On Labeling: Principle C and Head Movement

CARLO CECCHETTO
University of Milan-Bicocca
carlo.cecchetto@unimib.it

CATERINA DONATI
University of Urbino
caterina.donati@uniurb.it

In this paper, we critically re-examine the two algorithms that govern phrase structure building according to Chomsky (2005). We accept the algorithm that dictates that a lexical item transmits its label when it is merged with another object (the Head Algorithm) but reject the second algorithm proposed by Chomsky and replace it with the Probing Algorithm, which states that the probe of any kind of Merge is the label. In addition to capturing core cases of phrase structure building, these two algorithms shed light on Principle C effects and the syntax of wh constructions, which we analyze as cases of conflict between them. In these two configurations a lexical item (which should become the label in compliance with the Head Algorithm) is merged with a syntactic object that, being the probe of the operation, should become the label in compliance with the Probing Algorithm. In one case, this conflict produces two alternative outputs (a question or a free relative) that are both acceptable. In Principle C configurations, one of the resulting output (the one determined by the Head Algorithm) produces an object that is not interpretable. This way, Principle C effects are reduced to cases of mislabeling, with no need to postulate a specific condition to rule them out.

1. Introduction

One important assumption in the minimalist program, initially formulated by Chomsky (1995), is the Inclusiveness Condition, according to which narrow syntax merely operates on lexical items and cannot "add" interpretative material. This is usually interpreted as meaning that semantically active material such as indices, bar levels or labels cannot be inserted in the course of a derivation.

Still, there is an important theoretical notion that does not seem to de dispensable, namely that Merge yields labeled syntactic objects: when Merge forms a syntactic object, the features associated with one and only one of the assembled items are visible for further computation[1]. If the inclusiveness condition is to be taken seriously,

[1] Collins (2002) sketches a theory of syntax in which labels can be completely dispensed with. However his polemical objective is the notion of label as an extra object distinct from the two items that

this cannot be captured through the insertion of a new object distinct from the items that are merged, such as a label in standard X-bar theory. Rather, we shall define label as a subset of features, as in (1).

(1) Label: features of a syntactic object (SO) which are visible to further computation

Therefore syntax should have a simple, automatic way to calculate the label of any syntactic object (namely, a relevant subset of its features). Following Chomsky (2005), we shall call this the labeling algorithm(s). In this paper we discuss how these algorithms should be defined, keeping with the Inclusiveness Condition and taking seriously the unification of syntactic operations put forward in recent works, reducing movement to a special instance of Merge.
The paper is organized as follows: section 2 is focused on the issue of labeling from a theoretical point of view: we first discuss the two algorithms proposed in Chomsky (2005) providing a criticism and then propose a new pair of algorithms and show how they work both for External and for Internal Merge. Involving two different labeling algorithms, the system crucially predicts that cases of conflicts between them could arise. Two such case studies are discussed in the remnant of the paper: the first is discussed in section 3 and concerns a conflict arising with External Merge: a case where the tension between the two algorithms derives what is standardly known as Principle C; the second is discussed in section 4 and illustrates the same kind of conflict in connection to Internal Merge: interrogatives and free relatives are the case in point. Section 5 concludes the paper.

2. The Labeling Algorithms

Chomsky (2005) proposes that the two algorithms in (2) and (3) are necessary and sufficient to yield labeled syntactic objects in most derivations:

(2) In {H, α}, H a lexical item (LI), H is the label
(3) If α is internally merged to β forming {α,β}, then the label of β is the label of {α,β}.

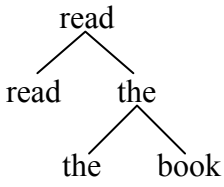
The status of these two principles is very different, as is their likelihood as syntactic primitives. Let us discuss them briefly in turn.
The concept of LI which is implicit in the algorithm in (2) is minimally simple: an LI is an item listed in the lexicon as such. Rephrased in standard X-bar terms, (2) claims

---

are merged, as was in Chomsky's (1995) version of bare phrase structure theory. In that early version of the theory, the output of merging of X and Y was not the minimally simple object {X,Y}, but was either {X,{X,Y}} or {Y,{X,Y}}, depending on which category projects. We believe that once a label is defined as a subset of the features of one of the two merging objects, the quest for simplification argued for by Collins can be satisfied. Still, differences between Collins's approach and the one based on labelling algorithms remain. They do not arise so much in the area of phrase structure theory, since the notion of label is replaced by the closely related notion of Locus, as Collins himself notice (p. 48), nor in the theory of subcategorization, for Collins assumes that lexical features like +/- V, +/- N do exist, although they do not project at the phrasal level. The area in which differences arise is the theory of locality, since a label-less theory *à la* Collins requires a reformulation of the Minimal Link Condition, with potentially different empirical predictions. We cannot make a complete comparison between our approach and Collins', due to reason of space.

that it is always a head that projects. Under minimal assumptions on the relation of syntax and lexicon, (2) is a very likely candidate of a syntactic primitive, and we shall adopt it as such[2]. To illustrate how (2) works, consider the derivation in (4), a case of External Merge of an LI to an SO: by virtue of (2), the SO generated by merging the LI with the SO gets the label of the LI (i.e., recall, a subset of its features: see (1)).

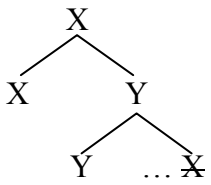
(4)



read the book

However, labeling is an issue concerning any kind of merge. If by merge we mean not only External Merge but also Internal Merge (i.e. movement), then we expect the algorithm in (1) to work indistinguishably in cases like (4) and in cases where movement is involved. Consider for example the abstract derivation in (5), where a simple lexical item is internally merged to a syntactic object.

(5)



By (2), X provides the label. As such, (5) illustrates an interesting consequence of the algorithm (2) when applied to Internal Merge: (2) predicts that what is traditionally called head movement has the property of modifying the label of its target. The algorithm in (3) is exactly meant to avoid such a consequence, and ensure that "in all movement operations it is always the target that projects": (3) explicitly sets apart External Merge, basically stipulating a residual of a "movement theory". This stipulation goes against the unification of syntactic operations explicit in the reduction of movement to Internal Merge. As such, (3) is a severe departure from the minimalist assumptions and ideally should be discarded. However, the algorithm in (2) alone is not enough to provide the computational system with an automatic device for labeling the core cases of syntactic objects created by merge. While we might expect labeling to be not always univocal, leaving some work to the interfaces, with (2) alone we would have too much indeterminacy, many suspicious and even wrong predictions. Let us see some of them in detail.

First of all, a system working with one an only one algorithm as (2) would have nothing to say about the very first step of any derivation, when two lexical items get merged, as in (6).

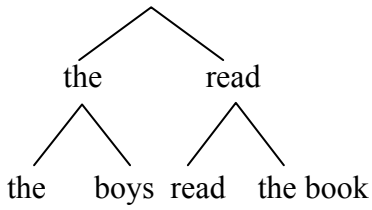
(6) {saw, John}

This would give us a weird grammar, in which any computation automatically runs at least two parallel derivations given any pair of lexical items, depending on which provides the label. In fact, this problem also arises if one assumes the pair of

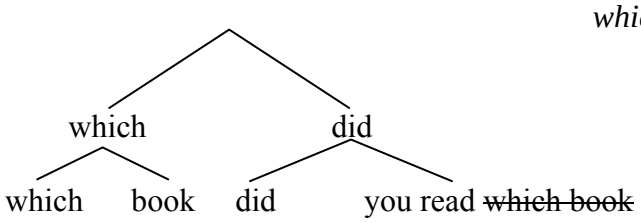
[2] Chomsky (2006) actually derives (2) from simple economy considerations: if the label is the lexical item, and by definition the label is the probe, then the search space for a probe is simplified w.r.t. c-command: simply reduced to the complement domain of a lexical item.

algorithms (2) and (3) proposed by Chomsky (2005). Chomsky acknowledges it but claims that a multiple spell out system like the theory of phases ensures that the "wrong" derivation will crash early enough. Still, the system would introduce the computational burden of maintaining two parallel derivations up to the next higher phase even in trivial cases like (6) that are not temporarily ambiguous in any reasonable sense.
More problematic cases systematically arising in a system containing only (2) are illustrated in (7) and (8), for External Merge and Internal Merge, respectively.

(7) *the boys read the book*

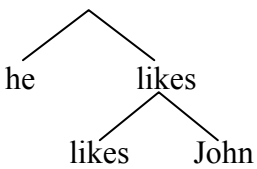


(8) *which book did you read*

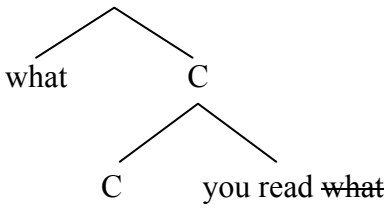


Both in (7) and (8) two objects are merged none of which is a lexical item: (2) might be taken to mean that they don't have any label: a clearly unwanted result[3]. Alternatively, a system that has (2) as its only labeling algorithm might be taken to mean that labeling cannot be decided in such cases and this is equally unsatisfactory.
Finally, the system yields wrong or at least very suspicious results in a number of contexts where a lexical item gets merged with a syntactic object, as illustrated in (9) and (10) for External Merge and Internal Merge, respectively.

(9) *he likes John*



(10) *what you read*



[3] However an interesting consequence of such a system would be that structures like (7), being unlabelable, are not tolerated and highly instable: this might provide a promising explanation for the systematic tendency to disrupt small clauses analyzed by Moro (2000).

Both in (9) and (10) the algorithm in (2) predicts that the label should be provided by the lexical head: a clearly wrong result in the case of (9), which is interpreted as a clause, not as a DP; a very suspicious result in (10), which can be interpreted as clausal in nature, not (necessarily) as a DP. Notice that the case in (10) is the reason why Chomsky (2005) stipulates the algorithm (3), in order to ensure that movement never changes the label of its target. In addition of being an unjustified stipulation, as already discussed above, the algorithm (3) does not solve the problem of (9), which does not involve movement.

This quick review of some representative cases of merge clearly shows that a system that contains only (2) as a labeling algorithm is clearly unsatisfactory. A closer look to the problematic cases can however give us a simple solution, though. Consider for example the cases of 'first merge'. It is very clear in (6) that the two lexical items selected from the numeration are not playing the same role in the computation: to put it very simple, a transitive verb like "saw" needs a direct object like "John", while "John" does not (necessarily) need "saw." A classical way to describe this asymmetry is to say that "John" **saturates** "saw", and not vice-versa. Given the strong unification thesis, it becomes very appealing to frame this asymmetric relation between the two members of a merging pair in terms of a Probe-Goal relation: in this spirit we might say that "saw" has an unvalued feature (a selectional feature, or, in Chomsky 2005's terms, an edge feature) — a Probe — which gets valued by some feature of "John", the Goal. Capitalizing on this asymmetry, we might propose the following algorithm, which should replace the stipulation in (3) and complement the labeling algorithm (2).

(11) In $\{\alpha, \beta\}$, α the probe of Merge between α and β, α is the label

What (11) basically says is that merge is always asymmetrically triggered and is governed by the features of the items involved. We shall call (11) **Probing Algorithm**, while we might call (2) **Head Algorithm** for ease of reference**,** since it basically replaces the definition of head in X-bar theory. To illustrate, in (6) "saw" provides the label because both algorithms converge: "saw" is the probe referred to in (11) and is a lexical item, in compliance with (2)[4].

Let us go back now to the other problematic cases. Suppose we (externally) merge a syntactic object with another syntactic object (case 7): the Head Algorithm (2) has nothing to say since no LI is involved. But there will always be one (and by hypothesis only one) of the two syntactic objects that has triggered the operation needing the valuation of its edge feature: this one ("read"*)* will label the output. No difference arises when the same configuration is given by Internal Merge (i.e. movement): in (8) the operation is triggered by the edge feature of the clause, and the output ends up being itself clausal.

An interesting consequence of this system is that, since it contains two labeling algorithms, it predicts that there might exist cases of conflict. The problematic cases discussed above (9 and 10) can be usefully analyzed as such. In both cases an LI (which should become the label in compliance with the Head Algorithm 2) is merged with an SO that, being the probe of the operation, should become the label in compliance with the Probing Algorithm (11). The remnant of the paper will be devoted to discussing these two cases in great detail.

[4] The alternative derivation, in which *John* is the label of {saw, John}, obeys the Head Algorithm (2) but violates the Probing Algorithm (11).

3. Principle C as a case of mislabeling

In this section we shall show that standard cases of Principle C can be reduced to symptoms of a mislabeling, dispensing with the canonical definition of Principle C, which is incompatible with the inclusiveness condition and is not minimally rooted as a syntactic primitive. For the purposes of this paper we will refer to the formulation in (12) as the canonical definition of Principle C.

(12) An R-expression cannot be c-commanded by a coindexed category

(12) is a negative condition on the distribution of indexes. The tacit assumption is that NPs can be freely assigned identical indexes unless this is explicitly blocked. Principle C introduces one such blocking condition. Apart from the dubious status of indexes in the minimalist program (see above), another possible concern with the canonical formulation of Principle C is that it is conceived as a primitive of the theory (whence the label *Principle* C), which is codified as such in UG. Although this is not unreasonable, since Principle C is likely to be a language universal, if Principle C were deducible from more primitive elements of UG, we would have an important simplification of the theory. In the same minimalist spirit, various attempts have been made to dispense with binding-theoretical principles. Kayne (2005) and Schlenker (2006) try to reduce Principle C from more primitive conditions[5], while Chomsky (1993), Hornstein (2006) and Reuland (2001), among others, offer minimalist reformulation of Principle A and Principle B. We will be concerned uniquely with Principle C and propose that its empirical coverage can be made to follow from to the interaction between the Head and the Probing Algorithms. This, as we will show, in addition to being conceptually desirable, is also preferable on empirical ground.

3.1 Principle C reduced to the labeling algorithms

In order to deduce Principle C from the labeling algorithms we introduce a new notion, namely "referential valuation". The intuition that we would like to build on is that grammatical relations are asymmetric. For example, a DP values the agreement morpheme of the verb (and not vice versa). Similarly, a DP values the φ-features of an adjectival expression and a noun values the edge feature of a determiner. We propose that something like that happens in a different domain, namely referential properties of DPs. For example, if a referential expression like a proper name and a pronoun have the same semantic value (i.e. they pick out the same individual), this relation is asymmetric in the sense that it is the semantic value of the proper name which determines the semantic value of the pronoun (and not vice versa). Trying to be slightly more formal we will assume that a category A has an intrinsic semantic value, namely it is a referential expression, if and only if its semantic value is independent from the function that assigns a value to free variables. It follows from this that, for example, a proper name has an intrinsic semantic value, while a pronoun does not.

[5] In Kayne's (2005) theory every case in which a pronoun and its antecedent have the same semantic value is reduced to an instance of movement out of a clitic doubling configuration. Principle C effects are then reduced to illicit cases of movement. See footnote 13 for a specific observation on Kayne's system. In Schlenker's (2006) approach Principle C (as well as the other binding-theoretic principles) follow from a non-standard interpretive procedure, which can mimic the relation of c-command in the semantic component. The basic condition that replaces Principle C is an interpretative filter which prevents any given object from appearing twice in any sequence of evaluation for a given sentence. See footnote 8 for a specific observation on Schlenker's system.

We will define the notion of referential valuation as follows: A referentially values B if the semantic component receives an instruction from narrow syntax which has the effect that all the assignment functions that do *not* assign the individual that is the intrinsic semantic value of A to B are disregarded. This way the semantic value of B is intrinsically determined by the semantic value of A.
Given the similarities with other asymmetric relations, we assume that referential valuation is just another case of Probe-Goal matching, in which the Probe (a pronominal expression) searches for the Goal (a referential expression). As a result, we are widening the notion of Probe with respect to the way it is standardly conceived: Probe-Goal matching does not involve only valuation of φ-features, *wh*-features etc. but also edge features and referential valuation: pretty much in the spirit of the strong unification we are trying to comply with in this paper.
Having introduced the notion of referential valuation, we are ready to discuss a standard case of Principle C violation like (13), in which "he" and "John" have the same semantic value (for the reader's convenience, we will continue to indicate that two categories have the same semantic value by coindexing them).

(13) *He$_i$ likes John$_i$

The sentence in (13) is a clear case of conflict between the two algorithms discussed above: according to the Head Algorithm (2), "he", being an LI, should provide the label. According to the Probing Algorithm (11), T, being the Probe of the Merging operation, should do so. The definition of label in (1) determines that only the label can access further computation. So, "he" can probe "John" for its referential valuation only if "he" is the label (in compliance with Head Algorithm 2). However, the derivation in which the Head Algorithm "wins" makes (13) uninterpretable under the relevant reading, because (13) would receive a nominal label, but it is not a DP. Given this "wrong" label, (13) cannot be interpreted as a clausal constituent, under the plausible assumption that syntax-semantics mapping rules dictates that CPs, and possibly TPs, but not DPs, have the semantic type of sentences[6]. This way the Principle C effect in (13) is reduced to a case of mislabeling.
On the other hand, if the Probing Algorithm "wins", T unproblematically provides the label but "he" cannot probe "John" (as indicated by the lack of coindexing in 14).

(14) He$_i$ likes John$_j$

Notice that the acceptability of (14) indicates that a pronoun is not forced to probe its sister node to get referentially valued by a matching Goal. Arguably, this introduces a difference with other cases of Probe-Goal relations. For example, T *must* (as opposed to *can*) search its sister for a matching DP category that values its φ-features. However, it is easy to see what is the basis for this difference between referential valuation and other cases of Probe-Goal matching. If a pronoun is not referentially valued by Probe-Goal matching, nothing goes wrong in the semantic component,

[6] In this paper, we cannot develop a comprehensive theory of the role of labeling at the syntax-semantics interface. However, it is a standard assumption that syntactic categories are mapped to a restricted set of semantic types. Although there is no rigid one-to-one mapping (i.e. CPs have different semantic types when they are independent sentences and when they are relative clauses), it is *not* normally assumed that anything goes. In particular, DPs don't have the semantic type *t* of sentences, banning special cases of propositional DPs (i.e. "He knows the time", which means "He knows what time it is"). Since propositional DPs are marked cases, we leave them for future research.

since this contains an independent procedure to assign a value to it, namely assignment functions to free variables. This mechanism is independently required for the many cases in which a pronoun does not receive a semantic value from the linguistic context. Therefore, it is so only natural that it applies whenever Probe-Goal matching does not take place and the pronoun is not valued in narrow syntax.
The next step is to show that this account does not extend inappropriately. Take (15) or (16) as representatives.

(15) He_i likes his_i friends

(16) $John_i$ likes his_i friends

If "he" and "his" were in a Probe-Goal relation in (15), the reading in which "he" and "his" have the same semantic value should be ruled out by the same reasoning that rules (13) out (namely, 15 should be another case of mislabeling). A similar problem would arise with (16), if "John" and "his" were in a Probe-Goal relation.
However, this problem does not arise, because the Probe-Goal relation is asymmetric. In every cases of Probe-Goal matching, intrinsic features of the Goal value those of the Probe. A referential expression like a proper name has intrinsic referential features (it has an assignment independent value), while a pronoun is not intrinsically referential (it has an assignment dependent value). So, "John" can be a Goal in (13) but "his" cannot be a Goal in either (15) or (16). Therefore, the mislabeling problem does not arise in (15) and (16). Of course, a legitimate question is how the relevant reading arises in (15) and (16). The importance of this question may not be completely apparent, because in a framework that assumes that indexes are freely distributed there is nothing special to say about (15) and (16). What happens is that "he" and "his" (or "John" and "his") receive the same indexes and this is the end of the story. But we are trying to avoid using indexes, in compliance with the Inclusiveness Condition. So, we must explain why the relevant reading arises in (15) and (16). We will do that in section 3.3.

3.2. When the canonical definition of Principle C and the definition in terms of mislabeling diverge

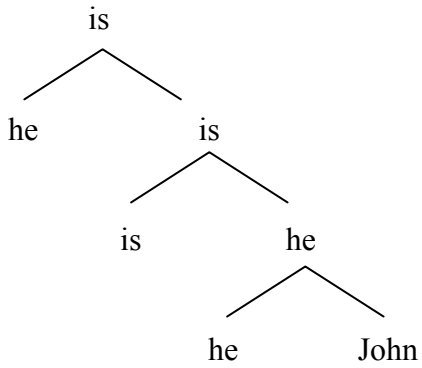
Up to now, we have been arguing that the approach to Principle C in terms of mislabeling is to be preferred on conceptual grounds, because Principle C would not be a primitive anymore and because any use of referential indexes would be avoided. In this paragraph, we try to make a case for the formulation of Principle C in terms of mislabeling to be empirically superior. We could identify two areas in which the two alternative approaches to Principle C make clearly divergent predictions.
The first domain is identity sentences like (17)[7]. The canonical formulation of Principle C make an embarrassingly wrong prediction, since (17) should be a patent violation of Principle C. In fact, it is perfectly OK.

(17) He_i is $John_i$

[7] We use the term "identity sentences" as a shortcut for sentences of the form [DP is DP]. Our analysis is orthogonal to the issue whether identity sentences are structurally distinct from other types of copular sentences.

Sentences like (17) are conveniently ignored in most discussions about Principle C[8]. The natural question is if the formulation of Principle C in terms of mislabeling fares any better than more canonical approaches. We will now show that it does. To see this, we have to focus on the initial step of the derivation of (17), when "he" and "John" are first merged. We will assume that, at least in the case of identity statements, the copula selects a headless small clause. Given this structure, both 'he' and 'John' can provide the label. Let us focus on the derivation in which 'he' does that. If 'he' "projects", it can search its sister node for a Goal that can value its unvalued feature. This Goal is 'John'. Given this derivation, the syntactic object created by merging 'he' and 'John' is a DP. Assuming that "he" later raises to T, (17) has the following structure**:**

(18)



The label at each step of the derivation is determined by algorithms (2) and (11) as follows:

(i) Label of {he, John} = label of "he" (by algorithm (2)) = D
(ii) Label of {is, {he, John}} = label of "is" (by algorithm (2) and (11)) = T
(iii) Label of {he, {is, {he, John}}}= label of "is" (by algorithm (11)) = T

[8] Heim and Kratzer (1998): 269-274 claim that identity sentences are in the same boat with "accidental coreference" cases like (i).
(i) Everyone likes John. Bill likes John, Mary likes John, Robert likes John. He_i likes $John_i$, too.
However, it is very dubious that (17) and (i) exemplify the same phenomenon. Rather special discourse contexts must be set-up to bring out the judgments that coreference is possible in (i) and similar cases but no special discourse context is required to make the same reading perspicuous in (17). Heim (1998) contains a related discussion. She elaborates on the well-known distinction (due to Frege 1892), between the proposition expressed by an identity statement and its cognitive value. For example, (17) has two reading. The first one is the tautological reading that states that John is identical to himself ($a=a$). The second reading ($a=b$) is more informative. Assuming that John can be associated to different guises (the guise "Bill's best friend", the guise "the guy I talked to you about", the guise "the person who is standing in front of me" etc.), the informative reading of (17) identify two different guises as being associated to the same person. Namely (17) says that the person of whom the interlocutor has a current visual impression is the same person (called "John") of whom the interlocutor carry in his/her memory an entry with various pieces of information. Schlenker (2006) elaborates on Heim's proposal to explain why the informative reading of (17) is not ruled out by Principle C. However, even if Heim's approach could be extended to the informative reading, it would have nothing to say about the tautological reading of identity sentences, which is *not* ruled out, contrary to what the standard formulation of Principle C predicts. Furthermore, as acknowledged by Schlenker, this approach runs into the risk of opening a Pandora's box. If we introduces guises to explain the absence of binding violations in identity sentences – why couldn't we *always* introduce different implicit descriptions to refer to a given individual, thus circumventing any kind of binding-theoretic violation?

The critical step is (iii). The crucial observation is that the unvalued referential feature of 'he' has already been valued in its base position. So "he" does not need to probe "John" at stage (iii) of the derivation and the algorithm (11) correctly dictates that the root gets a T label. Assuming that the copy of "he" in its base position is not interpreted, the structure in (18) reflects the fact that the copula identifies two categories which both have a DP label. As such, the structure in (18) is compatible with the semantic analysis of copular sentences proposed by Partee (1987), who proposes that in identity sentences an entity of type *e* is mapped onto the singleton set of entities identical with the individual. Thus, "John" is mapped onto the set of individuals who are identical with John.
Let us move to other cases of copular sentences, to double check if the approach that we are pursuing can account for them as well. First, let us focus on (19), in which "he" and "John" cannot have the same semantic value.

(19) *He_i is [the friend of $John_i$]

In order for the relevant reading to arise, "he" has to probe "John". This can either happen when "he" is first merged or when "he" raises to the Spec,T. No matter when "he" probes, a problem will arise, though. In fact, if probing is constrained by a locality requirement, the reading in which "he" and "John" have the same semantic value is blocked by an intervention effect, because the closest DP that "he" can probe is the DP "the friend of John", instead of the DP "John"[9]. This also explains why "he" and "the friend of John" can (in fact *must*, given the semantics of copular sentences) have the same semantic value (cf. 20).

(20) He_i is [the friend of $John_j$]$_i$

We think that the explanation in terms of intervention for the pattern in (19) and (20) is very intuitive. However, the concept of intervention is syntactic in nature. Therefore, this simple explanation can only be maintained if the referential valuation of the pronoun is the result of a syntactic operation, like Probing is. In this sense, the pattern in (19) and (20) is strong evidence for the approach that claims that referential valuation takes place as a result of a syntactic operation.
Finally, let us focus on a predicative copular sentence like (21). In these copular constructions, the obviation of Principle C effects observed with identity copular sentences is not observed.

(21) *He_i is [envious of $John_i$]

This can be explained in our approach as follows. In order for the relevant reading to arise, "he" has to probe "John". If this happens when "he" raises to the Spec,T, the familiar mislabeling problem, illustrated by sentence (13), arises since the sentence incorrectly got a D label. If "he" probes when it is first merged, a different mislabeling

[9] The fact that "he" and "John" cannot have the same semantic value can be reduced to a Relativized Minimality effect, if intervention is defined in terms of containment (in addition to the classical definition in terms of c-command, due to Rizzi 1990). Alternatively, the intervention effect exemplified by (19) can be expressed in terms of a violation of the *i*-within-*i* filter.

problem will arise, because the small clause formed when "he" is merged with "envious of John" will get a D label, rather than being an adjectival category[10].
Let us now move to the second area in which the approach to Principle C in terms of mislabeling and the traditional one make divergent predictions. This is exemplified by sentences like (22) and (23):

(22) My father voted for my father

(23) *He$_i$ voted for [my father]$_i$

The canonical definition of Principle C rules out both (22) and (23). On the other hand, our approach excludes (23) as a case of mislabeling but does not preclude (22). In fact, (22) does not contain any pronominal expression, therefore the reasoning based on referential valuation simply cannot apply to this case. We would like to argue that this consequence of our approach is welcome. It is certain that (22) is odd, probably because a grammaticalized way to express the relevant information exists, namely the sentence "my father has voted for himself". Still, the status of (22) cannot be equated to the status of (23). This becomes particularly clear in contexts that remove the oddity of (22) but cannot rescue the ungrammaticality of (23):

(24) In this election, each person voted for himself. This means for example that…
  a. √ My father voted for my father
  b. * He voted for my father

In this section, we have shown that there are at least two areas in which the canonical definition of Principle C is problematic, while our approach fares better.

3.3. Referential valuation, semantic binding and accidental coreference

We have introduced the notion of referential valuation, which we propose to be the result of the syntactic configuration of Probe-Goal matching. At the semantic interface, referential valuation is read as an instruction to exclude one class of assignment functions, namely those that do *not* assign the same individual to the category that evaluates (the Goal) and to the one that gets evaluated (the Probe).
We will now discuss how the notion of referential valuation relates with the notions of semantic binding and accidental coreference. Let us start from the previous one: A semantically binds B if A reduces the assignment dependency of B. Following Reinhart (2000), we will say that binding is the procedure of closing a property, which can be implemented as binding a free variable to a λ-operator, namely:

(25) A binds B iff A is the sister of a λ-predicate whose operator binds B.

[10] There is an interesting difference between Principle A and Principle C effects, as pointed out to us by Andrea Moro. "He" and "the friend of John" cannot have the same semantic value in (i), due to intervention, but an anaphor is licensed in a structurally similar configuration, cf. (ii).
(i) *He$_i$ is an [admirer of John$_i$]$_j$
(ii) He$_i$ is an [admirer of himself$_i$]
Although we cannot offer a comprehensive analysis of Principle A effects, this pattern is not a direct problem for our proposal, since we assume that Probe-Goal matching (which is sensitive to intervention) takes place only in (i), in which the proper name "John" is a suitable Goal. In (ii) no Goal is present, so Probe-Goal matching cannot take place.

In section 3.1, we left the question open of how "John" and "his" can have the same semantic value in sentence (16), repeated as (26), in absence of indexes that can be freely assigned to each of the two categories. The answer is that this reading arises if "John" semantically binds "his"[11]:

(26) John likes his friends
John (λx (x likes x's friends))

Of course, the same binding operation takes place in sentences like (27), in which the binder is a quantificational expression:

(27) [Every boy]$_i$ thinks that he$_i$ can win the competition

With this in mind, we have to go back to paradigmatic cases of Principle C effects like (13) because a serious problem for our theory seems to arise. Couldn't the illicit reading that we have excluded as a case of mislabeling *when Probe-Goal matching takes place* result from semantic binding, in absence of Probe-Goal matching? The standard assumption is that the binder must c-command the bindee and in Principle C configuration this does not happen, at least at Spell-Out. However, the proper name in (13) or the quantification expression in (28) in principle should be able to undergo QR and reach a position from which they c-command (and bind) the pronoun.

(28) *He$_i$ likes [every boy]$_i$

So, the question remains open of why semantic binding is impossible in (13) and (28). Of course, we are dealing with a well known problem, namely strong crossover. In the Government and Binding framework (cf. Chomsky 1981), strong crossover cases were reduced to Principle C violations, because the trace left by the binder was equated to a referential expression. However, we are *not* assuming Principle C as a primitive, so we cannot take the easy way to reduce strong crossover to Principle C. A natural explanation in terms of intervention is available, though. In fact, one can argue that what is wrong with (28) is that, after QR, the pronoun in the subject position creates an intervention effect for the chain that relates the quantificational expression and its copy[12]:

(29) *[Every boy]$_i$ he$_i$ likes ~~every boy~~
Binder intervener bindee

Let us close this section by briefly discussing (accidental) coreference. Two categories A and B corefer when they accidentally denote the same individual as a consequence of the value that the assignment function assigns to free variables.

[11] By treating (26) as a case of semantic binding, we assume that proper names can undergo QR. See Heim and Kratzer (1998), chapter 8 for motivation.

[12] The approach based on the intervention effects recasts the intuition underlying the Bijection Principle (cf. Koopman and Sportiche 1982), which states that what is wrong with strong (and weak) crossover configurations is that there is just one binder for two categories that need to be bound. However, a literal version of the Bijection Principle is problematic, for there are cases, like (i), in which an operator unproblematically binds two variables:
(i) [Every boy]$_i$ said that he$_i$ hates his$_i$ brother
The approach based on intervention effects can distinguish between (i) and (29), since only in the latter case a chain created by movement (Internal Merge) gets disrupted by an intervention effect.

Every theory has to limit the occurrence of accidental coreference, because, if accidental coreference could occur totally freely, it would not be possible to formulate any constraint on referential dependencies. In the literature, two ways have been proposed to block accidental coreference[13]. The first one is Lasnik's (1976). In fact, Lasnik's early formulation of Principle C, unlike the standard GB formulation in (12), prevents coreference rather than coindexing. We cannot follow Lasnik, because we want to dispense with Principle C altogether. However, we can assume the approach proposed by Reinhart (1983), based on Rule-I.

(30) Rule-I
NP A cannot corefer with NP B if replacing A, at LF, with a variable A-bound by B, yields an indistinguishable interpretation.

As discussed in the literature, Rule-I has the effect of blocking accidental coreference in canonical cases of Principle C violation like (13). This is all we need to assume.
Let us summarize. Principle C effects might be reduced to cases of mislabeling. If this proposal is on the right track, the status of Principle C changes. Instead of being a primitive (arguably, with a universal import), Principle C would be the result of the interaction between the two fundamental algorithms that govern phrase structure building. We argued that this reformulation of Principle C can be shown to be empirically superior to canonical formulations, most notably in the area of copular constructions.

4. Labels and movement

In the previous section we have discussed a case of conflict between the two labeling algorithms where only one of the two possible outputs is interpretable, and the other is excluded as a mislabeling at the interface. However another type of situation might in principle hold as well: one where, given a conflict, both possible outputs are acceptable and interpreted as (different) syntactic/semantic objects. This is what we shall discuss in the next section of the paper.

4.1 Does "Head Movement" exist in syntax?

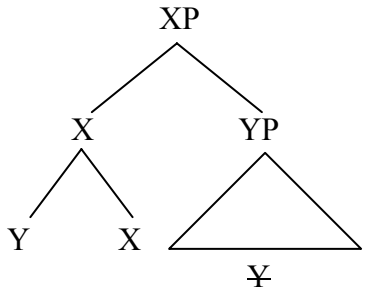
Recall we have been assuming the strong unification thesis that the operation responsible for movement and for structure expansion is the same: Merge, simply defined as an operation putting together lexical items or syntactic objects. One consequence that has been implicit throughout the paper is that any of the following sets should be available, both for Internal and for External Merge.

[13] Kayne (2005) explicitly assumes that there is nothing like accidental coreference. In this system, a pronoun *must* receive its semantic value from a (possibly null) antecedent and no space is left for pronouns in isolation. Coherently, for example, Kayne takes a sentence like "He is smart" with an unstressed *he* to be unacceptable in isolation, since *he* cannot receive a semantic value in his system. If we understand it correctly, Kayne's theory denies the existence of the mechanism of assignment function to free variables. We won't follow Kayne in such a radical departures from standard semantic assumptions since it opens various problems that are not explicitly addressed, i.e. the way in which *bound* variables are interpreted in absence of assignment functions.

(31) {LI, LI}[14] (order irrelevant assuming that the linear dimension
{SO, LI} falls within the phonological component)
{SO, SO}

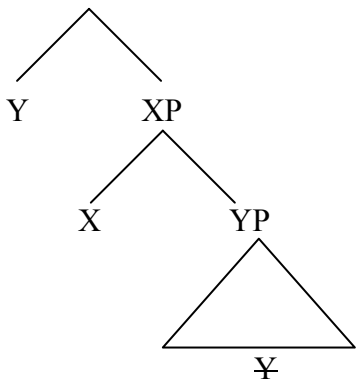
On the 'movement' side, this implies that what is traditionally called head movement and what is traditionally called phrasal movement should both be available to computation, *contra* recent attempts to ban head movement from syntax (e.g. Chomsky (2001)[15]. These approaches crucially rest on the assumption that head movement is not to be considered as syntactic in nature because it lacks the essential cyclic character of syntactic operations. This is certainly true of the standard head-adjunction configuration given in (32).

(32)



The traditional motivation for this configuration is that in core cases of head movement considered in the literature (V-to-T movement or V-to-C movement) the two heads conflate and behave as a single constituent, hence the assumption that they form a sort of a "derived lexical item" represented in the head adjunction configuration above. But this is not the only configuration head movement can in principle produce. Suppose we have an SO X endowed with an edge feature that needs to be valued. Nothing prevents in principle to internally merge (a copy of) a head (=LI) Y endowed with a matching feature to the root of the structure, as in (33).

(33)



The configuration in (33) is a structure obtained by merging a new item to the root of

[14] Actually this is not available to Internal Merge by definition: a lexical item cannot contain another lexical item to be internally merged to it.
[15] See Matushanski 2006 and Donati 2006 for a detailed criticism of this position showing that it holds on wrong assumptions, both theoretical and empirical, that it is incompatible with other aspects of Chomsky's theory (e.g. phases) and that it brings undesirable consequences.

the tree, hence complying with the extension condition. Given (33), the head conflation effect correlated with many cases of head movement can be the result of an independent process, perhaps phonological, call it affixation, which has nothing to do in principle with head movement: affixation is something that can happen to two adjacent heads independently from how and why they ended up being adjacent. (33) is the kind of head movement that is predicted by the definition of movement in terms of Internal Merge, and this is the kind of head movement we shall be considering here.

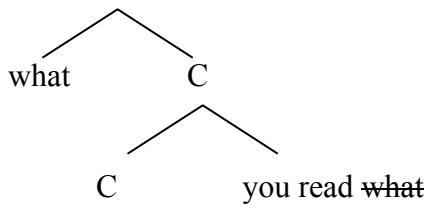
4.2. Labeling properties of head and phrasal movement

If we have two options available, head movement and phrasal movement, the question of what triggers the choice between the two becomes an interesting one. What we want to propose here is that it is the labeling algorithms that provide the relevant mechanism. But let us proceed step by step.

The most standard proposal for discriminating between the two movement options and accounting for their complementary distribution relies on locality differences: in a nutshell, head movement is claimed to be constrained by a specific locality condition (the Head Movement Constraint (Travis 1984)), and thus available only in a very restricted set of cases, namely when the Goal is a feature of the head of the complement of the Probe. However this account, which has been challenged on empirical grounds by many (Lema & Rivero 1990; Borsley et al 1996; Carnie 1995; Roberts 1994; Manzini 1994 a.o.) is really not an option given the minimalist approach to movement we are adopting. Recall that locality is a condition on the search procedure establishing Probe-Goal relations, which is only defined in terms of features: it is an unvalued feature the one which acts as a Probe, and it is a matching feature the one which gets searched as a Goal. Internal Merge is not part of this searching procedure, but really a separate though parasitic operation triggered by some extra mechanism (generalized pied-piping: Chomsky 2003, i.e. the need for the target to get its Edge feature valued: Chomsky 2005). As a consequence, the phrasal status of what is internally merged cannot be determined by the search procedure, nor by its locality constraints.

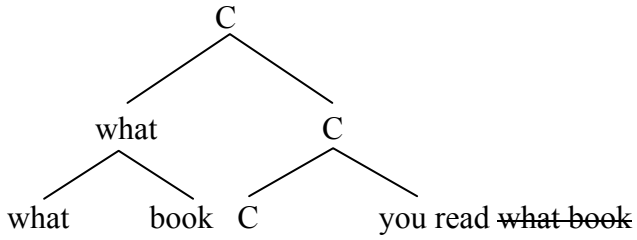
The solution to this problem lies in the labeling algorithms we have been discussing so far: while head movement and phrasal movement cannot be distinguished on the basis of the Probe-Goal relation they establish, they have very different effects on labeling. Consider again the labeling conflict (10), repeated here as (34), from which our discussion began:

(34) *what you read*



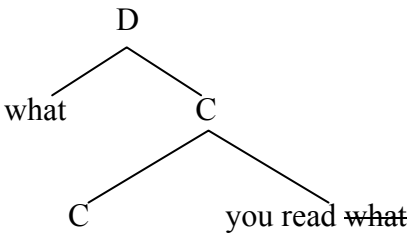
(34) is derived by internally merging a single lexical item “what” (head movement) to the edge of a clause, an option, as we have seen, that we have no reasons to exclude. The result is a conflict between the two algorithms we have been considering: by the Head Algorithm (2), “what”, being a lexical item, should provide the label. By the Probing Algorithm (11), “C”, being the probe of the merging operation, should do so. This kind of conflict never arises when a phrase is internally merged, as in (35).

(35) *what book you read*

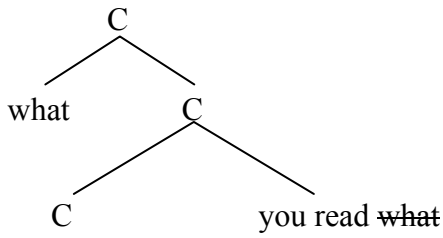


Here Merge holds between two SOs, and the Head Algorithm (2) is not relevant: by (11), it will be "C", the probe of the merging, which labels the entire construction.
The prediction is that the minimal difference between (34) and (35) should be reflected in the distribution and interpretation of the two structures: more precisely, (34) is predicted to have two possible labels, illustrated in (36), while the only labeling available to (35) has been given in the corresponding diagram.

(36) a.



b.



This prediction is reflected by the systematic ambiguity of a phrase like "what you read": it can be interpreted as a free relative and be embedded under a verb selecting a DP: (37).

(37) a. I read what you read/a book
b. I read the thing that you read

However, it can also be interpreted as an indirect interrogative clause, and be embedded under verbs selecting for clausal complements: (38).

(38) a. I wonder what you read/ if the sun will shine tomorrow
b. I wonder what book you read

These two readings and distributions correspond to the two labeling possibilities: in (36a) the clause gets the D category of 'what' and the structure of a (free) relative clause *à la* Kayne (1994) and Bianchi (1999); in (36b) the clause gets the C category of C and the structure of an interrogative clause.
Crucially, no ambiguity at all, neither in interpretation nor in distribution, arises when phrasal movement is involved: a clause corresponding to (35) can only occur in environments for clauses, and can only be interpreted as a simple interrogative, as shown in (39).

(39) a. I wonder what book you read
b. *I read what book you read

We are now in the position of answering the question we raised at the beginning of the paragraph: what is the difference between the two movements available to syntax? What we have been claiming here is that this difference has to do with their labeling consequences: while phrasal movement had no effect on the label of its target (by algorithm 11), head movement systematically gives rise to an extra labeling possibility (activating algorithm 2), in which the moved head "relabels" the target[16].

4.3 On the distribution of head and phrasal movement

We showed that in our system it is possible to distinguish between what is traditionally called head movement and what is traditionally called phrasal movement. However, we have not yet derived the complementary distribution of the two movement options. In particular, if what head movement does is generating more labeling possibilities, why isn't it always selected in any derivation, especially given its more minimal status? Let us try to answer this important question.

First of all, by what we have seen in the preceding section the alleged complementary distribution of the two types of movements is not that systematic: much of it is an illusion due to how phrase structure worked in pre-minimalist terms. In standard X-bar theory accounts, the head movement cases discussed above would be analyzed as phrasal movement cases, under the assumption that a simple lexical item cannot occupy a position where a phrase can sit. This way of looking at things is completely incompatible with the inclusiveness condition we have tried to adhere to in the paper. There is no way under this strict condition to assume an ambiguity in the phrase structure status of a head: a head is a lexical item, namely an element listed in the lexicon as such. The idea of a single lexical item being a phrase simply makes no sense, if phrases are defined as the syntactic objects obtained by merging *two* things[17]. This implies that each time we see a lexical item being displaced we will have to analyze it as head (=LI) movement.

We can illustrate this way of looking at things by reflecting on the distribution of clitics in Romance languages: while standardly analyzed as instances of phrasal movement[18], clitics are typically heads being displaced from their thematic position to an inflection-related position, as illustrated in (40). There is no reason (and no way) in the present approach not to analyze clitic movement as head movement.

(40) Maria lo conosce ~~lo~~.
Maria him knows him
'Maria knows him'

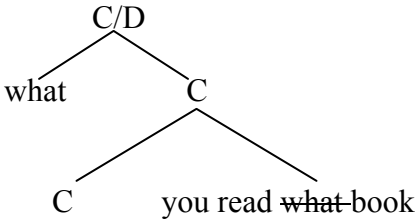
[16] Or better 'apparently relabels' the target: what we have been claiming here is that head movement simply expands the structure, adding an extra label to the label it merges with.

[17] Chomsky (1995) assumed that a head with the distribution of a phrase is an ambiguous element, being both minimal, since it is an LI, and maximal (he assumes a relational definition according to which a maximal projection is a category that does not project any further). Our system is simpler in that it dispenses with the relational definition of maximal projection.

[18] Clitics in Romance have always been a problem. Given standard assumptions on phrase structure, clitics are clearly 'maximal projections' in the position where they are generated, but they end up being afffixed to an inflectional head. This tension is at the root of standard and influential analyses such as Kayne (1989)'s, where the clitic starts up as a phrase but moves and adjoins as a head, changing its phrase structure status in the course of the derivation (violating a condition like that of the uniformity of chains, Chomsky 1995). A head movement analysis is by far more minimal.

Given this important proviso, we can just say that in canonical cases of wh-movement both head movement and phrasal movement can be displayed, in a distribution that is not complementary at all. On the other hand, only head movement of a wh-element is able to yield free relatives as the one discussed in relation to (34)[19].
However we still need to explain why head movement is not the *only* option in any environment, and in interrogatives in particular. More precisely, we need to explain why (41) is an impossible derivation.

(41) ** what you read book*



In (41) the lexical item "what" is extracted from the phrase it labels/heads, and internally merged to the root. By algorithm (11), "C", the probe, should provide the label and the result should be an interrogative clause. By algorithm (2), "what", an LI, should also provide the label and the result might also be a relative clause. The configuration should be ambiguous, but it is not. As shown in (42), it can neither be embedded under a context selecting for an interrogative clause nor under a context selecting for a relative clause: plainly, it is ungrammatical.

(42) a. *I wonder what you read book.
b. *I read what you read book.

There are at least two possible approaches to explain this restriction. One is to claim that (41) is an improper movement in that it extracts a subconstituent out of a constituent[20]. This would amount to exclude any instance of head movement that does not involve an intransitive head, in any context. This approach appears however to be too strong in the light of familiar cases of verb movement: V-to-T or T to C

[19] There is an (apparent) exception to this generalization: a class of free relatives which appear to allow phrasal Wh- movement, illustrated in (i) and (ii), respectively for English and Italian:
(i) Mangerò [qualunque pane] vorrai [t]
I-will-eat whatever bread you-will-want
(ii) I shall visit [whatever town] you will visit [t]
The fact that this movement pattern correlates with the presence of some extra-material in the head of the clause, the suffix-*ever/-unque*, suggests an obvious solution, namely that these relatives are only apparently " free", i.e. defective, but rather correspond to full relative clauses. As such, they are generated as the complement of an external determiner (Kayne 1994, Bianchi 1999) , the universal quantifier *-ever/-unque*.
(iii) I shall visit $[_{DP}$ $[_{D}$ ever] $[_{CP}$ $[_{DP}$ what town] […… ..]]]
For some reason due to its universal value (Larson 1987), *unque/ever* triggers the raising and head adjunction of the wh-determiner, yielding (iv).
(iv) $[_{DP}$ $[_{D}$ what $[_{D}$ ever]] $[_{CP}$ $[_{DP}$ ~~[what]~~ town] […… ..]]]
See Battye 1989 for a series of empirical arguments demonstrating that these relatives ought to be treated as "pseudo free relatives". See also Kayne 1994:154n for a similar analysis proposed on totally different grounds. The same analysis extends naturally to other 'maximalizing relatives' (Grosu 2002) like (v), whose interpretation is only compatible with the presence of a silent *ever*-type determiner.
(v) I will read what books you will tell me.

[20] This sub-extraction might be an illicit case of intervention, if intervention is defined as containment: see footnote 8.

movements are exactly extractions of a head out of its constituent, at least under standard accounts[21]. From this point of view, (43) provides a very interesting contrast.

(43) a. *I wonder what you read ~~what~~ book
b. Did you ~~did~~ read that book?

Observing the contrast in (43) we might elaborate an alternative capitalizing on what makes the difference, namely the environment of head extraction: in (43a) the head is extracted from a DP, while in (43b) it is extracted from TP. There must be something special in DP that makes it difficult to extract its head. Notice in fact that such extraction, although difficult, is not completely impossible: the literature on Wh-movement is full of cases like the ones illustrated by French sentence (44) and by German sentence (45c):

(44) Combien as-tu lu de livres
How-NEUT have-you read of books
'How many books have you read?'

(45) a. Welche Bucher hat Johann gelesen?
Which-PL books has Johann read

b. *Welche has Johann Bucher gelesen?
Which-PL has Johann bokks read

c. Was hat Johann fur Bucher gelesen?
What-NEUT has Johann for books read
'Which books has Johann read?'

These data have been analyzed in a number of different ways in the literature, all assuming phrasal wh-movement given the standard X bar theory restrictions we discussed above. In our system, we shall analyze all these cases as instances of licit head movement of a wh-determiner out of its DP. From this point of view the data above suggest that only ***some*** determiners are disallowed to move out of their DP, and the relevant factor appears to be agreement: to illustrate, in German the only case when a wh-head can move out of its constituent is when it does not agree with its associate: in (45a-b) 'whelche' agrees with its NP and moving it as a bare head is impossible; in (45c) 'was' does not and head extraction is OK. A similar contrast is visible in Italian (46).

(46) a. Quanto sono alti?
How-NEUT-SING are tall-MASC-PL.
'How tall are they'?

b. Quanti libri hai letto?
How-MASC-PL books have.2SG read

[21] An alternative is the remnant movement approach, by which what really moves is the entire VP only containing the verb because all other constituents are extracted: see Koopman (1995), Mueller (1998), a.o.

c. *Quanti hai letto libri?
How-MASC-PL have.2SG read books
'How many books have you read?'

In (46a) 'quanto' does not agree with the AP 'alti' and moving it as a bare head is the only option (*quanto alti sono* is deviant) possible. In (46b-c) the same head does agree with its complement and its extraction is disallowed.
Why should agreement have this freezing effect? Suppose agreement in DP is a local Probe-Goal relation between the D (its unvalued phi-features) and the N. If it is so, the pattern in (44)-(46) might be described by saying that when Probe-Goal matching (in agreement features) takes place in the base configuration, it cannot be destroyed by moving the Probe. A weaker alternative is to derive this freezing effect from a spell out condition in the spirit of D'Alessandro and Roberts (2007), who assume that members of an agreement chain must be spelled-out in the same phase.
Exploring all the consequences of this approach in details goes far beyond the scope of this paragraph and requires further research. However, no matter which technical implementation turns out to be more optimal, it is clear that head movement of the D head out of the DP is possible in principle (although it is sometimes blocked by an interfering factor). This makes movement of the D head out of the DP parallel to more familiar cases of "head" movement (say, T out of TP and v out of vP). In turn, this reinforces the conclusion that the distribution of "head" movement is much less limited than it is usually assumed, a conclusion that allows a unified theory of phrasal and head movement operations.
Summarizing, the ambiguity of sentences involving bare wh-words, which are compatible with both the distribution of interrogatives and with the distribution of free relatives, can be reduced to cases of conflicts between the two labeling algorithms proposed here. Since they arise only when head movement in involved, this way of looking at things provide a new understanding of the very nature of this typology of movement as opposed to phrasal movement.

5. General conclusion

One persistent goal of the research in syntax in the last fifteen years has been the attempt to simplify phrase structure building rules. The aim was to preserve the empirical coverage of X-bar theory by dispensing with its rich apparatus. A first step has been Kayne's (1994) approach, in which much of X-bar theory was reduced to a single axiom (Linear Corresponding Axiom). A further step was Bare Phrase Structure theory, which starting from the version in Chomsky's (1995) has undergone various reformulations until Chomsky's (2005) version, in which only two algorithms govern phrase structure building. In this paper, we critically re-examined these two algorithms and claimed that the algorithm that dictates that a lexical item transmits its label when it is merged with another object (we called it the Head Algorithm) conforms to minimalist assumptions. A second algorithm proposed by Chomsky (2005) does not obey minimalist requirements because it is specifically restricted to movement configuration and, by doing so, it does not allow reduction of movement to (Internal) Merge. Therefore, we replaced the second algorithm with what we called the Probing Algorithm, which holds equally for Internal and External Merge. In addition to capturing core cases of phrase structure building, the Head and the Probing Algorithms allowed us to shed light on two distinct areas, namely

Principle C effects and the syntax of *wh* constructions, which we analyzed as cases of conflict between the two algorithms. What unifies these two configurations is the fact that a lexical item (which should become the label in compliance with the Head Algorithm) is merged with a syntactic object that, being the probe of the operation, should become the label in compliance with the Probing Algorithm. In one case, this conflict produces two alternative outputs (a question or a free relative) that are both legible at the syntax/semantics interface. In Principle C configurations, one of the resulting output (the one determined by the Head Algorithm, in which the pronoun and the referential expression have the same semantic value) produces an object that is not interpretable at the syntax/semantics interface. This way, Principle C effects are reduced to cases of mislabeling, with no need to postulate a specific condition to rule them out.

We hope to have shown that the simplification of the apparatus, in addition to complying to minimalist assumptions, can reinforce the deductive power of the theory. In particular, in this paper we carried out a simplification of phrase structure theory rules that allowed us to adopt the very same explanation for two apparently unrelated phenomena, such as constraints on the interpretation of pronouns and the categorial status of *wh* constructions.

References

Battye, A. (1989). Free relatives, pseudo-free relatives and the syntax of CP in Italian. *Rivista di linguistica* 1: 219-250.

Bianchi, V. (1999). *Consequences of antisymmetry. Headed relative clauses*. Berlin: Mouton de Gruyter.

Borsley, R, M.L. Rivero and J. Stephens (1996). Long head movement in Breton. In R. Borsley and I. Roberts (eds.), *The syntax of the Celtic languages: a comparative perspective*, 53-74. Cambridge: Cambridge University Press.

Carnie, A. (1995). Nonverbal predication and head movement. Doctoral Dissertation, MIT.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht : Foris.

Chomsky, N. (1995). "Bare Phrase Structure," in Gert Webelhuth (ed.) *Government and Binding Theory and the Minimalist Program*. Cambridge: Blackwell, 383-439.

Chomsky, N. (2000), "Minimalist Inquiries: The Framework." In *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, edited by Robert Martin, David Michaels and Juan Uriagereka, 89-155. Cambridge, MA: The MIT Press, 2000.

Chomsky, N. (2001), "Derivation by Phase" in M. Kenstowicz (ed.), *Ken Hale: A Life in Language* 1-52, Cambridge, MA: The MIT Press.

Chomsky, N. (1993), ``A Minimalist Program for Linguistic Theory." In K. Hale and S. J. Keyser, eds., *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, MA: The MIT Press, 1-52.

Chomsky, Noam. (1995), "Bare Phrase Structure", in Gert Webelhuth (ed.) *Government and Binding Theory and the Minimalist Program*. Cambridge: Blackwell. 383-439.

Chomsky, N. (2005). On Phases. Manuscript, MIT.

Chomsky, N. (2006). Approaching UG from below. Manuscript, MIT.

Cinque, G. (1990). *Types of A'-dependencies*. Cambridge, Mass., The MIT Press.

Collins, C. (2002). Eliminating Labels. In Samuel Epstein and Daniel Seely (eds.), Derivation and Explanation in the Minimalist Program. Blackwell. 42-61.

D'Alessandro and Roberts. (2007). Movement and agreement in Italian past participles and defective phases. Manuscript, University of Cambridge.

Donati, C. (2006). On *Wh*-head movement. In L. Cheng and N. Corver, eds., *Wh-movement: moving on*, Cambridge, Mass: The MIT Press, 21-46.

Fox, D. (1999) *Economy and Semantic Interpretation*. Cambridge, MA: The MIT Press.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50.

Grosu, J. (2002). Strange relatives at the interface of two millennia. Glot International 6, 145-167.

Heim, I. (1998). Anaphora and Semantic Interpretation: A Reinterpretation of Reinhart's Approach. In U. Sauerland and O. Percus (eds.), *The Interpretive Tract*. MIT Working Papers in Linguistics 25, 205-246.
Heim, I and A. Kratzer (1998) *Semantics in Generative Grammar*, Oxford. Blackwell.
Hornstein, N. (2006). Pronouns in a Minimalist Setting, manuscript, University of Maryland.
Kayne, R. (1989). Facets of Romance past participle agreement. In P. Benincà (ed.), *Dialect variation and the theory of grammar*, Dordrecht: Foris, 85-105.
Kayne, R. (1999). *The antisymmetry of syntax*, Cambridge, MA: The MIT Press.
Kayne, R. (2005). Pronouns and their Antecedents, In S. Epstein and D. Seely (eds.), *Derivation and Explanation in the Minimalist Program*, Malden, MA: Blackwell, 133-166.
Koopman, H. (1995). On verbs that fail to undergo Verb-Second. *Linguistic Inquiry* 26:137-163.
Koopman, H. and D. Sportiche (1982). Variables and the Bijection Principle. In J. Kaye, H. Koopman and D. Sportiche (eds.), *Projet sur les langues kru: Premier rapport*, Université du Québec à Montreal, 176-202.
Lema, J. and M. L. Rivero (1990). Long head movement: ECP vs. HMC. In J. Carter, R:M: Déchaine and K. Kusumoto (eds.), *Proceedings of NELS* 28, Amherst, University of Massachusets, GLSA, 219-234.
Matushansky, O. (2006). Head movement in linguistic theory, *Linguistic Inquiry* 37: 69-109.
Moro, A. (1997) *The raising of predicates: predicative noun phrases and the theory of clause structure*, Cambridge: Cambridge University Press.
Moro, A. (2000) *Dynamic Antisymmetry: movement as a symmetry breaking phenomenon,* Cambridge, MA: The MIT Press.
Müller, G. 1998. *Incomplete category fronting*. Dordrecht: Kluwer.
Partee, B. 1987. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jong, and Martin Stokhof, eds., *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers,* Dordrecht: Foris, 115–143.
Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
Reinhart, T. (2000). Strategies of anaphora resolution. In Hans Bennis, Martin, Everaert, and Eric Reuland (eds.), *Interface strategies*, Amsterdam: Royal Academy of Arts and Sciences, 295–325.
Reuland, E. (2001). Primitives of Binding. *Linguistic Inquiry*32: 439–492.
Roberts, I. (1994). Two types of head movement in Romance. In D. Lightfoot and N. Hornestein (eds.), *Verb Movement*, Cambridge: Cambridge University Press, 207-242.
Rothstein, S. (2001) *Predicates and their Subjects*. Dordrecht: Kluwer.
Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: The MIT Press.
Schlenker, P. (2003). Non-Redundancy: Towards a Semantic Reinterpretation of Binding Theory. *Natural Language Semantics*13: 1-92.
Travis, L. (1984). Parameters and effects of word order variation, Doctoral dissertation, MIT.