Luigi Rizzi*

Labeling, maximality and the head – phrase distinction

Abstract: Combining the labeling algorithm of Chomsky (2013) with bare phrase structure raises the question of how heads (simple or complex) and phrases can be distinguished. I propose a notational device which draws the distinction in a way which solves technical problems for the labeling algorithm. Focusing on phrasal movement, I show how the "halting problem" for wh-movement, and in particular the freezing effects arising in criterial positions, can be derived from labeling and a maximality principle, restricting movement to maximal elements with a given label. Looking then at head movement, I argue that it can be made consistent with the No Tampering condition, and work out the labeling algorithm for structures derived by head movement. Finally, I argue that the ban against excorporation in head movement can be analyzed as a case of freezing, and traced back, much as freezing in phrasal movement, to the maximality principle relativized to the head – phrase distinction.

Keywords: Syntax, Labeling, Freezing, Head Movement, Maximality

1 Introduction

A standard assumption throughout the history of generative grammar is that syntactic representations are hierarchical structures expressible as labeled bracketings, or trees. The labels of the pairs of brackets, or of the nodes in the tree, are the names of syntactic constituents. While labels of nodes are automatically provided by X-bar theory in more traditional approaches, a system based on recursive merge as the fundamental structure-building device requires a labeling algorithm. The labeling algorithm introduced in Chomsky (2013) capitalizes on the distinction between heads and phrases: heads, but not phrases count as potential labelers of structures created by merge; for instance, when a verb and a nominal expression are merged, the new constituent is labeled by the verb as a verbal projection, a VP in traditional notation.

Under current assumptions, syntactic representation are "bare", in the sense that they do not express bar level distinctions, as in Bare Phrase Structure (BPS:

***Corresponding author: Luigi Rizzi,** University of Geneva, Geneva, Switzerland; University of Siena, Siena, Italy, E-mail: luigi.rizzi@unige.ch

Chomsky 1995: Ch. 4). BPS is, in turn, a consequence of the Inclusiveness Condition, stating that the computational system can only see and use properties expressed in the lexical items entering the computation, without introducing new specifications. So, current conceptions of phrase structure differ in at least two respects from traditional X-bar theoretic accounts: syntactic representations are bare, and they are labeled by an algorithm distinct from the structure-building device (merge). Now, technical problems arise when these two ideas are combined: the labeling algorithm requires that the distinction between heads and phrases be readily available to single out potential labelers, but representations based on bare phrase structure do not express the head – phrase distinction.

In this paper, after an illustration of the functioning of the labeling algorithm (Sections 2 and 3), I would like to introduce a notational device which expresses the head-phrase distinction in a way consistent with the Inclusiveness Condition (Section 4). I will then turn to phrasal movement and review recent contributions on the "halting" problem for wh-movement, the fact that stepwise successive-cyclic movement is forced to continue from certain positions, while it is forced to stop in other positions, which give rise to freezing effects. The halting and freezing positions are criterial positions, defined by heads such as Q, Foc, Top, etc., expressing scope-discourse properties (Rizzi 1997). A maximality principle, stating that only maximal objects with a given label can be moved, interacts with the labeling algorithm to capture the freezing effects (Section 5). In the last part of the paper I turn to head movement, which can be made consistent with a slightly modified No Tampering condition (Section 6). At first sight, maximality bans head movement, as it only allows movement of maximal objects (maximal projections, in terms of traditional X-bar notation); in fact, if the principle is relativized to the head-phrase distinction, head movement becomes consistent with maximality (Section 7). Moreover, maximality offers a principled explanation for an important property of head movement, the ban against excorporation: when a head is incorporated into another head, only the derived complex head can be moved further, and no excorporation of the moved (or host) head is possible (Section 8). The ban against excorporation and the freezing effects on phrasal movement can thus be seen as two sides of the same coin: both are derived from the maximality principle relativized to the head – phrase distinction.

2 On the labeling algorithm

I will assume, following Chomsky (2013, 2015), that syntactic trees must be uniformly labeled at the interfaces. So I will assume the following well-formedness constraint to hold:

(1)   Uniform labeling: at the interfaces, a tree must be completely labeled.

Why should (1) hold? One possible motivation has to do with selection. If selectional requirements (including categorial selection, in the sense of Grimshaw 1978) are checked at the interface with semantics under strict locality (sisterhood), labels must be present at that level. More generally, uniform labeling could be a consequence of interpretive principles, which may need labels to properly interpret structure. Intuitively, this makes sense: a DP, a VP and a CP are interpreted differently, and interpretive principles may be sensitive to the "canonical structural realizations" of semantic types.[1]

    The second assumption that I will borrow from Chomsky (2013) is that the labeler of a category created by Merge is the closest head:

(2)   *α* created by merge receives the label of the closest head

In Rizzi (2015a) I have proposed that the notion "closest head" can be understood in terms of familiar intervention locality:

(3)   *α* created by merge receives the label of head H_1 such that:
                        I. *α* contains H_1, and
                        II. there is no other head H_2 such that
                                            i. *α* contains H_2, and
                                            ii. H_2 c-commands H_1.

---

1 The point may be considered controversial, though, as standard interpretive systems assumed in formal semantics (Heim and Kratzer 1998) typically do not use phrasal labels. There could also be other reasons for (1). Once a structural chunk has been computed, the complex object created by syntactic computation is put in a memory repository, where it can be consulted and used by performance systems. The encoding of a node in memory may be optimally efficient if there is a name for that node, the label, so that an unlabeled node may be "unstable" at the interface.

    Previous approaches to labeling stipulated that labeling was necessary to permit further applications of merge, i.e., that merge only applied to labeled structures. So, in the worst case, in a system specifying (1) we would have that a stipulation replaces another stipulation. The two systems clearly differ in empirical predictions: (1) permits the syntactic computation to continue even if a node has remained unlabeled, and labeling can be delayed till the end of the phase. This possible delay is crucially capitalized by the system to permit salvaging strategies, with clear empirical consequences which we will discuss later. From now on, I will assume (1) to hold, and will not further explore its motivation. See Cecchetto and Donati (2010, 2015) for different assumptions on the labeling algorithm.

In plain words, a head is the labeler of a given node when there is no other head which intervenes between the head and the node, where intervention is expressed in the usual hierarchical terms of c-command. (3) builds intervention locality, precisely defined in terms of Relativized Minimality (Rizzi 1990), into the labeling algorithm. A more elegant solution would have the algorithm refer to locality stated as an independent principle. This can be achieved by appealing to the notion of minimal configuration (Rizzi 2004):

(4)   X is in a minimal configuration with Y with respect to local relation R only if there is no Z such that
         i. Z c-commands Y and Z does not c-command X, and
         ii. Z is of the same type as X with respect to R.

We can then define the labeling algorithm as

(5) α receives the label of a head contained by α, and in a minimal configuration with it.

Relation R here is the relation between a category created by merge and a potential labeler, a head contained in it. A given head is in a minimal configuration with α when there is no other element of the same type, a potential labeler, i.e., another head, which intervenes between the given category and the given head.

3 Interactions with types of merge

The algorithm interacts with the typology of merge. Let us first see how it works by using an informal notation which, much as traditional X-bar theory, encodes the distinction between heads and phrases.
   There are three cases of merge to consider:

I. Head – Head Merge:

(6)
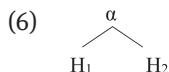
This is the case in which two elements are taken from the (functional or contentive) lexicon and combine. (6) is already problematic for labeling as each head would prevent the other one from being the closest head to α.

Chomsky, op. cit. suggests that this cases of primary merge may be restricted to the merger of an unlabeled lexical root with a categorizing functional head (n. v, a: see Marantz 2013 and much related work): as only the latter has a category to contribute, there is no competition and the categorizing head wins (this would mean that H in (3) and "a head" in (5) should be understood as "a head with a label"). Further assumptions may be needed to cover other imaginable cases of primary merge (e.g., when two elements are taken from the functional lexicon and are merged together, e.g., a determiner and a number specification in French: *le* + *s* 'the + Pl')); I will not discuss such cases here. Notice that we want to be able to say that a complex object formed by a root and a categorizing functional head still counts as a head for selection, labeling, attraction of movement, etc. This is made possible by the formalism worked out below.

II. Head − Phrase Merge:

(7)

$$\alpha$$
$$H_1 \quad Phrase_2$$
$$H_2$$
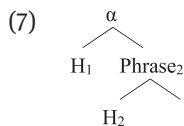
Here things are straightforward: H_1 is closer to α than H_2 (or any other lower head) hence α gets the label of H_1. So, for instance, when T is merged with an AspP, α is labeled by T, as AspP is not a head, hence it is not taken into account, and its head Asp is too far away to interfere:

(8)   (a) α (b)   TP

T   AspP   ➔   T   AspP

Asp   Asp
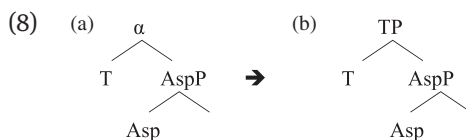
This is the standard case of recursive merge, which in traditional terms of X-bar theory yields [$_{VP}$V DP], [$_{AspP}$ Asp vP], [$_{CP}$ C TP], etc.
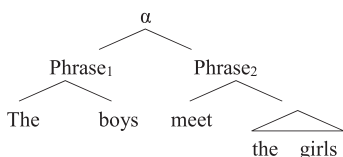
III. Phrase − Phrase Merge:

Merge must be able to combine two phrases already formed by previous applications of merge yielding a configuration like the following:

(9)

$$\alpha$$
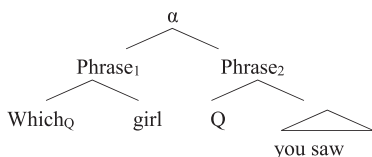$$Phrase_1 \quad Phrase_2$$
$$H_1 \qquad H_2$$
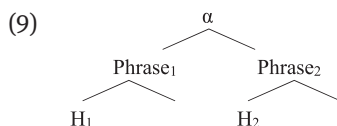
This configuration may arise both through external and internal merge. A case of external merge is the merger of an external argument and a predicate (a vP), both of which may be of arbitrary complexity:

(10) External merge:

```
                    α
              ┌─────┴─────┐
         Phrase₁        Phrase₂
          ┌──┴──┐      ┌───┴───┐
        The   boys   meet    ╱──────╲
                             the   girls
```

A case of phrase – phrase internal merge is provided by any instance of phrasal movement, e.g., wh-movement:

(11) Internal Merge:

```
                    α
              ┌─────┴─────┐
         Phrase₁        Phrase₂
          ┌──┴──┐      ┌───┴───┐
       Which_Q  girl   Q     ╱──────╲
                            you saw __
```

As far as labeling is concerned, in case of Phrase – Phrase merge, the situation is ambiguous, as both H_1 and H_2 in (9) qualify as the closest head to the new node created by merge (both are in a minimal configuration with the node, according to (5)), so the algorithm gives inconsistent indications in (9), and α remains unlabeled. But this can only be a temporary state of affairs: under Uniform Labeling (1), α must receive a label before being passed on to the interpretive systems. So, something must happen here to make labeling possible. Chomsky (2013) envisages two devices to achieve labeling here:

1. Movement of one of the two phrases: if in (9) one of the two phrases moves further, the head of the remaining phrase remains without a competitor, and labels α. This is what happens in (10): the external argument moves from its thematic position, and the head of Phrase2 labels α (here, as vP). This method of salvaging the structure is inspired by Moro's (2000) dynamic antisymmetry approach, in which movement can salvage a structure which would otherwise disallow linearization, under dynamic antisymmetry (an approach inspired in turn by Kayne 1994). See Rizzi 2015a: 326 on why this salvaging strategy for labeling is consistent with the copy theory of traces.
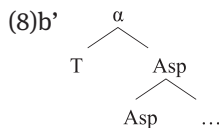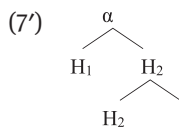
2. The creation of a criterial configuration, in the sense of Rizzi (1997): in such configurations, Phrase1 and Phrase2 agree with respect to a criterial feature, a feature expressing a scope-discourse property: Q, Top, Foc, etc. Here both phrases (and their heads) give consistent indications, the criterial feature has categorial status, and α gets labeled accordingly; in (11), Phrase1 and Phrase2 are headed by an element bearing Q, hence α gets labeled as Q, i.e., a question.

Before coming back to the two devices permitting labeling in the Phrase – Phrase configuration, let us consider certain technical problems which are raised by the labeling algorithm if it is combined with bare phrase structure.

4 Distinguishing heads and projections

Bare phrase structure (BPS) in its original version (Chomsky 1995: Ch. 4) does not distinguish between heads, intermediate projections, maximal projections: there is just one type of label used throughout. This is in compliance with the Inclusiveness Condition, according to which the computational system does not add properties which are not already specified in the lexicon: so, categorial labels are admitted, as they are inherited from lexical items, while bar levels are not.

But now, in order for the labeling algorithm to work properly, we need a way to distinguish between heads and projections. Otherwise, cases like (7) would really look like (7′), and the concrete case (8)b would be like (8)b', with a shape analogous to (6)

(7′)
$$
\begin{array}{c}
\alpha \\
\diagup\diagdown \\
H_1 \quad\; H_2 \\
\diagup\diagdown \\
H_2
\end{array}
$$

(8)b'
$$
\begin{array}{c}
\alpha \\
\diagup\diagdown \\
T \quad\; Asp \\
\diagup\diagdown \\
Asp \quad \ldots
\end{array}
$$

A simple distinction between "simple" and "complex" objects (objects created by merge) would not suffice, as we want to be able to express the fact that some heads may be complex objects (see below).

A first approximation to draw the head-phrase distinction within BPS could be the following:

(12)   An element drawn from the lexicon is a head, everything else is a phrase.

Let us structure a bit this way of expressing the head-phrase distinction. Elements which are going to be merged with other elements can be taken from three repositories:
1.   the lexicon (functional or contentive)[2]
2.   a temporary work space containing a structure already built by merge;
3.   a second temporary workspace, containing another structure built by merge.

If merge is a binary operation there is no need for any other temporary repository (whereas if merge were n-ary, one would need n such repositories; so, presumably binary merge is the most economical structure building device with the necessary expressive power). Head-head merge takes two elements from 1, head-phrase merge takes one element from 1 and the content of 2 (or 3), phrase – phrase merge takes the content of 2 and the content of 3.
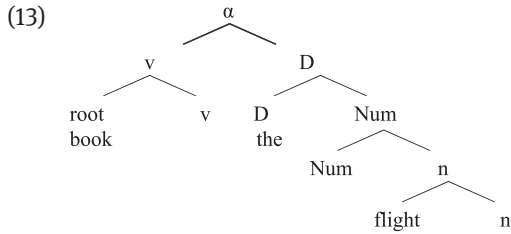
Definition (12) would work for (8)b': T is drawn from the functional lexicon, whereas its sister node Asp already is a complex syntactic object formed by previous applications of Merge, so Asp is not a head here and T has no competitor.

Nevertheless, there are more complex cases in which (12) is not general enough because we may want a syntactic object which has already undergone merge to count as a head, a complex head. Consider for instance a phrase in which the verb has been formed by merging v and a lexical root as in (13) (part of a sentence like *John will book the flight*). We want the complex entity *book* + *v* to count as a head here, capable of selecting an object DP and of labeling its mother node as vP (in informal notation):[3]

---

2 Or, possibly, the "numeration", a preliminary choice of the lexical items preceding syntactic computations, in the sense of Chomsky (2000).

3 Here I have inserted the Num(ber) head into the object DP just to represent the rich functional structure of DP's, as it emerges from cartographic research (see Cinque 2002 and, for an updated overview, Rizzi and Cinque 2015).

As an alternative to (13) one may consider the option of merging the object DP directly with the unlabelled root *book*; but in that case, the structure would be incorrectly labelled D, as the root, by definition, could not have a label to provide. So, I assume that unlabeled roots must first be merged with categorizing functional heads, with the complex category root + x inheriting the label of x and the selectional properties of the root (but see the proposal in fn. 5 for certain cases of compounds).

(13)

$$\alpha$$

```
           α
        /     \
      v           D
    /   \       /    \
  root   v    D      Num
  book       the    /    \
                  Num      n
                  /  \
              flight    n
```

But here, the problem would arise of distinguishing (13) from (9) (or any other configuration created by Phrase – Phrase merge), which under bare phrase structure would look like the following:

(14)

```
           α
        /     \
      H₁        H₂
    /   \      /   \
  H₁          H₂
```
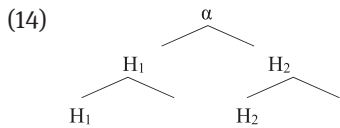
Clearly, definition (12) does not suffice to distinguish between (13) and (14) (or the concrete cases of (14) such as (10) or (11)).

Still, I think the idea that a head is an element drawn from the lexicon can be used in a more indirect way.

I will assume the following notational device:

(15)   An element drawn from the lexicon bears the feature "lex"

So, now a head is a category with the lex feature. When a lex category undergoes merge with another category, the lex feature may project with the categorial feature, or not. In the former case we get a complex structure labeled with a lex category, a complex lexical item, a complex head; in the latter case we get a non-lex category, a phrasal category. For concreteness, consider the case of a lexical root merged with a categorizing functional head.

(16)

```
        v_lex
      /      \
  root_lex    v_lex
  book
```
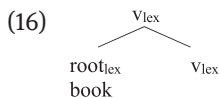
Here both v and the root are lex, as they are drawn from the lexicon. v wins the competition, as the root has no categorial feature; the category created by merge can be labeled as lex. We thus get a derived lexical item with label v_{lex},, which

will function as a head in further computation. It can undergo head – phrase merge and be combined with a direct object:

(17)
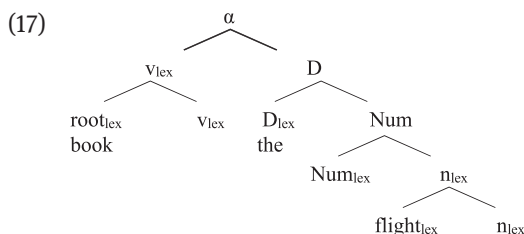
Under what condition is feature lex passed on to the mother category in the labeling process? The simplest assumption seems to be that the inheritance of lex is optional. Of course, the option will be constrained by well-formedness principles. In particular, it typically is the case that complex heads do not contain phrasal material: for instance, an element of a compound cannot be productively modified (e.g., "three truck drivers" can mean "three drivers of trucks", but not "drivers of three trucks", etc.). Let us state this as a uniformity condition:[4]

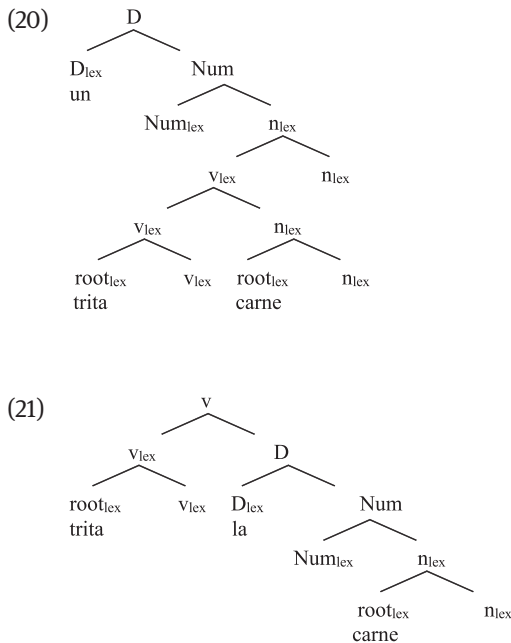(18)   Lexical uniformity: a lex category cannot contain non-lex material

i.e., heads can be made very complex through repeated applications of merge of lex material, but as soon as the labeling algorithm does not transmit the lex feature to the mother node, the structure leaves the head zone, and enters the phrasal zone: at that point it cannot come back to being a complex head, a property reminiscent of the cyclic principle, Adriana Belletti observes (p.c.). The lex feature thus demarcates the zone of the tree in which syntactic processes apply "below the word", at the sublexical level, and above the word, at the phrasal level. Within X-bar theory, the distinction between sub-lexical and phrasal syntax can be expressed by bar levels, e.g., by indicating affixes which are heads but not complete lexical items with "negative" bar levels, X^{-1}, as in Rizzi and Roberts (1989). The system proposed here has no bar levels, but the highest category bearing lex in a tree demarcates the sublexical and the phrasal zone.

---

4 Such complex nouns like *je-m-en-foutiste* (French), *menefreghista* (Italian) ("I don't care – ist"), containing a whole sentence followed by the nominal affix *–iste, -ista*, clearly are frozen exceptions to be listed in the lexicon.

As an illustration, compare a V-N compound, such as (19)a, with a regular verb phrase such as (19)b in Italian:

(19)   a   Questo strumento è un trita carne
           'This instrument is a grind meat = a meat grinder'
       b   Questo strumento trita la carne
           'This instrument grinds the meat'

The two structural representations are roughly as follows:

(20)

$$
\begin{array}{c}
\text{D} \\
\diagdown \\
\text{D}_{lex}\quad\quad \text{Num} \\
\text{un} \\
\quad \text{Num}_{lex}\quad \text{n}_{lex} \\
\quad\quad \text{v}_{lex}\quad \text{n}_{lex} \\
\quad \text{v}_{lex}\quad\quad \text{n}_{lex} \\
\text{root}_{lex}\ \text{v}_{lex}\ \text{root}_{lex}\ \text{n}_{lex} \\
\text{trita}\quad\quad \text{carne}
\end{array}
$$

(21)

$$
\begin{array}{c}
\text{v} \\
\text{v}_{lex}\quad\quad \text{D} \\
\text{root}_{lex}\ \text{v}_{lex}\ \text{D}_{lex}\quad\quad \text{Num} \\
\text{trita}\quad\quad \text{la} \\
\quad\quad \text{Num}_{lex}\quad \text{n}_{lex} \\
\quad\quad \text{root}_{lex}\quad \text{n}_{lex} \\
\quad\quad \text{carne}
\end{array}
$$

In (20) *trita carne* is a complex noun, hence dominated by a lex node (and containing only lex material, because of lexical uniformity); in (21) *trita la carne* is a verb phrase, dominated by the phrasal node v.[5]

---

5 One may ask the question of how the compound is labeled in (20). Clearly, the compound *trita carne* is a noun, and still its semantic head is the verbal part *trita* (grind), assigning a thematic role to the nominal part *carne*, much as the plain verb *trita* does in the phrasal structure (21); the result is obtained by brute force in (20), by merging the compound with the categorizing head n. A simpler alternative would be to assume that *trita* is a bare lexical root which is merged with the noun *carne* yielding

Going back to (17): the complex verbal element formed by merge as in (16) is lex, hence a head. When it is merged with the DP *the flights*, it wins the competition (its sister node is not a head), hence it labels α as v, a vP in informal notation. Here the feature lex cannot be passed on to α because of lexical uniformity, as α contains phrasal material (the object DP, in informal notation).

Why couldn't the object DP in (17) in fact be D_{lex}, a complex head, with the lex feature being passed on all the way up to the projected D node? I assume that functional elements quite generally select phrases, not heads as complements (but not always of course, as the categorizing heads v, n, a select lexical roots). If D_{lex} has the property of selecting a phrase, Num must be phrasal, and at that point the tree enters the phrasal zone, and it cannot bear lex anymore. Hence the projected D node cannot be lex, and when the derivation reaches (17) the root node cannot be lex because of lexical uniformity.

Analogously, (8)b' would have representation (8)b''':

(8)b'''
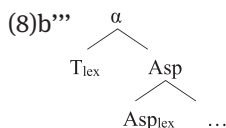
Here the Asp node is phrasal because of lexical uniformity (it contains phrasal material, the vP), so that T_{lex} is the closest head to α, hence α is labeled as T. It cannot be lex because of lexical uniformity, as it contains phrasal material, the AspP, in informal notation.

In conclusion the lex feature provides a device to distinguish heads from projections, thus making the labeling algorithm consistent with bare phrase structure. The lex mechanism is consistent with Inclusiveness, as the computational system does not introduce any specification not already contained in the lexical elements (in fact, the option is to lose the lexical specification lex: that the computational system may not carry over the whole set of lexical specifications of the head is fully consistent with Inclusiveness).

___

(i)
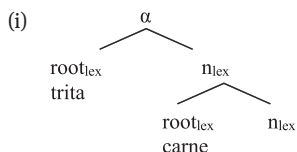
at this point, as the bare root *trita* does not provide any label, α is labeled by the closest head with a label, which is $n_{lex.}$ The result that the compound is a noun is thus straightforwardly obtained.

5 Labeling and the "halting problem" for wh-movement

Wh-movement is successive cyclic because of locality (Chomsky 1973). But in some cases, wh-movement necessarily continues from an intermediate C-system to a higher one, while in other cases it necessarily stops, and there are freezing effects: the "halting problem" for wh-movement, in the terminology of Rizzi (2015a). Once a particular C-system is reached, whether movement must continue, or must stop, depends on the nature of the selecting verb.

A verb like *think*, selecting a declarative complement, requires movement to continue, i.e., the intermediate movement step (22)b cannot surface as such, and movement must proceed to the main C-system, yielding (22)c:[6]

(22) a   John thinks [C_{decl} [Bill read [which$_Q$ book]]]
      b   * John thinks [$_\alpha$ [which$_Q$ book] [C_{decl} [Bill read ___]]]
      c   [$_\beta$ [which$_Q$ book] [Q [John think [$_\alpha$ ___ C_{decl} [Bill read]]]

Chomsky (2013) captures the necessary continuation of movement in (22)b through labeling: if *[which$_Q$ book]* stops in the embedded C-system, an XP-YP configuration is created, and a labeling problem arises for α. As C is a declarative complementizer here, a criterial configuration cannot be created, hence the only possibility is that the wh-phrase continues to move. After movement has taken place, α can be labeled as C_{decl}, a declarative clause. The main clause category β in c now forms a criterial configuration (both *which book* and the clause headed by Q share the criterial feature Q), hence β can be labeled as Q, a main question.

The mirror image effect is observed when the embedded clause is the complement of a verb selecting an indirect question:

(23) a   John wonders [Q [Bill read [which$_Q$ book]]]
      b   John wonders [$_\alpha$ [which$_Q$ book] [Q [Bill read ___]]]
   c   * [$_\beta$ [which$_Q$ book] [Q [John wonders [$_\alpha$ ___ C [Bill read]]]

---

6 Configurations akin to (22)b are possible in languages permitting "partial wh-movement", typically with an unmarked wh-element occurring in the main C-system. See McDaniel (1988), Dayal (1994), and Boskovic (2008b), among many other references. I will not address this case here. Ian Roberts (p.c.) raises the question of why Decl never enters a criterial configuration. One possibility is that there is no Decl feature, and "declarative" is the default interpretation. Roberts observes that this idea may be supported by the fact that there are no "reverse" residual V2 languages, i.e., languages with V2 in declaratives but not in questions.

Here the embedded complementizer is Q, hence when movement applies a criterial configuration is created in (23)b, and α can be labeled as Q, an indirect question. But here not only is it the case that wh-movement can stop: it must stop, there is a freezing effect, as the ill-formedness of (23)c shows.[7]

In order to capture the freezing effect by capitalizing on the labeling idea, it is proposed in Rizzi (2015a) that one could appeal to the familiar fact that phrasal movement must involve maximal projections (i.e., in terms of classical X-bar theory, we have DP movement but not D' movement, AP movement, but not A' movement, CP movement, but not C' movement, etc.: typically one cannot move the X' constituents stranding the respective specifiers). In terms of BPS, a maximal projection must be understood dynamically, as the maximal node with a given label. So, the observed restriction on movement can be captured by a maximality principle like the following:

(24)   Maximality: only maximal objects with a given label can be moved.

Consider now the representation of (23)b after labeling has applied

(25)   wonder....
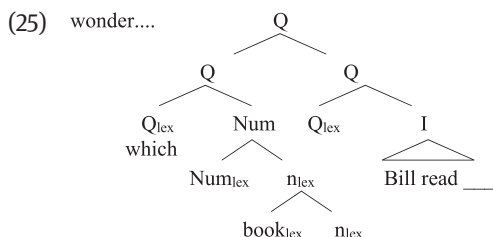
After labeling of the clausal node as Q has taken place, *which book* ceases to be a maximal node: under the dynamic interpretation of maximality enforced by BPS the whole clause now is the maximal node with label Q. So, under the maximality principle (24) *which book* ceases to be a freely movable element: only the maximal node, the clause, can be moved at this point, e.g. to be topicalized:[8]
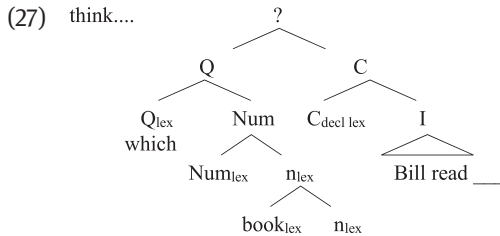
---

7 The freezing effect in simple cases like (23)c could be naturally amenable to an "inactivation" analysis along the lines of Bošković (2008a), or to an interpretive filter at LF, but these solutions are not straightforwardly applicable to the more complex cases discussed in Rizzi (2006), and Rizzi and Shlonsky (2007) and subsequent work, in which two distinct criterial features are specified in the moved phrase.

8 The syntactic object *which book* bears other features in addition to Q, e.g., it presumably is a DP. So, in order to derive the freezing effect from maximality we must understand this principle as crucially requiring maximality of the criterial feature. So, the assumption is that maximality is required for each categorial feature (hence including criterial features, which in the criterial approach are categorial).

(26)   Which book Bill read, I really don't know __

The freezing effect thus follows from the labeling algorithm, under maximality.

No problem with maximality arises in cases like (22). Here, after the first application of wh-movement, the representation is:

(27)   think....

```
                          ?
                  ┌───────┴───────┐
                  Q               C
              ┌───┴───┐       ┌───┴───┐
          Q_lex     Num    C_decl lex   I
          which   ┌──┴──┐          ┌────┴────┐
              Num_lex   n_lex    Bill read ___
              ┌────┴────┐
          book_lex    n_lex
```

In (27) the node "?" cannot be labeled because of the non-criterial XP-YP configuration. *Which book* must move in order to permit labeling of "?" as a declarative clause; and in fact *which book* can move further, under maximality, because it is the maximal node labeled Q.[9]

There is a timing issue here, a point raised by an anonymous reviewer. If labeling can be delayed in general, why can't it be delayed in (25) as well? If labeling of the clausal node as Q could be delayed there, *which book* would be movable as a maximal element, and the explanation of the freezing effect via maximality would be voided. So a delay must be ruled out in (25). In Rizzi (2015a: 330) I assume that labeling applies in accordance with Pesetsky's Earliness Principle (see Pesetsky and Torrego 2001: 400), i.e., as soon as it can apply. So, labeling applies in (25) as soon as the criterial configuration is created by internal merge, hence further movement of *which book* is excluded by maximality, as desired. On the contrary, labeling of the clausal node cannot take place in the non-criterial XP – YP configuration of (27), hence further movement of *which problem* is possible and takes place, thus solving the labeling problem for the clausal node.[10]

---

9 Representation (27) raises the issue of how intermediate steps of wh-movement are triggered. If they are triggered by an uninterpretable Q feature in the intermediate C-system, we may assume, following Chomsky (1995), that the uninterpretable Q feature is deleted once it is checked, hence it has no impact on labeling. In general, it seems to be the case that only interpretable features, giving rise to criterial configurations, are taken into account by the labeling algorithm. If intermediate movement is not feature-triggered, and only the final step is triggered by the criterial feature, the issue does not arise.

10 The reviewer also observes that in principle also the YP constituent (the projection of C) could be moved from (27) under maximality. This is correct, but other reasons would rule out

Labeling, interacting with maximality, thus offers a comprehensive solution to the halting problem for wh-movement. This approach was extended in Rizzi (2015a, 2015b) to capture fixed subject effects, treated as criterial freezing effects in subject position. If the (high) subject position is a criterial position, it is a possible "halting" site for subject movement; much as other halting positions, it also is a freezing position, so that a phrase moved there cannot be moved further. This captures fixed subject effects such as that-trace effects, which are also amenable to an explanation in terms of labeling and maximality. See also Rizzi (2015b), and Shlonsky and Rizzi (2015) for extensions of the same ideas to other case of criterial freezing, primarily in the low focus position (Belletti 2004) in inverse copular sentences (Moro 1997) and other constructions. See Chomsky (2015) for an approach to fixed subject effects also capitalizing on labeling, similar in spirit to the one presented here and in Rizzi 2015b, but not relying on maximality.

6 Head movement and No Tampering

The maximality principle, as stated, proscribes head movement: a head is not the maximal node with a given label, hence head movement is excluded, much as the movement of an intermediate phrasal projection. This may be seen as a welcome result: head movement raises problems for the No Tampering condition, as the derived structure it creates modifies the structure already constituted by merge, an unexpected state of affairs under No Tampering, so that ruling out head movement in principle may seem desirable.

Nevertheless, the empirical evidence for head movement is robust and varied: a verb can pick up various inflectional specifications (of agreement, tense, aspect, etc.), and proceed all the way to C as a bare element, i.e., without carrying any dependent (complement or specifier). Assuming that all such cases are cases of phrasal movement in disguise (remnant movement)

---

this option. Suppose that the constituent *C Bill read* is topicalized from (27), yielding something like the following:

(i)   [C Bill read] I think [$_\alpha$ [what$_Q$ book]–]

Among other problems that would arise here, α would be labeled as Q, thus violating the selectional restrictions of *think*.

  Boskovic (2015) assumes a timing of labeling such that X-YP configurations are labeled immediately, while labeling of XP-YP is delayed. As far as I can tell, this assumption is consistent with the Earliness approach which I have adopted. On a distinct timing issue arising in the context of head movement, see Shlonsky (2015).

raises the problem of how we can make sure that all the dependents of the moved head can be moved out from the phrase, so that the head remains alone in the phrase to be moved, in order to properly mimic head movement. And the problem is worsened in cases of successive head movement (say, V to T to C): at each movement step, independent applications of movement should evacuate all the material contained in a projection except the head. Another possible approach is to assume that head movement exists *qua* movement of the head alone, but it takes place in the PF branch of the grammar (Chomsky 1995; see also Boeckx and Stjepanovic 2001), hence it is not a core syntactic phenomenon. But see Roberts (2010) for detailed evidence that head movement affects interpretation in ways that would not be expected under a PF approach (see also Lambova 2002; Lechner 2005).

Here I will continue to adopt the traditional assumption that head movement (Head – Head internal merge) exists as a core syntactic phenomenon, as distinct from phrasal movement (Phrase – Phrase internal merge), and will explore the consequences of this assumption for labeling. Let us first take a closer look at the status of head movement w.r.t. the No Tampering condition, stating that the structure created by merge cannot be modified.

A minimal modification of the condition can be envisaged which would make the it consistent with head movement:

(28)   No Tampering (revised): The complement of the probe cannot be modified

Formulation (28) permits modification of the probe itself, while necessarily preserving the structure of its complement. This seems to me to make sense conceptually. The fundamental motivation of No Tampering seems to be to reduce computational complexity by making structures already computed unmodifiable, so that the computational system can exclude a priori a number of conceivable operations, and does not overload its memory resources (this is the fundamental rationale of the notion of cycle in a bottom up derivation: at each stage, the only things that can happen, happen at the root, and the rest of the structure remains unchanged). If this is so, it makes sense to keep the probe, which has just entered syntax and is in the "focus of attention" of computation, accessible to modification. So, according to (28), movement can target the whole root structure (phrasal movement), or just the root head, the probe, and this permits head movement. It should also be noticed that a system based on (28) still captures two fundamental results of classical No Tampering with respect to movement: 1., the fact that movement is always to a higher position in the tree (when a probe-goal relation is established, the goal can only be attached to the probe, or to the whole structure: no lowering is permitted); and 2. the copy theory of traces ("movement" cannot be radical displacement

because that would modify the complement of the probe, so a full occurrence of the "moved" element must remain in the complement). As stated, (28) refers to internal merge; whether or not it can be extended to external merge (hence, merge tout court), depends on whether external merge can also be stated as involving a preliminary probe-goal relation (as is argued in Cecchetto and Donati 2015); I will not pursue this issue any further here.

7 Head movement and maximality: the role of lex

Consider now cases of head movement in connection to labeling. For instance, T_{lex} (or, more plausibly, some lower inflectional head), attracts v_{lex} in (29)a, yielding (29)b (here the moved head can be attached to the attracting head without violating the revised No Tampering condition):

(29)　a　T_{lex}　　$[[_{vlex} root_{lex} v_{lex}] DP] \rightarrow$
　　　　b　$[_\beta [_{vlex} root_{lex} v_{lex}]$　$T_{lex}] [<[_{vlex} root_{lex} v_{lex}]> DP]$

How is the complex head β, created by movement, labeled here? Clearly, we want T_{lex} to win the competition.

(30)

$$\beta$$

$$v_{lex} \quad T_{lex}$$

$$root_{lex} \quad v_{lex}$$
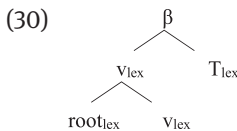
Here both T_{lex} and v_{lex} are heads, according to our assumptions, so in principle they compete for labeling. But a difference between T_{lex} and v_{lex} is that the former is a simple head, drawn from the functional lexicon, while the latter is a complex head, built via merge: we may assume that in such circumstances the simple head wins the labeling competition.[11]

---

11 Another difference between the two competing heads is that the sole occurrence of T_{lex} c-commands all the occurrences of v_{lex}, while not all the occurrences of v_{lex} c-command T_{lex}, so T_{lex} is in this sense closer to β than v_{lex}, and wins the labelling competition. Notice that this way of computing intervention would offer an alternative option to address the labeling problem in, e.g., (8)b': the sole occurrence of T c-commands all the occurrences of Asp, whereas not all the occurrences of Asp c-command T; for this case, this particular way of computing intervention may thus be an alternative to the notational device we have discussed in Section 4. It should also be noticed that the approach mentioned in this footnote does not cover case (13), so that it cannot be seen as a global alternative to the lex approach.

A different approach to this formal problem would capitalize on the fact that attractor and attractee share a feature, and this feature projects: so, maybe the attracted v has a T feature, which ensures that v will be attracted by T. Then we would have:

(31) a T_{lex} [[$_{vlex,T}$ root$_{lex}$ $v_{lex,T}$] DP] →

   b   [$_\beta$ [$_{vlex,T}$ root$_{lex}$ $v_{lex,T}$] T_{lex}] [<$_{vlex,T}$ root$_{lex}$ $v_{lex,\ T}$]> DP]

Let us consider, for more clarity, the derived structure of head movement under this view:

(31)b'

$$
\begin{array}{c}
\beta \\
\diagup\diagdown \\
v_{lex,\ T} \quad T_{lex} \\
\diagup\diagdown \\
root_{lex} \quad v_{lex,\ T}
\end{array}
$$
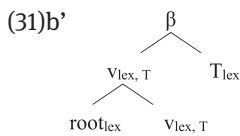
Here both H1 and H2 would share the feature T, which would then project (in a way akin to what happens in criterial configurations, in which the same feature is also shared by both elements undergoing merge), and the complex head resulting from head movement would then be labeled as T.
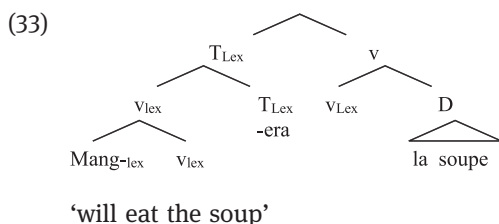
Whatever mechanism is adopted for the proper labeling of β in (30), the complex head thus created can further be head-moved to C, and then the new complex head will be labeled as C (or Fin, in a cartographic representation of the C-system as in Rizzi 1997), with the familiar properties of head movement (respecting the Mirror Principle of Baker 1988, etc.).

As was mentioned above, an issue arises for head movement once the maximality principle is introduced. Why is head movement possible in the first place, under maximality? Consider, for an illustration, v to T movement in French, as in Pollock (1989):

(32)

$$
\begin{array}{c}
\diagup\diagdown \\
T_{Lex} \quad\quad v \\
\text{-era} \quad \diagup\diagdown \\
v_{lex} \quad\quad D \\
\diagup\diagdown \quad \diagup\diagdown \\
mang\text{-}_{lex} \quad v_{lex} \quad la \ \ soupe
\end{array}
$$

'will eat the soup'
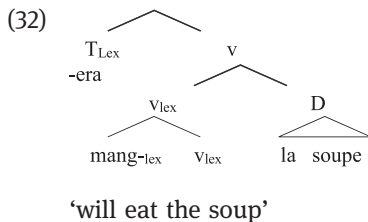
here *mang-* is not the maximal v node, so how can it move alone to T, to form the inflected verb? Clearly, all occurrences of head movement violate an unqualified version of the maximality principle.

If head movement exists, a natural possibility to make it consistent with maximality is to capitalize on the lex feature. Head movement may be possible because what gets moved is the maximal lex category. So maximality is relativized to the lex/non-lex distinction, and head movement applying in (32) can yield (33):

(33)

$$T_{Lex}\qquad\qquad v$$

$$v_{lex}\qquad T_{Lex}\quad v_{Lex}\qquad D$$
$$\text{-era}$$

$$\text{Mang-}_{lex}\quad v_{lex}\qquad\qquad\qquad \text{la soupe}$$

   'will eat the soup'

And then the complex T_{Lex} thus created may move further to C, e.g., in questions

(34)   Mangera-t-il la soupe?
       'Will he eat the soup?'

Here again the maximal lex category T_{lex} is moved, in accordance with the (relativized) maximality condition.

8 No excorporation

A familiar property of structures created by head movement is the "no-excorporation" prohibition (see Roberts 2001 for discussion):

(35)   No excorporation: When a head H_1 is incorporated into a head H_2, neither can be excorporated.

I.e., after incorporation, the only possibility is that $H_1 + H_2$ moves further via head movement. This is illustrated, for instance, by the fact that when the negative marker is cliticized onto an auxiliary verb in English, it cannot be stranded, but must be taken along if the auxiliary moves to C in a question:

(36)   a   John has not left
       b   Has John __ not left?
       c * Has not John __ __ left

(37)   a   John hasn't left
       b   Hasn't John __ left?
       c * Has John __ n't left?

i.e., in (36), where *not* has not cliticized, and presumably is in the Spec of a negative phrase, (36)c is excluded because *has not* is not a constituent; in (37) the negative element has cliticized onto the auxiliary, and it must be taken along, as in (37)b, and cannot be excorporated, as in (37)c.

Analogously, if a complement clitic has cliticized onto an inflected verbal element in French, the complex head cl + V is moved to C as a whole, as in (38)b, and the complement clitic cannot be stranded, as in (38)c:

(38)   a   Il lui a donné un livre
            'He to + him has given a book'
       b   Lui a-t-il __ donné un livre?
            'To + him has he given a book?'
       c   * A-t-il lui__ donné un livre?
   'Has he to + him given a book?'

And, quite generally, if a verb is associated with some lower inflectional specification, say an aspectual specification, then it cannot be excorporated to reach alone a higher specification, say T, but the whole v + Asp complex must move further, as in (39)d:

(39)   a   ... T... Asp... v
       b   ... T... v + Asp... __
       c   *... v + T... __ + Asp... __
       d   ... v + Asp + T... __... __

The ban against excorporation now looks very much like a case of freezing; so, it is natural to try to relate it to the same explanatory principle responsible for freezing with phrasal movement.

In fact, if the lex feature is taken into account, only the maximal lex node will be movable, under maximality.

For instance, for cases like (38) we would have (omitting many details):

(40)

    'He to + him has ... '
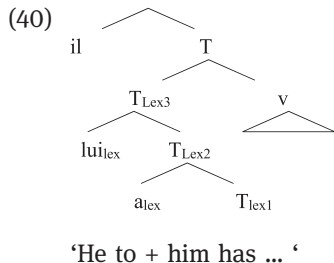
At this point, T_{Lex} is attracted to C, and by maximality, the maximal T_{lex} (hence T_{lex3} in (40)) must be attracted, whence no excorporation.[12] The same account generalizes to the other cases of no-excorporation.

More generally: there are two kinds of labels: X and X_{lex}. Maximality is relativized to the kind: if the attractor attracts X_{lex}, the maximal X_{lex} must move; if the attractor attracts X, the maximal X is moved. Non maximal elements, at the X and X_{lex} level, are inaccessible to movement, whence the freezing effects of phrasal movement in criterial positions, and the ban against excorporation for head movement.

9 Conclusion

X-bar theory encoded the distinction between head, intermediate projection and maximal projection in terms of bar levels. Bare phrase structure radically simplified the representational system by using uniform labels derived from the lexical specifications, hence consistent with the Inclusiveness Condition. Nevertheless, the labeling algorithm of Chomsky (2013, 2015) seems to crucially need the distinction between head and projection, as only the former can act as a labeler; and the analysis of criterial freezing effects in terms of maximality in Rizzi (2015a, 2015b) requires the distinction between maximal and non-maximal projections. As for the latter distinction, it can be expressed in dynamic terms within bare phrase structure: the maximal projection is the maximal node with a given label; this natural assumption makes it possible to capture the freezing effects in criterial configurations through maximality. As for the head – phrase distinction, I have introduced a notational device consistent with Inclusiveness: an element taken from the lexicon bears the feature lex, which may be passed on to a higher node through labeling.

---

12 The case illustrated in the text involves no excorporation of the host head (T in this case). An anonymous reviewer raises the question of how excorporation of the incorporating head (e.g., the auxiliary, or the clitic, in (40)) is banned. One natural possibility is that lex is considered a full-fledged component of the label: hence under maximality, once a complex head has been created, only the maximal element bearing lex, the whole complex head, can be (further) moved.

An alternative to head movement based on remnant phrasal movement might consider deriving the ban against excorporation as a Left Branch violation, as the remnant moved constituent would always be on a left branch. But Left Branch effects vary a lot (for instance no violation is detected in Smuggling cases, in the sense of Collins 2005), while the ban against excorporation looks very stable (but see Roberts' (2010) analysis of clitic climbing). So, deriving the effect from an apparently inviolable principle, like maximality, may be advantageous.

This identifies the heads as potential labelers, and permits the constitution of complex heads through merge. Internal merge of heads (head movement) is made consistent with a modified version of the No Tampering condition. The maximality principle is relativized to the head – phrase (lex – non-lex) distinction, a step which makes head movement consistent with it: so under maximality, the maximal lex (= head) or non-lex (= phrase) categories are the only licit targets of movement. Maximality relativized in this way captures the ban against excorporation from complex heads, which is thus assimilated to the freezing effects at the phrasal level.

References

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: Chicago University Press.

Belletti, Adriana. 2004. Aspects of the low IP area. In Luigi Rizzi (ed.), *The structure of CP and IP: The cartography of syntactic structures*, vol. 2, New York: Oxford University Press.

Bošković, Željko. 2008a. On the operator freezing effects. *Natural Language and Linguistic Theory* 26. 455–496.

Bošković, Željko. 2008b. On successive cyclic movement and the freezing effect of feature checking. In Jutta M. Hartmann, Veronika Hegedüs & Henk van Riemsdijk (ed.), *Sounds of silence: Empty elements in syntax and phonology*, 195–233. Amsterdam: Elsevier North Holland.

Bošković, Željko. 2015. On the timing of labeling: Deducing comp-trace effects, the subject condition, the adjunct condition and tucking in from labeling. Ms. University of Connecticut.

Boeckx, Cédric & Sandra Stjepanović. 2001. Head-ing toward PF. *Linguistic Inquiry* 32. 345–355

Cecchetto, Carlo & Caterina Donati. 2010. On labeling: Principle C and head movement. *Syntax* 13. 241–278.

Cecchetto, Carlo & Caterina Donati. 2015. *(Re)labeling*. Cambridge, MA: The MIT Press.

Chomsky, Noam. 1973. Conditions on transformations. In S. Anderson & P. Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt Rinehart and Winston.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries. In R. Martin, D. Michaels & J. Uriagereka (eds.), *Step by step: Minimalist syntax in honor of Howard Lasnik*, 3, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2008. On Phases. In R. Freidin, C. P. Otero & M. L. Zubizarreta (eds.), *Foundational issues in linguistic theory. Essays in honor of Jean-Roger Vergnaud*, 291–321. Cambridge MA: MIT Press.

Chomsky, Noam. 2013. Problems of Projection. In *Lingua*, 130, Special Issue "Core Ideas and Results in Syntax". 33–49.

Chomsky, Noam. 2015. Problems of projection: Extensions. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, strategies and beyond – studies in honour of Adriana Belletti*, 3–16. Amsterdam & Philadelphia: John Benjamins.

Cinque, Guglielmo. 1999. *Adverbs and inflectional heads*. New York: Oxford University Press.

Cinque, Guglielmo. (ed.). 2002. *The structure of CP and DP*. New York: Oxford University Press.

Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8. 81–120.

Dayal, Veneeta. 1994. Scope marking as indirect wh-dependency. *Natural Language Semantics* 2. 137–170.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: The MIT Press.

Lambova, Mariana. 2002. On A'-movements in Bulgarian and their interaction. *The Linguistic Review* 18. 327–374

Lechner, Winfried. 2005. Interpretive effects of head-movement. Ms. University of Tubingen (lingBuzz/000178).

Marantz, Alec. 2013. Verbal argument structure: Events and participants. *Lingua* 130(Special Issue) "Core Ideas and Results in Syntax".

McDaniel, Dana. 1989. Partial and multiple Wh-movement. *Natural Language and Linguistic Theory* 7(4). 565–604.

Moro, Andrea. 1997. *The raising of predicates*. Cambridge: Cambridge University Press.

Moro, Andrea. 2000. *Dynamic antisymmetry*. Cambridge, MA: MIT Press.

Pesetsky, David & Esther Torrego. 2001. T-to-C Movement: Causes and Consequences. In M. Kenstowicz (ed.). *Ken Hale: A Life in Language*, 355–426. Cambridge, Mass.: MIT Press.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry* 20. 365–424.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337, Dordrecht: Kluwer.

Rizzi, Luigi. 2004. Locality and left periphery. In Adriana Belletti (ed.), *Structures and beyond*. 223–251. New York: Oxford University Press.

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In L. Cheng & N. Corver (eds.), *On Wh movement*. 97–133. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2015a. Cartography, criteria, and labeling. In Ur Shlonsky (ed.), *Beyond the functional sequence*, 314–338. New York: Oxford University Press.

Rizzi, Luigi. 2015b. Notes on labeling and subjects. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, strategies and beyond – studies in honour of Adriana Belletti*, 17–46. Amsterdam & Philadelphia: John Benjamins.

Rizzi, Luigi & Guglielmo Cinque. 2015. Functional categories and syntactic theory.To appear *Annual Review of Linguistics*.

Rizzi, Luigi & Ian Roberts. 1989. Complex inversion in French. *Probus* 1. 1–30.

Rizzi, Luigi & Ur Shlonsky. 2007. Strategies of subject extraction. In H.-M. Gärtner & Uli Sauerland (eds.), *Interfaces + Recursion = Language? Chomsky's minimalism and the view from syntax-semantics*, 115–160. Berlin: Mouton de Gruyter.

Roberts, Ian. 2001. Head movement. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 113–147. Oxford: Blackwell.

Roberts, Ian. 2010. *Agreement and head movement*. Cambridge, MA: The MIT Press.

Shlonsky, Ur. 2015. A note on labeling, berber states and VSO order. In Sabrina Bendjaballah, Noam Faust, Nicola Lampitelli & Mohamed Lahrouchi (eds.), *The form of structure, the structure of form*, 349–360. Amsterdam: John Benjamins.

Shlonsky, Ur & Luigi Rizzi. 2015. Criterial Freezing in small clauses and copular constructions in Italian and Hebrew. Ms., University of Geneva, University of Siena.