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 \( CH = \{\text{Cosmo}, \text{Cosmo}\} \), 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).

\[
\begin{align*}
(1) \quad a. & \quad \text{Cosmo was arrested.} \\
& \quad [_{TP} \text{Cosmo} \, [_{T} \, \text{was} \, \, [_{VP} \, \text{arrested Cosmo}]]]
\end{align*}
\]

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 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).\(^1\) 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

\(^1\) Although, see Martin & Uriagereka (2014) for an exception.
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. Once this new copy is merged with the main structure, as in (2c), the representation contains two constituents carrying the same index.

\[(2)\]
\begin{align*}
a. & \quad K = \left[ T_P \text{ was}^i \left[ V_P \text{ arrested}^i \text{ Cosmo}^k \right] \right] \\
b. & \quad K = \left[ T_P \text{ was}^j \left[ V_P \text{ arrested}^j \text{ Cosmo}^k \right] \right] \\
& \quad L = \text{ Cosmo}^k \\
c. & \quad K = \left[ T_P \text{ Cosmo}^k \left[ T_r \text{ was}^i \left[ V_P \text{ arrested}^i \text{ Cosmo}^j \right] \right] \right]
\end{align*}

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

\[(3)\] \text{Indexical-S}
Two constituents \( \alpha \) and \( \beta \) 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 \emph{Inclusiveness Condition}.

\[(4)\] \text{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.\footnote{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.}

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.

\footnote{See Nunes (1995; 2004) for discussion.}
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.

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 <Att,Val> where
   b. Att is drawn from the set of attributes, {A, B, C, D, E, …}
   c. and Val is drawn from the set of values, {a, b, …}

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 ∅ instead of a value.

(7) Unvalued feature
   a. An unvalued feature is an ordered pair <Att,∅> where
   b. Att is drawn from the set of attributes, {A, B, C, D, E, …}
   c. and ∅ needs to be replaced with an element from the set of values, {a, b, …}
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 <Cat,V>.

As stated in (7c), an unvalued feature <Att,∅> requires replacing the empty set ∅ 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 <Case,∅> 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 <κ,∅>, allowing it to enter in an Agree relation with matrix T, and an ω-feature <ω,∅>, 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): who⁴ is externally merged in a small clause SC carrying both unvalued features <κ,∅> and <ω,∅>; who³ is a copy of who⁴ generated through successive cyclic movement; who² is a copy of who³ that A-moved to matrix Spec,T and received nominative Case; and finally, who¹ is a copy of who² that A’-moved to Spec,C valuing its ω-feature.⁴

(11) a. [CP Who¹ [TP who² seems [TP who³ to be [SC who⁴ happy]]]]
b. Who¹ {<κ,NOM>,<ω,Q>,...} ... who² {<κ,NOM>,<ω,Q>,...} ... who³ {<κ,∅>,<ω,∅>,...} ... who⁴ {<κ,∅>,<ω,∅>,...}...

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 CH = {who¹, who², who³, who⁴}. 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 Vα is the set of morphosyntactic values of a constituent α, and that the sets V¹, V², V³ and V⁴ correspond to the values of who¹, who², who³ and who⁴, respectively. According to (11b): the sets V⁴

⁴ Numerical indexes on copies are introduced for expository purposes only.
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., $\{\ldots\} = \{\ldots\}$)  
b. $V^3 \subset V^2$ (i.e., $\{\ldots\} \subset \{\text{NOM}, \ldots\}$)  
c. $V^2 \subset V^1$ (i.e., $\{\text{NOM}, \ldots\} \subset \{\text{NOM}, Q, \ldots\}$)

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 $\text{who}$ in (11) are related through inclusion, i.e., the values of $\text{who}_3$ contain the values of $\text{who}_4$ (cf. (13a)); the values of $\text{who}_2$ contain the values of $\text{who}_3$ (cf. (13b)); and the values of $\text{who}_1$ contain the values of $\text{who}_2$ (cf. (13c)).

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

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 $\beta$ is non-distinct from a constituent $\alpha$ if for every value of $\beta$ there is an identical value in $\alpha$.

The definition in (14) is a more formal version of a very intuitive idea: if the features of $\alpha$ contain the morphosyntactic information encoded in the features of $\beta$, then $\beta$ is indistinguishable from (part of) $\alpha$. 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.\(^5\)

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

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

---

\(^5\) 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 $\alpha$ for a constituent $\beta$ that is non-distinct from $\alpha$. Abort the search if a constituent $\delta$ is found, such as the set of values $V^\delta$ of $\delta$ is a proper superset of the set of values $V^\alpha$ of $\alpha$). There are no meaningful differences between both approaches.
transitivity. That is, if the constituents $\alpha$ and $\beta$ comply with the conditions in (15), and $\beta$ and $\gamma$ also comply with (15), then $\alpha$, $\beta$ and $\gamma$ should be part of the same chain.\(^6\)

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 $\langle k,\text{NOM} \rangle$ while Cosmo\(^2\) carries its unvalued counterpart $\langle k,\emptyset \rangle$.

(16)  
\begin{align*}
  a. & \quad \text{Cosmo was arrested.} \\
  b. & \quad [\_{\text{TP}} \text{Cosmo}^1 [\_{\text{T}} [\_{\text{VP}} \text{arrested Cosmo}^2]]] \\
  c. & \quad [\_{\text{TP}} \text{Cosmo}^1 [\langle k,\text{NOM} \rangle, \ldots] [\_{\text{VP}} \ldots \text{Cosmo}^2 [\langle k,\emptyset \rangle, \ldots]]]
\end{align*}

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., $\{\ldots\}$, is a subset of the set of values of Cosmo\(^1\), i.e., $\{\text{NOM}, \ldots\}$, so this requirement is also satisfied. Finally, there is no intervenier $\delta$ 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 $\text{CH} = \{\text{Cosmo}^1, \text{Cosmo}^2\}$, 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,\(^7\) 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)  
\begin{align*}
  a. & \quad \text{Cosmo arrested Cosmo.} \\
  b. & \quad [\_{\text{TP}} \text{Cosmo}^1 [\_{\text{T}} [\_{\text{VP}} \text{Cosmo}^2 [\_{\text{VP}} \text{arrested Cosmo}^3]]]] \\
  c. & \quad [\_{\text{TP}} \text{Cosmo}^1 [\langle k,\text{NOM} \rangle, \ldots] [\_{\text{VP}} \ldots \text{Cosmo}^2 [\langle k,\emptyset \rangle, \ldots] [\_{\text{VP}} \ldots \text{Cosmo}^3 [\langle k,\text{ACC} \rangle, \ldots]]]]
\end{align*}

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., $\{\ldots\} \subseteq \{\text{NOM}, \ldots\}$), and (iii) there are no interveners between them, thus Cosmo\(^1\) and Cosmo\(^2\) form the chain $\text{CH} = \{\text{Cosmo}^1, \text{Cosmo}^2\}$. Cosmo\(^3\) cannot be considered non-distinct from the elements in $\text{CH}_1$ since it carries a value ACC that is absent in both Cosmo\(^1\) and Cosmo\(^2\) (e.g., $\{\text{ACC}, \ldots\} \not\subset \{\ldots\}$). Therefore, Cosmo\(^2\) forms a trivial chain of its own $\text{CH}_2 = \{\text{Cosmo}^3\}$. Again, this is the wanted result.

\(^6\) 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 $\alpha$ and $\gamma$ are in distinct but adjacent phases, so the conditions in (15) cannot apply to them at the same time. In this scenario, $\alpha$ and $\gamma$ may still be part of the same chain under certain conditions. That is, (i) there is a constituent $\beta$ in the edge between both phases (i.e., in a position accessible to both $\alpha$ and $\gamma$), and (ii) $\beta$ forms a chain link with both $\alpha$ and $\gamma$; 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.

\(^7\) 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) $[\_{\text{TP}} [\_{\text{VP}} \text{IA}_{<k,\text{ACC}>} [\_{\text{VP}} \text{EA}_{<k,\emptyset>} [\_{\text{VP}} \text{V} \, [\_{\text{VP}} \text{IA}_{<k,\emptyset}>]]]]$

This configuration poses 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. (iii)). 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) $[\_{\text{TP}} [\_{\text{VP}} \text{EA}_{<k,\emptyset>} [\_{\text{VP}} \text{Voice} [\_{\text{VP}} \text{IA}_{<k,\text{ACC}>} [\_{\text{VP}} \text{V} \, [\_{\text{VP}} \text{IA}_{<k,\emptyset}>]]]]]$ 

(iii) $[\_{\text{TP}} [\_{\text{VP}} \text{EA}_{<k,\emptyset>} [\_{\text{VP}} \text{IA}_{<k,\text{ACC}>} [\_{\text{VP}} \text{V} \, [\_{\text{VP}} \text{VIA}_{<k,\emptyset}>]]]]$
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\(^2\) 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\(^4\) through A-movement.

(18)  
\(a.\) Cosmo said that Cosmo was arrested.  
\(b.\) \([\text{TP} \text{Cosmo}^1 \text{T} \text{VP} \text{Cosmo}^3 \text{[\text{v} \text{said that} \text{TP} \text{Cosmo}^5 \text{[\text{v} \text{was} \text{VP} \text{arrested} \text{Cosmo}^4]]}]]\)  
\(c.\) \([\text{TP} \text{Cosmo}^1 \{<\text{k},\text{nom}>\} \ldots \text{[\text{VP} \text{Cosmo}^2 \{<\text{k},\theta>,\ldots\}] \ldots \text{[TP} \text{Cosmo}^3 \{<\text{k},\text{nom}>\} \ldots \text{[VP} \ldots \text{Cosmo}^4\{<\text{k},\theta>,\ldots\}]\}]\)

This sentence contains the chains \(\text{CH}_1 = \{\text{Cosmo}^3, \text{Cosmo}^2\}\) 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., \(\{\text{nom}, \ldots\} \subseteq \{\text{nom}, \ldots\}\)), 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., \(\{\ldots\} \subseteq \{\ldots\}\)), 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\(^2\) and who\(^6\) carry both unvalued Case \(<\text{k},\theta>\) and left-peripheral \(<\omega,\theta>\) features; who\(^2\) and who\(^6\) are copies generated through A-movement; who\(^1\) and who\(^4\) are copies generated through A'-movement.

(19)  
\(a.\) Who said who was arrested?  
\(b.\) \([\text{CP} \text{Who}^1 \{<\text{k},\text{nom}>\} \ldots \text{[\text{TP} \text{who}^2 \text{T} \text{VP} \text{who}^3 \text{[\text{v} \text{said} \text{CP} \text{who}^4 \{<\text{k},\text{nom}>\} \ldots \text{[TP} \text{who}^5 \{<\text{k},\text{nom}>\} \ldots \text{[VP} \ldots \text{who}^6\{<\text{k},\theta>,\ldots\}]\}]\}]\)  
\(c.\) \([\text{CP} \text{Who}^1 \{<\text{k},\text{nom}>\} \ldots \text{[\text{TP} \text{who}^2 \{<\text{k},\text{nom}>\} \ldots \text{[VP} \text{who}^3 \{<\text{k},\text{nom}>\} \ldots \text{[TP} \text{who}^4 \{<\text{k},\text{nom}>\} \ldots \text{[VP} \ldots \text{who}^5\{<\text{k},\theta>,\ldots\}]\}]\]}

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 \(\text{CH}_1\) is formed. The same applies for the Non-Distinctiveness relation between who\(^2\) 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 representationaical 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. 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 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.\(^8\)

---

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

9 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).
(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.\(^{10}\)

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.\(^{11}\) 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)).

(22) **German** (McDaniel 1986: 183)

\begin{verbatim}
Wen glaubt Hans wen Jakob gesehen hat?
who thinks Hans who Jakob seen has
\end{verbatim}

‘Who does Hans think Jakob saw?’

(23) **German**

\begin{verbatim}
Wen glaubt Hans dass Jakob gesehen hat?
who thinks Hans that Jakob seen has
\end{verbatim}

‘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) $[[\text{CP} \text{Wen}^{i} \text{C}_{\text{INT}} \text{glaubt Hans} [\text{CP} \text{wen}^{i} \text{C}_{\text{DECL}} \text{Jakob} [\text{VP} \text{wen}^{i} \text{gesehen}]] \text{hat}]]$

Under Indexical-S (cf. (3)), the three copies of the wh-pronoun $\text{wen}$ must necessarily form a single chain $CH = \{\text{wen}^{i}, \text{wen}^{i}, \text{wen}^{i}\}$. 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.


a. Chain Reduction (i.e., the operation deleting chain members at PF) is costly.

b. 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.

c. 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

\(^{10}\) See Kennedy (2002) and Landau (2006) for related proposals.

\(^{11}\) The present discussion focuses on German data. For examples from other languages, see Fanselow (2006) and references therein.
application of *Fusion* (Halle and Marantz 1993), an operation that combines two terminal nodes into one. For 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)

\[[x, \alpha] \lor [y, \beta] \rightarrow [x/y, \alpha, \beta]\]

where \(\alpha\) and \(\beta\) are features of \(X\) and \(Y\).

Regarding a structure like (24), Nunes proposes that the intermediate copy of *wen* and the embedded complementizer \(C_{\text{DECL}}\) undergo Fusion and form a single terminal node \([wen + C_{\text{DECL}}]\). 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 \([wen + C_{\text{DECL}}]\). 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)

\[\text{Wen denkst Du wen sie meint wen Harald liebt?} \]
\[\text{who think you who she believes who Harald loves} \]
\[\text{‘Who do you think that she believes that Harald loves?’} \]

(28) *German* (Nunes 2004: 39)

\[\text{*Wen glaubt Hans wen Jakob wen gesehen hat?} \]
\[\text{who thinks Hans who Jakob who seen has} \]
\[\text{‘Who does Hans think Jakob saw?’} \]

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.

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

\[\text{*Welchen Mann glaubst Du welchen Mann sie liebt?} \]
\[\text{which man believe you which man she loves} \]
\[\text{‘Which man do you believe that she loves?’} \]

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

³ 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.

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.
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.  [CP Wen¹ CINV glaubt Hans [CP wen² CDECL Jakob [VP wen³ gesehen] hat]]
   b.  [CP Wen¹ {<k,ACC>,<ω,θ>,...} ... [CP wen² {<k,ACC>,<ω,θ>,...} ... [VP wen³ {<k,ACC>,<ω,θ>,...}]]]

Narrow syntax generates this phrase marker and delivers it to the interfaces for interpretation. At LF, Inclusion-S generates the chain \( CH_{LF} = \{wen¹, wen², wen³\} \) 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., \( \{\text{ACC}, \ldots\} \subseteq \{\text{ACC}, \text{Q}, \ldots\} \)), and the values of *wen*² contain the values of *wen*³ (i.e., \( \{\text{ACC}, \ldots\} \subseteq \{\text{ACC}, \ldots\} \)), 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 \( [wen² + C_{DECL}] \). 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 \( C_{DECL} \).

(31) PF representation after Fusion  
[CP Wen¹ {<k,ACC>,<ω,θ>,...} ... [CP [wen² + C_{DECL}] {<k,ACC>,<ω,θ>,C,...} ... [VP wen³ {<k,ACC>,<ω,θ>,...}]]]

According to Inclusion-S, the copies of *wen* in this representation should form two distinct chains at PF. Since the values of \( [wen² + C_{DECL}] \) contain the values of *wen*² (i.e., \( \{\text{ACC}, \ldots\} \subseteq \{\text{ACC}, C, \ldots\} \)), these elements form the chain \( CH_{PF2} = \{[wen² + C_{DECL}], wen³\} \). However, the set of values of *wen*¹ does not contain the values of \( [wen² + C_{DECL}] \) (i.e., \( \{\text{ACC}, C, \ldots\} \not\subseteq \{\text{ACC}, Q, \ldots\} \)), therefore *wen*¹ forms a chain of its own \( CH_{PF1} = \{wen¹\} \). Given that *wen*¹ and \( [wen² + C_{DECL}] \) 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). Consider a case where a head \( Y \) carrying a feature \( \beta \) moves to a head \( X \) carrying a feature \( \alpha \). The resulting structure contains two copies of \( Y \), i.e., \( Y² \) in the original base position and \( Y¹ \) adjoined to \( X \).

(32)  [XP [\( X, Y¹ \)] \( (\beta) \) \( X(\alpha) \)] \( [VP Y² \)] ... ]

In a standard case of head movement, \( Y² \) should not be pronounced. This follows in both proposals from forming the chain \( CH = \{Y¹, Y²\} \) and pronouncing only the head of the chain, as usual.\(^ {15} \)

\(^ {13} \) 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 \( \{\text{ACC}, C, \ldots\} \). 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).

\(^ {14} \) I am grateful to Jonathan Bobaljik (p.c.) for this observation.

\(^ {15} \) I follow Chomsky (1986) in assuming that \( X \) has two segments in an adjunction configuration as (32). For \( X \) to dominate \( Y¹ \), every segment of \( X \) must dominate \( Y¹ \). Since this is not the case, the first node in the structure dominating \( Y¹ \) is XP. Therefore, \( Y¹ \) c-commands its copy \( Y² \), so they can form a chain.
Suppose that Fusion applies to \(Y^1\) and \(X\) after head movement. This would form a node \([Y^1 + X]\) that carries both features \(\alpha\) and \(\beta\).

\[(33) \quad [\text{XP} [Y^1 + X]_{\langle \alpha, \beta \rangle} [\text{YP} Y^2_{\langle \beta \rangle} \ldots]]\]

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 \(CH_{\text{XP}} = \{[Y^1 + X], Y^2\}\) 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.

\[(34) \quad \text{Bambara} (\text{Kastenholz} 1989; \text{as cited in} \text{Julien} 2002: 307)\]

a. dünan yé doọ min
guest AFF.PAST millet.beer drink
‘the guest drank millet beer.’

b. ń müso má jëge fëere
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.

\[(35) \quad \text{German} (\text{Fanselow & Mahajan} 2000: 196)\]

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?

‘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 do you think I saw?’

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

Wat denk je die ik gezien heb?

‘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) \[ \text{die} \ D \ [\text{wie} \ \phi \ [\text{wat} \ Q]] \]

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 \textit{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) a. \[ K = [\text{XP} \ X \ldots [\text{YP} \ldots [\text{die} \ D \ [\text{wie} \ \phi \ [\text{wat} \ Q]] \ldots ]]] \]
   b. \[ K = [\text{XP} \ X \ldots [\text{YP} \ldots [\text{die} \ D \ [\text{wie} \ \phi \ [\text{wat} \ Q]] \ldots ]]] \]
   \[ L = [\text{wie} \ \phi \ [\text{wat} \ Q]] \]
   c. \[ K = [\text{XP} \ [\text{wie} \ \phi \ [\text{wat} \ Q]] \ [\text{XP} \ X \ldots [\text{YP} \ldots [\text{die} \ D \ [\text{wie} \ \phi \ [\text{wat} \ Q]] \ldots ]]] \]

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 \(<\kappa,\emptyset> \text{ or } <\omega,\emptyset> \text{ are omitted for simplicity.}

(42) a. \(\text{wat} = \{Q, \ldots\} \)
   b. \(\text{wie} = \{Q, \phi, \ldots\} \)
   c. \(\text{die} = \{Q, \phi, \text{d}, \ldots\} \)

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).16 In these terms, a derivation like (41) involves (i) a structure K containing a node with the set of features \(\{Q, \phi, \text{d}, \ldots\} \) corresponding to "die" as in (43a), (ii) copying from this node the subset of features \(\{Q, \phi, \ldots\} \) as in (43b), and (iii) merging this newly generated set into the main structure K as in (43c).

(43) a. \(K = [\text{xp} \ X \ldots [\text{yp} \ldots \{Q, \phi, \text{d}, \ldots\} \ldots ]]] \)
   b. \(K = [\text{xp} \ X \ldots [\text{yp} \ldots \{Q, \phi, \text{d}, \ldots\} \ldots ]]] \)
   \[ L = \{Q, \phi, \ldots\} \]
   c. \([\text{xp} \ {Q, \phi, \ldots} [\text{xp} \ X \ldots [\text{yp} \ldots \{Q, \phi, \text{d}, \ldots\} \ldots ]]]\]

---

16 Barbiers et al. (2010) argue that applying Partial Copying to the features of a constituent leads to a violation of \textit{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.
Given that the sets of features \{Q, \varphi, \ldots\} and \{Q, \varphi, d, \ldots\} satisfy the Vocabulary Insertion rules of *wie* and *die*, respectively, the relevant nodes are spelled-out as sketched in (44).

\[(44) \quad [_{\text{XP}} \quad \text{wie} \quad [_{X} \quad \ldots \quad [_{\text{VP}} \quad \ldots \quad \text{die} \quad \ldots \quad ]]]\]

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.

\[(45) \quad \text{Analysis of (36) in terms of Partial Copying} \quad [_{\text{CP}} \quad \text{Wat}^1_{\{Q, \varphi, \ldots\}} \text{C}_{\text{INT}} \quad \text{denk je} \quad [_{\text{CP}} \quad \text{wie}^2_{\{Q, \varphi, \ldots\}} \text{C}_{\text{DECL}} \quad \text{ik} \quad [_{\text{VP}} \quad \text{wie}^3_{\{Q, \varphi, \ldots\}} \text{gezien}] \text{heb}]]\]

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 \(\varphi\)-features are not copied in the latter case.

\[(46) \quad \text{Analysis of (37) in terms of Partial Copying} \quad [_{\text{CP}} \quad \text{Wie}^1_{\{Q, \varphi, \ldots\}} \text{C}_{\text{INT}} \quad \text{denk je} \quad [_{\text{CP}} \quad \text{die}^2_{\{Q, \varphi, d, \ldots\}} \text{C}_{\text{DECL}} \quad \text{ik} \quad [_{\text{VP}} \quad \text{die}^3_{\{Q, \varphi, d, \ldots\}} \text{gezien}] \text{heb}]]\]

\[(47) \quad \text{Analysis of (38) in terms of Partial Copying} \quad [_{\text{CP}} \quad \text{Wat}^1_{\{Q, \varphi, \ldots\}} \text{C}_{\text{INT}} \quad \text{denk je} \quad [_{\text{CP}} \quad \text{die}^2_{\{Q, \varphi, d, \ldots\}} \text{C}_{\text{DECL}} \quad \text{ik} \quad [_{\text{CP}} \quad \text{die}^3_{\{Q, \varphi, d, \ldots\}} \text{gezien}] \text{heb}]]\]

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.

\[(48) \quad \text{a.} \quad *\quad [_{\text{CP}} \quad \text{wie}^2_{\{Q, \varphi, \ldots\}} \ldots \quad [_{\text{CP}} \quad \text{wat}^1_{\{Q, \ldots\}} \ldots]] \quad (\text{cf. (39a)})
\quad \text{b.} \quad *\quad [_{\text{CP}} \quad \text{die}^3_{\{Q, \varphi, d, \ldots\}} \ldots \quad [_{\text{CP}} \quad \text{wie}^2_{\{Q, \varphi, \ldots\}} \ldots]] \quad (\text{cf. (39b)})
\quad \text{c.} \quad *\quad [_{\text{CP}} \quad \text{die}^3_{\{Q, \varphi, d, \ldots\}} \ldots \quad [_{\text{CP}} \quad \text{wat}^1_{\{Q, \ldots\}} \ldots]] \quad (\text{cf. (39c)})\]

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.

\[(49) \quad \text{a.} \quad [_{\text{CP}} \quad \text{Wat}^1_{\{Q, \ldots\}} \ldots \quad [_{\text{CP}} \quad \text{wie}^1_{\{Q, \ldots\}} \ldots \quad [_{\text{VP}} \quad \text{wie}^1_{\ldots}\ldots]] \quad (\text{cf. (36)})
\quad \text{b.} \quad \text{CH} = \{\text{wat}^1, \text{wie}^1, \text{wie}^1\}\]

\[(50) \quad \text{a.} \quad [_{\text{CP}} \quad \text{Wie}^1_{\{Q, \ldots\}} \ldots \quad [_{\text{CP}} \quad \text{die}^1_{\{Q, \ldots\}} \ldots \quad [_{\text{VP}} \quad \text{die}^1_{\ldots}\ldots]] \quad (\text{cf. (37)})
\quad \text{b.} \quad \text{CH} = \{\text{wie}^1, \text{die}^1, \text{die}^1\}\]

\[(51) \quad \text{a.} \quad [_{\text{CP}} \quad \text{Wat}^1_{\{Q, \ldots\}} \ldots \quad [_{\text{CP}} \quad \text{die}^1_{\{Q, \ldots\}} \ldots \quad [_{\text{VP}} \quad \text{die}^1_{\ldots}\ldots]] \quad (\text{cf. (38)})
\quad \text{b.} \quad \text{CH} = \{\text{wat}^1, \text{die}^1, \text{die}^1\}\]
Under standard assumptions, these chains should behave exactly in the same way as any chain $CH = \{XP_1, XP_2, XP_3\}$ 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 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 don’t you think she has invited?’

(54)   *Dutch* (Barbiers et al. 2010: 40)
  Wat denk je niet wie zij uitgenodigd heeft?
  ‘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 don’t you think she has invited?’

(56)   *German* (Rett 2006: 359)
  *Wen glaubst du nicht wen sie liebt?
  *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 com-
plementizer 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 inter-
vention 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 fea-
tures (i.e., \{(q, \varphi, \ldots) \subseteq \{(q, \varphi, \ldots)\}); however, since the features of wie^2 are not a subset
of the features of wie^1 (i.e., \{(q, \varphi, \ldots) \not\subseteq \{(q, \varphi, \ldots)\}), wie^1 forms a trivial chain of its own
CH_{PF1} = \{wie^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 wie^1 forms the trivial chain CH_{LF1} = \{wie^1\}.
These results are summarized in (58).

(58)  Chain formation at the interfaces according to Inclusion-S (cf. (36))
\begin{itemize}
  \item[a.] \([CP \; \text{wat}^1(0, \ldots) \ldots [CP \; \text{wie}^2(0, \varphi, \ldots) \ldots [VP \; \text{wie}^3(0, \varphi, \ldots) \ldots]]\]
  \item[b.] CH_{PF1} = \{wat^1\}; CH_{PF2} = \{wie^2, wie^3\}
  \item[c.] CH_{LF1} = \{wat^1\}; CH_{LF2} = \{wie^2, wie^3\}
\end{itemize}

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 \(\alpha\) and \(\beta\) are non-distinct if (i) \(\alpha\) c-commands \(\beta\) and (ii) \(\alpha\) contains the
information encoded in \(\beta\). 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., \{(q, \varphi, \ldots) \not\subseteq \{(q, \varphi, \ldots)\}), so they form two separate chains
as shown in (59). In the latter, the features of \( \text{die}^2 \) are not a subset of the features of \( \text{wat}^1 \) (i.e., \( \{Q, \varphi, \ldots\} \not\subset \{Q, \ldots\} \)), 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. \left[\text{CP} \right] \text{wie}^1_{\{Q, \varphi, \ldots\}} \ldots \left[\text{CP} \right] \text{die}^2_{\{Q, \varphi, \ldots\}} \ldots \left[\text{VP} \right] \text{die}^3_{\{Q, \varphi, \ldots\}} \ldots\)
\[
b. \left[\text{CH}_{\text{PF}1}\right] = \{\text{wie}^1\}; \left[\text{CH}_{\text{PF}2}\right] = \{\text{die}^2, \text{die}^3\}
\[
c. \left[\text{CH}_{\text{LF}1}\right] = \{\text{wie}^1\}; \left[\text{CH}_{\text{LF}2}\right] = \{\text{die}^2, \text{die}^3\}
\]

\[(60)\] Chain formation at the interfaces according to Inclusion-S (cf. (41))
\[
a. \left[\text{CP} \right] \text{wat}^1_{\{\ldots\}} \ldots \left[\text{CP} \right] \text{die}^2_{\{Q, \varphi, \ldots\}} \ldots \left[\text{VP} \right] \text{die}^3_{\{Q, \varphi, \ldots\}} \ldots\)
\[
b. \left[\text{CH}_{\text{PF}1}\right] = \{\text{wat}^1\}; \left[\text{CH}_{\text{PF}2}\right] = \{\text{die}^2, \text{die}^3\}
\[
c. \left[\text{CH}_{\text{LF}1}\right] = \{\text{wat}^1\}; \left[\text{CH}_{\text{LF}2}\right] = \{\text{die}^2, \text{die}^3\}
\]

Extending Partial Copying to capture wh-copying patterns is conceptually simple. \(^{17}\) Assume a feature \( \{F\} \) that is common to \( \text{wat}, \text{wie} \) and \( \text{die} \). 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 \( \{Q, \varphi, F, \ldots\} \) corresponding to the wh-pronoun \( \text{wie} \) 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 \( \{Q, \varphi, \ldots\} \) is generated and merged into matrix Spec,C. Since this newly formed set of features also corresponds to the phonological exponent \( \text{wie} \), a third occurrence of this pronoun is generated in the structure.

\[(61)\] Analysis of (52) in terms of Partial Copying
\[\left[\text{CP} \right] \text{Wie}^1_{\{Q, \varphi, \ldots\}} \text{N} \text{NT} \text{en} \text{je} \left[\text{CP} \right] \text{wie}^2_{\{Q, \varphi, F, \ldots\}} \text{C} \text{ decl.} \text{i} \text{k} \left[\text{VP} \right] \text{wie}^3_{\{Q, \varphi, \ldots\}} \text{gezien} \text{[heb]}\]

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

\[(62)\] Chain formation at the interfaces according to Inclusion-S (cf. (52))
\[
a. \left[\text{CP} \right] \text{wie}^1_{\{Q, \varphi, \ldots\}} \ldots \left[\text{CP} \right] \text{wie}^2_{\{Q, \varphi, F, \ldots\}} \ldots \left[\text{VP} \right] \text{wie}^3_{\{Q, \varphi, \ldots\}} \ldots\)
\[
b. \left[\text{CH}_{\text{PF}1}\right] = \{\text{wie}^1\}; \left[\text{CH}_{\text{PF}2}\right] = \{\text{wie}^2, \text{wie}^3\}
\[
c. \left[\text{CH}_{\text{LF}1}\right] = \{\text{wie}^1\}; \left[\text{CH}_{\text{LF}2}\right] = \{\text{wie}^2, \text{wie}^3\}
\]

Given that \( \left[\text{CH}_{\text{PF}1}\right] \) and \( \left[\text{CH}_{\text{PF}2}\right] \) 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 \( \left[\text{CH}_{\text{LF}1}\right] \) and \( \left[\text{CH}_{\text{LF}2}\right] \) in (58), (59), (60) and (62). In each of these cases, the trivial chain \( \left[\text{CH}_{\text{LF}1}\right] \) 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 \( \left[\text{CH}_{\text{LF}2}\right] \) does receive a \( \theta \)-role as one of its members occupies a thematic position; however, none of the wh-elements pertaining to \( \left[\text{CH}_{\text{LF}2}\right] \) is in a spec-head configuration with \( \text{C}_{\text{NT}} \). Since there is no overt syntactic relation between \( \left[\text{CH}_{\text{LF}2}\right] \) 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

\(^{17}\) See Hiemstra (1986) for a similar account of wh-copying patterns.
be licensed by establishing a covert dependency with the interrogative C-domain. The German sentence in (63), for instance, is supposed to involve an abstract relation linking the wh-pronoun wo ‘where’ and the interrogative complementizer $C_{\text{INT}}$ 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 $C_{\text{INT}}$ hat Luise wo gesehen?  
In a similar fashion, the head of each of the wh-chains $CH_{\text{LF2}}$ 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 $C_{\text{INT}}$ and $wie^2$.

(65)  
$[\text{CP} \ \text{Wat}^1 C_{\text{INT}} \text{denk je } [\text{CP} \ \text{wie}^2 \text{ ik } \text{wie}^3 \text{ gezien heb}]]$? (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 $C_{\text{INT}}$ and $wie^2$ prevents the licensing of the wh-chain $CH_{\text{LF2}} = \{wie^2, wie^3\}$.

(66)  
*$[\text{SC} \ \text{Wat}^1 C_{\text{INT}} \text{denk je niet } [\text{SC} \ \text{wie}^2 \text{ zij } \text{wie}^3 \text{ uitgenodigd heeft}]]$ (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  
’Whom have you seen?’

\[\text{I remain agnostic whether this covert dependency involves LF-movement (Lasnik & Saito 1992), feature movement (Pesetsky 2000), or a focus-related computation (Beck 2006).}\]
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 is a genius did he believe?
   b. *[DP Which argument [CP that Cosmo is a genius]] did he believe
   [DP which argument [CP that Cosmo 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 Cosmo made]] did he 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 made]] did he 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) \[\text{[DP The claim that Cosmo}_1\text{ was asleep] seems to him}_1\text{ to be correct.}\]

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.\(^{19}\)

(75) \[\text{[DP The [NP claim that Cosmo}_1\text{ was asleep]] seems to him}_1\text{ to be [DP the] correct.}\]

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) \[\text{[DP His}_1\text{ picture of the president}_2\text{] seemed to every man}_1\text{ to be seen by him}_2\text{ to be an intrusion.}\]

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 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.\(^{20}\)

(77) \[\text{[DP His}_1\text{ picture of the president}_2\text{] seemed to [every man}_1\text{ [DP his}_1\text{] to be seen by him}_2\text{ [DP his}_1\text{] to be [DP his}_1\text{ an intrusion.}\]

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) \[a. \text{CH} = \{[\text{DP which argument that Cosmo made}], [\text{DP which argument}]\} \quad (\text{cf. (71)}) \]

\[b. \text{CH} = \{[\text{DP the claim that Cosmo was asleep}], [\text{DP the}]\} \quad (\text{cf. (74)}) \]

\[c. \text{CH} = \{[\text{DP his picture of the president}], [\text{DP his}], [\text{DP his}], [\text{DP his}]\} \quad (\text{cf. (76)}) \]

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

\(^{19}\) 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.

\(^{20}\) 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.
constituent containing the relevant R-expression. Consider the following sample derivation. A constituent $\alpha P$ is generated in a position where it is c-commanded by a pronoun; since $\alpha P$ does not contain an R-expression, Condition C is respected (cf. (79a)). Later in the derivation, $\alpha P$ moves to a position where it c-commands the pronoun; both copies of $\alpha P$ share the same index (cf. (79b)). As a third step, a $\beta P$ containing an R-expression is inserted into $\alpha P$ as in (79c). At this point, the pronoun and the R-expression can be correferential, and since both occurrences of $\alpha P$ share the same index, they form a chain.

(79)  
\[ [xp X \ldots [yp Pronoun_i \ldots [zp \ldots [\alpha^i]]]] \]

b. \[ [xp [\alpha^i] [x X \ldots [yp Pronoun_i \ldots [zp \ldots [\alpha^i]]]] \]

c. \[ [xp [\alpha^i [\beta P \text{R-expression}_i] [x X \ldots [yp Pronoun_i \ldots [zp \ldots [\alpha^i]]]] \]

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., $\beta P$ in (79c) must be an adjunct), while Wholesale Late Merger extends the empirical domain of the former to complements (i.e., $\beta P$ 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.

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 $\phi$-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 $\phi$-features. This DP agrees with T and receives nominative Case.  

(81) \[ [DP \text{The \{<}_{k,NOM}, <}_{k,N}, ...} \text{ seems to him, to be correct \{<}_{k,B}, <}_{k,B}, ...} \text{.}] \]

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

---

21 While the configuration in (81) seems to require Upward Agree (Zeijlstra 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**.
values of the full-DP contain the values of \( D_{\min/\max} \) (i.e., \( \{\ldots\} \subseteq \{\text{NOM, SG,} \ldots\} \)), 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.\(^{22}\)

\[
\begin{align*}
(82) & & [\text{DP} \text{His picture of the president}]_{1} \{<\kappa,\text{NOM}>, <\text{Num},\text{SG}>, \ldots\} \text{ seemed to every man,} \\
& & [\text{DP} \text{his}]_{1} \{<\kappa,\emptyset>, <\text{Num},\emptyset>, \ldots\} \text{ to be seen by him,} \\
& & [\text{DP} \text{his}]_{1} \{<\kappa,\emptyset>, <\text{Num},\emptyset>, \ldots\} \text{ an intrusion.}
\end{align*}
\]

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 \( \omega \)-feature, it enters in an Agree relation with \( C_{\text{int}} \). Notice, however, that the DP still lacks a value for its \( \kappa \)-feature.

\[
\begin{align*}
(83) & & [\text{CP} \text{Which argument [ADU that Cosmo made]}]_{1} \{<\kappa,\emptyset>, <\omega,\text{Q}>, \ldots\} \text{ did he believe [\text{DP} \text{which argument}]_{1} \{<\kappa,\text{ACC}>, <\omega,\emptyset>, \ldots\}].
\end{align*}
\]

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

\[
\begin{align*}
(84) & & [\text{CP} \text{Which argument [ADU that Cosmo made]}]_{1} \{<\kappa,\text{ACC}>, <\omega,\text{Q}>, \ldots\} \text{ did he believe [\text{DP} \text{which argument}]_{1} \{<\kappa,\emptyset>, <\omega,\text{Q}>, \ldots\}].
\end{align*}
\]

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 \( \varphi \)-features. Therefore, this type of element is unable to value the \( \varphi \)-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:

\[\text{There are reasons to believe that his is not the head of this DP, but the specifier of a null determiner, e.g., } [\text{DP his }]_{1} \{<\kappa,\text{every [NP idea]} >\}. \text{ Translating this analysis to the sentence in (76) is fairly simple within Inclusion-S: a null } D_{\min/\max} \text{ and the DP his picture of the president form a chain in exactly the same way as in (82). However, an account based on Wholesale Late Merger would have to assume a more complex derivation, e.g., that both the specifier his and the complement NP may be introduced countercyclically.}\]
(85) Only caseless chain members (i.e., “traces” of A-movement) may be $D^{\min/\max}$.

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 $\rightarrow$ (Det [Pred $\lambda y(y = x)$])
   b. Determiner Replacement: (Det) [Pred $\lambda y(y = x)$] $\rightarrow$ the [Pred $\lambda y(y = x)$]

This rule transforms a wh-phrase into a definite description with anaphoric value.23 The subpart in (86a) introduces a predicate $<e,t>$ that functions as a variable and is interpreted compositionally with a complete nominal predicate, i.e., another $<e,t>$ expression, through *Predicate Modification* (Heim & Kratzer 1998). Incomplete nominal predicates (i.e., nouns lacking some argument) are not $<e,t>$ 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) $[\text{CP [DP *which woman]} \{<\kappa,\text{acc}>,<\omega,\text{Q}>,\ldots\} \text{ C} \text{int} \text{ did Elaine meet} [\text{DP *which girl]} \{<\kappa,\text{ACC}>,<\omega,\text{Q}>,\ldots\}]]$

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.24

(89) *Which woman $\lambda x$. Elaine met the girl x

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

23 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.

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

(i) $[\text{CP wie}_{(0,\ldots,\ldots)} \ldots [\text{CP wat}_{(0,\ldots,\ldots)} \ldots]]$ (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’.
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.25

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 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.

Abbreviations

ACC = accusative, AFF = affirmative, C = Complementizer, C DECL = declarative Complementizer, C INT = interrogative Complementizer, CP = Complementizer phrase, D = Determiner, DP = Determiner phrase, EA = external argument, IA = internal argument, NEG = negation, NOM = nominative, NUM = number, REL.PRON = relative pronoun, SG = singular, T = Tense, TP = Tense phrase, V = Verb, VP = Verb phrase.

Acknowledgements

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 Verdeccia, Pablo Zdrojewski, Romina Trebisacce, Viola Schmitt, 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.
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.

Competing Interests

The author has no competing interests to declare.

References


---

**How to cite this article:** Pérez, Carlos Muñoz. 2018. Recognizing copies: On the definition of Non-Distinctiveness. *Glossa: a journal of general linguistics* 3(1): 40.1–29, DOI: https://doi.org/10.5334/gjgl.271

**Submitted:** 30 September 2016  **Accepted:** 04 September 2017  **Published:** 23 March 2018

**Copyright:** © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.