operations as Late Merger and Wholesale Late Merger, a very welcome result from a conceptual point of view.

4 Concluding remarks

Copy Theory is based on the idea that elements forming a chain are non-distinct. In this paper, I offered a definition of the Non-Distinctiveness relation based on the featural content of constituents in a phrase marker: Inclusion-S. According to it, two constituents are non-distinct for the purposes of chain formation if the morphosyntactic properties of one of them constitute a subset of the morphosyntactic properties of the other. This condition is part of a representational algorithm of chain recognition that applies independently and in parallel at both interface levels.

Apart from offering a principled definition of Non-Distinctiveness, Inclusion-S introduces a number of empirical and conceptual advantages over a mere indexing mechanism. As discussed, it allows understanding wh-copying as a phenomenon in which a morphological reanalysis operation affects how chains are computed at PF. That is, LF takes a set of occurrences of a wh-pronoun to

chains are interpreted as anaphoric definite expressions. Suppose the following structure, in which *wie* and *wat* are not transformationally related.

- (i) $[\text{CP } \text{wie}_{\{Q, \phi, \dots\}} \dots [\text{CP } \text{wat}_{\{Q, \dots\}} \dots]]$ (cf. (39a))

These pronouns would form the chain CH = (wie, wat) under Inclusion-S. However, once Trace Conversion transforms the wh-pronoun *wat* ‘what’ into a definite description, the resulting representation would contain an uninterpretable operator-variable dependency paraphrasable as **which is the person x such that you think I saw the thing x*.

²⁶ Stanton (2016), for instance, proposes that there is a preference for unpronounced members of A-chains to be bare determiners D^{min/max} instead of full-DPs. While she interprets this preference as a tendency for applying Wholesale Late Merger as late as possible in the derivation, it can also be understood as a result of representational economy.

form a single chain, while the same elements form two (or more) chains at PF, which derives the doubling pattern.

Something similar has been argued to happen in non-identical wh-doubling constructions in Dutch. In this case, however, both interfaces form two chains from a set of wh-elements. The distribution of the pronouns *wat*, *wie* and *die* indicates that the doubling pattern is attested in those cases in which the overt pronouns cannot form a chain at PF according to Inclusion-S. Moreover, these constructions display an intervention effect triggered by negation that shows that the pronouns do not form a chain at LF either.

Finally, anti-reconstruction phenomena have been used to show that non-isomorphic constituents may be part of the same chain in certain contexts. This follows from Inclusion-S, as it predicts that two elements may form a chain even if they are not derivationally related through the Copy operation.

References

- Adger, David. 2010. A minimalist theory of feature structures. In Anna Kibort & Greville Corbett (eds.), *Features*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199577743.003.0008>.
- Adger, David & Peter Svenonius. 2011. Features in minimalist syntax. In Cedric Boeckx (ed.), *The Oxford handbook of linguistic minimalism*. Oxford, OUP.
<https://doi.org/10.1093/oxfordhb/9780199549368.013.0002>.
- Barbiers, Sjef, Olaf Koeneman & Marika Lekakou. 2010. Syntactic doubling and the structure of wh-chains. *Journal of Linguistics* 46. 1-46. <https://doi.org/10.1017/S002226709990181>.
- Beck, Sigrid. 1996. Quantified structures as barriers for LF-movement. *Natural Language Semantics* 4. 1-56. <https://doi.org/10.1007/BF00263536>.
- Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14. 1-56. <https://doi.org/10.1007/s11050-005-4532-y>.
- Bošković, Željko. 2000. Sometimes in Spec,CP, sometimes in-situ. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press.
- Bošković, Željko & Jairo Nunes. 2007. The copy theory of movement: a view from PF. In Norbert Corver & Jairo Nunes (eds), *The copy theory of movement*. Amsterdam: John Benjamins.
<https://doi.org/10.1075/la.107.03bos>.
- Brody, Michael .1998a. Projection and phrase structure. *Linguistic Inquiry* 29. 367-398.
<https://doi.org/10.1162/002438998553798>.
- Brody, Michael, 1998b. The minimalist program and a perfect syntax. *Mind and Language* 13. 205-214. <https://doi.org/10.1111/1468-0017.00074>.
- Cheng, Lisa. 2000. Moving just the feature. In Uli Lutz, Gereon Müller & Arnim von Stechow (Eds.), *Wh-scope marking*. Amsterdam: John Benjamins.

- Chomsky, Noam. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. (1986). *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. (1993). A minimalist program for linguistic theory. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.
- Chomsky, Noam. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, Juan Uriagereka & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: MIT Press
- Chomsky, Noam. 2001. Derivation by Phase. Michael Kenstowicz (ed.), *Ken Hale: A life in language*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On Phases. In Robert Freidin, Carlos Peregrín Otero & María Luisa Zubizarreta (eds), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9780262062787.003.0007>.
- Dayal, Veneeta. 1994. Scope Marking as indirect wh-dependency. *Natural Language Semantics* 2. 137–170.
- Dayal, Veneeta. 2000. Scope marking: cross-linguistic variation in indirect dependency. In Uli Lutz, Gereon Müller & Arnim von Stechow (eds.), *Wh-scope marking*. Amsterdam: John Benjamins.
- Donati, Caterina. 2003. Merge copy. In Kerstin Schwabe & Susanne Winkler (eds.), *The interfaces: Deriving and Interpreting omitted structures*. Amsterdam: John Benjamins.
- Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.
- Fanselow, Gisbert. 2006. Partial wh-movement. In Martin Everaert & Henk Van Riemsdijk (eds.), *The Blackwell companion to syntax, volume III*. Malden, CA: Blackwell Publishing.
- Fanselow, Gisbert & Anoop Mahajan. 2000. Towards a minimalist theory of wh-expletives, wh-copying, and successive cyclicity. In Uli Lutz, Gereon Müller & Arnim von Stechow (eds.), *Wh-scope marking*. Amsterdam: John Benjamins.
- Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33. 63-96. <https://doi.org/10.1162/002438902317382189>.
- Fox, Danny. 2003. On logical form. In Randall Hendrick (ed.), *Minimalist syntax*. Oxford: Blackwell.
- Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag. 1985. *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.
- Halle, Morris & Alec Marantz .1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.
- Harley, Heidi. 1995. *Subjects, events and licensing*. Cambridge, MA: MIT dissertation.

- Harley, Heidi. 2009. The morphology of nominalizations and the syntax of vP. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, definiteness and nominalization*. Oxford: OUP.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.
- Hiemstra, Inge. 1986. Some aspects of wh-questions in Frisian. *Nowele* 8. 97-110.
- Julien, Marit. 2002. *Syntactic heads and word formation*. Oxford: Oxford University Press.
- Kandybowicz, Jason. 2007. On fusion and multiple copy spell-out. In Norbert Corver & Jairo Nunes (eds.), *The copy theory of movement*. Amsterdam: John Benjamins.
<https://doi.org/10.1075/la.107.06kan>.
- Kastenholz, Raimund. 1989. *Grundburks Bambara (Manding) mit Texten*. Cologne: Rüdiger Köpfe.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kennedy, Christopher. 2002. Comparative deletion and optimality in syntax. *Natural Language & Linguistic Theory* 20. 553-621. <https://doi.org/10.1023/A:1015889823361>.
- Koizumi, Masatoshi. 1995. *Phrase structure in minimalist syntax*. Cambridge, MA: MIT dissertation.
- Kratzer, Angelika & Junco Shimoyama. 2002. Indeterminate phrases: the view from Japanese. In Yokio Otsu (ed.), *Proceedings of the Third Tokyo Conference on Psycholinguistics*. Tokyo: Hituzi Syobo.
- Landau, Idan. 2006. Chain resolution in hebrew V(P)-fronting. *Syntax* 9. 32-66.
<https://doi.org/10.1111/j.1467-9612.2006.00084.x>.
- Lapointe, Steven. 1980. *A theory of grammatical agreement*. Amherst, MA: University of Massachusetts dissertation.
- Lasnik, Howard. 1999. *Minimalist analysis*. Oxford: Blackwell.
- Lasnik, Howard & Juan Uriagereka. 1988. *A course in GB syntax*. Cambridge, MA: MIT Press.
- Lasnik, Howard & Mamoru Saito. 1992. *Move α: Conditions on its application and output*. Cambridge, MA: MIT Press.
- Lebeaux, David. 1988. *Language acquisition and the form of grammar*. Amherst, MA: University of Massachusetts dissertation.
- López, Luis. 2012. *Indefinite objects: Scrambling, choice functions, and differential marking*. Cambridge, MA: MIT Press.
- Martin, Roger & Juan Uriagereka. 2014. Chains in minimalism. In Peter Kosta, Steven L. Franks, Teodora Radeva-Bork & Lilia Schürcks (eds.), *Minimalism and beyond: Radicalizing the interfaces*. Amsterdam: John Benjamins. <https://doi.org/10.1075/lfab.11.07mar>
- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37. 69-109.
<https://doi.org/10.1162/002438906775321184>.
- McDaniel, Dana. 1986. *Conditions on wh-chains*. New York, NY: City University of New York dissertation.
- McDaniel, Dana. 1989. Partial and multiple wh-movement. *Natural Language and Linguistic Theory* 7. 565-604. <https://doi.org/10.1007/BF00205158>.

- Muñoz Pérez, Carlos. 2016. Three explanatory challenges for copy theory. In Kate Bellamy, Elena Karovskaya, Martin Kohlberger & George Saad (eds.), *ConSOLE XXIII: Proceedings of the 23rd conference of the Student Organization of Linguistics in Europe*. Leiden: Leiden University Centre for Linguistics.
- Muñoz Pérez, Carlos. 2017. *Cadenas e interfaces*. Buenos Aires: University of Buenos Aires dissertation.
- Neeleman, Ad & Hans van De Koot. 2010. A local encoding of syntactic dependencies and its consequences for the theory of movement. *Syntax* 13. 331-372. <https://doi.org/10.1111/j.1467-9612.2010.00143.x>.
- Nunes, Jairo. 1995. *The copy theory of movement and the linearization of chains in the minimalist program*. College Park, MD: University of Maryland dissertation.
- Nunes, Jairo. 1998. Bare X-bar theory and structures formed by movement. *Linguistic Inquiry* 29. 160-168. <https://doi.org/10.1162/002438998553707>.
- Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.
- Nunes, Jairo. 2011. The Copy Theory. In Cedric Boeckx (ed.), *The Oxford handbook of linguistic minimalism*. Oxford, OUP. <https://doi.org/10.1093/oxfordhb/9780199549368.013.0007>.
- Pankau, Andreas. 2014. *Replacing copies: The syntax of wh-copying in German*. Utrecht: Universiteit Utrecht dissertation.
- Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press.
- Rett, Jessica. 2006. Pronominal vs. determiner wh-words: Evidence from the copy construction. In Olivier Bonami & Patricia Cabredo-Hofherr (eds.), *Empirical issues in syntax and semantics* 6. 355-374.
- Rizzi, Luigi. 1986. On chain formation. *Syntax and Semantics* 19. 65-95.
- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.
- Ross, John. 1969. Guess who? *Proceedings of the Fifth Regional Meeting of the Chicago Linguistic Society*. Chicago: University of Chicago.
- Saab, Andrés. 2008. *Hacia una teoría de la identidad parcial en la ellipsis*. Buenos Aires: University of Buenos Aires dissertation.
- Sabel, Joachim 2000. Expletives as features. In Roger Billerey and Brook Danielle Lillehaugen (eds.), *Proceedings of WCCFL 19*. Somerville, MA: Cascadilla Press.
- Stanton, Juliet. 2016. Wholesale late merger in A'-movement: Evidence from preposition stranding. *Linguistic Inquiry* 46, 89-126. https://doi.org/10.1162/LING_a_00205.
- Takahashi, Shoichi & Sarah Hulsey. 2009. Wholesale late merger: Beyond the A/Ā distinction. *Linguistic Inquiry* 40, 387-426. <https://doi.org/10.1162/ling.2009.40.3.387>.
- Zeijlstra, Hedde. 2012. There is only one way to agree. *The Linguistic Review* 29. 491-539. <https://doi.org/10.1515/tlr-2012-0017>.