

# Head Movement in Linguistic Theory

*Ora Matushansky*

In this article, I address the issue of head movement in current linguistic theory. I propose a new view of the nature of heads and head movement that reveals that head movement is totally compliant with the standardly suggested properties of grammar. To do so, I suggest that head movement is not a single syntactic operation, but a combination of two operations: a syntactic one (movement) and a morphological one (m-merger). I then provide independent motivation for m-merger, arguing that it can be attested in environments where no head movement took place.

*Keywords:* head movement, c-selection, Agree, Merge, Head Movement Constraint, cyclicity

## 1 Introduction

In order to consider issues of head movement in linguistic theory, we must first have a definition of a head. The standard Principles-and-Parameters view, repeated in Chomsky 1994, is that a head is defined as a terminal node (which automatically entails that in order to be distinguished from terminals, nonterminal nodes must contain information other than what heads contain). However, once we consider the impact of the Distributed Morphology framework (Halle and Marantz 1993, 1994, Marantz 1993), according to which morphological structure is composed in syntax, does *terminal* mean ‘syntactically indivisible’ (atomic) or ‘merged from the lexicon/numeration as a single feature bundle and/or a single lexical item’? While in systems where feature bundles are inserted *as is* there is no difference between the two notions, the exact definition plays a role when it comes to derived heads (i.e., heads arising as a result of head movement) in the Distributed Morphology framework.

Suppose that heads are defined as nodes merged from the numeration. Consider then an item like the French finite verb *fin-ir-a* ‘finish-FUT-3SG’. From the syntactic point of view, this item is a head (as witnessed by the fact that it moves to C<sup>0</sup> in questions). However, it clearly has an internal morphological structure (root + tense +  $\phi$ ) and thus has not been merged *as is* from the Distributed Morphology point of view. Therefore, heads must be defined as syntactically atomic, which does not prevent them from being internally complex.

Many thanks to Cedric Boeckx, Danny Fox, Morris Halle, Ken Hiraiwa, Anoop Mahajan, Alec Marantz, David Pesetsky, Jay Rifkin, Eddy Ruys, and Philippe Schlenker for many fruitful discussions and pointers in different (right!) directions; to the audiences at MIT Ling-Lunch (December 2001) and the Scandinavian Conference of Linguistics 19 (January 2002) for their questions; to the two *LI* reviewers, whose comments made the argument both tighter and better; and in particular to Noam Chomsky for arguing against my hypotheses, to their improvement.

(1) *Definition*

A head is a syntactically indivisible bundle of formal features.

How do we arrive at this indivisibility? How do we obtain the nontrivial fact that from the syntactic point of view, a derived head behaves like a nonderived one? Assuming the methodological principles proposed in the Minimalist Program (Chomsky 1995), the atomicity of derived heads should follow from independently motivated properties of the grammar. I will argue that the operation standardly referred to as *head movement* (which creates derived heads) consists in fact of two operations, one of which is able to create complex nodes that are syntactically atomic. This approach will also allow us to remove several other problematic issues associated with head movement, especially in the minimalist approach.<sup>1</sup>

The article is organized as follows. In section 2, I will first present the empirical differences between head movement and phrasal movement. Then I will argue that head movement is, like phrasal movement, a complex operation, consisting of Search-F + (Re)Merge and obeying the c-command condition on movement. In the same section, I will defend the view that the analogue of Agree (feature valuation) in head movement is c(ategorial)-selection,<sup>2</sup> a view that will derive head movement's strict locality conditions (Travis 1984). In section 3, I will propose that the landing site of head movement is the specifier of the root (where at every point of syntactic derivation the root is defined as the topmost node, which has no parent) and that it is followed by the operation of m-merger, which results in the head adjunction structure traditionally associated with head movement. In section 4, I will demonstrate that these two assumptions, in combination with those made in section 2, result in the correct ordering of head movement and phrasal movement. I will derive the remaining special properties of head movement from the proposal that m-merger is a morphological operation—an assumption that I will motivate in section 5. Section 6 is the conclusion.

In the appendix, I will discuss Chomsky's hypothesis that head movement occurs after Spell-Out, on the phonological branch of the derivation. I will show that the claim that head movement has only PF effects is irrelevant or incorrect, and that assigning head movement to the phonological branch of the computation creates more problems than it solves. In the appendix, I also address the (apparent) lack of semantic and syntactic effects of head movement.

Before we proceed, a clarification is in order. This article is not intended to be a statement about the status of borderline cases (e.g., long head movement, discussed in Lema and Rivero 1990, Embick and Izvorski 1995, Bošković 1997, Toyoshima 2001, etc.), but rather a formalized description of paradigm instances of head movement (e.g., French V<sup>0</sup>-to-T<sup>0</sup> movement). I will therefore assume the standard restrictions on head movement (the Head Movement Constraint, lack of excorporation, etc.) and leave their apparent violations as something to be explained. Affix hopping and other potential cases of lowering are also left outside the picture.

<sup>1</sup> I will follow the proposal made in Chomsky 2001b and leave incorporation out of the picture under the assumption that it is much more poorly understood. A view of incorporation as compounding (Rosen 1989) supports this distinction. The proposed division of head movement into two components might provide new insights into the nature of incorporation.

<sup>2</sup> The link between c-selection and head movement was independently developed in Julien 2000, 2002b.

## 2 A Unified Theory of Movement

In this section, I defend the claim that head movement and phrasal movement are triggered by the same factor and are in fact instances of the same phenomenon (feature valuation followed by (Re)Merge). I argue that the differences between the two are due to different features triggering the movement and should not be encoded in the theory via special notions of ‘‘head movement’’ and ‘‘phrasal movement.’’

The locality restriction on head movement (Travis’s (1984) Head Movement Constraint) is not special to head movement but also governs c-selection: a head can c-select only its complement—and the same is true for lexical selection (Pesetsky 1982, 1995). I will argue that c-selection is to head movement what Agree is to phrasal movement: the trigger. This assumption will allow us to derive the locality constraints from the way projection is implemented. A welcome consequence of the unified theory of movement defended here is that it permits us to view all movement as generalized pied-piping, which admittedly still has to be defined (though see Heck 2004).<sup>3</sup>

### 2.1 Head Movement versus Phrasal Movement

Head movement differs from phrasal movement in several important ways, detailed in (2) (of which some are due to Mahajan (2000) and Harley (2004), and others are gathered from various sources). These differences, and the clear undesirability of having two types of movement in linguistic theory, have led many syntacticians, including but not limited to Chomsky (2000), Koopman and Szabolcsi (2000), and Mahajan (2000, 2001), to exclude head movement from the narrow syntax, either by relegating it to the phonological component of the grammar (Chomsky 2000; see also Harley 2004 for a very different implementation based on Hale and Keyser 2002, which allows head movement to take place in syntax but assumes that no actual movement is involved, and Parrott 2001 for a morphology-based approach) or by reanalyzing head movement as remnant movement (e.g., Koopman and Szabolcsi 2000, Mahajan 2000, 2001, Nilsen 2003).

#### *Head movement versus phrasal movement*

- a. The probe and the target act as one constituent after head movement, but not after phrasal movement.
- b. Neither the probe nor the target can be extracted after head movement.
- c. Head movement is more local than phrasal movement.<sup>4</sup>
- d. Head movement feeds affixation; phrasal movement does not.
- e. Head movement seems to have no semantic or syntactic effects, but phrasal movement does.

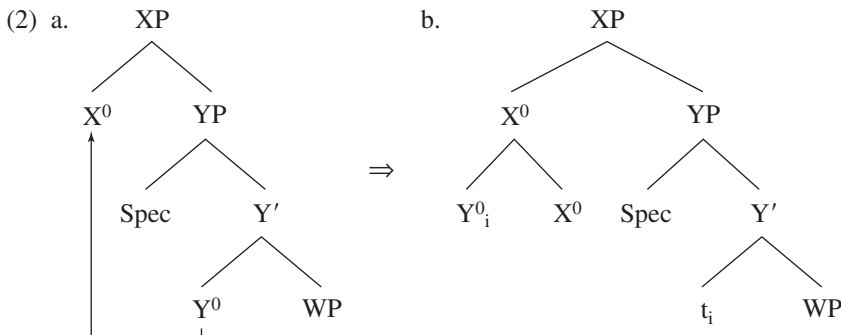
<sup>3</sup> Johnson and Lappin (1997:295) also argue that the notion of overt movement as generalized pied-piping does not follow from anything in the Minimalist Program.

<sup>4</sup> The so-called long head movement (e.g., Lema and Rivero 1990, Embick and Izvorski 1995, Bošković 1997, Toyoshima 2001) could be viewed as an argument against either the property (b) (the target and the moving head do not have to remain a constituent after movement (excorporation)) or the property (c) (the locality constraint on head movement (Travis 1984) can be violated). I will discuss long head movement in section 5.1.3.

The last two points in this list are misleading. The validity of property (d) is undermined by the existence of phrasal affixes, such as the Saxon genitive marker 's. On some accounts (Giorgi and Longobardi 1991 and references cited therein), Saxon genitives are created by movement of the possessor, agent, or theme DP from its argument position below the possessive noun to Spec,DP. If this view is correct, phrasal movement can not only feed affixation (*his*) but even trigger suppletion (*our*).<sup>5</sup>

Property (e) is also not uncontroversial. Syntactic effects of head movement have been proposed in a sizable body of work (including Larson 1988 and Koopman 1984, as well as considerable research on Scandinavian object shift), and I will explain its visible lack of semantic effects in the appendix. Be that as it may, the fact that the probe and the target act as one constituent after the movement is difficult to view as anything but a syntactic effect.

To account for the fact that items associated by head movement act as a single constituent, head movement was theorized to result obligatorily in adjunction to a head, as in (2), as opposed to the regular movement to a specifier. Both the higher head  $X^0$  and the lower one  $Y^0$  may be null.



One of the special properties of adjunction is the impossibility of extracting either the adjoining element ( $Y^0$  in (2b)) or the target element (the lower segment of  $X^0$  in (2b)) separately.<sup>6</sup> Postulating head adjunction therefore accounts for properties (a) and (b) of head movement, but says nothing about property (c). Moreover, head movement acquires two additional theory-internal properties distinguishing it from phrasal movement.

#### *Head movement versus phrasal movement*

- f. The landing site of head movement does not c-command the extraction site.
- g. (The Uniformity Condition on Chains) Only a head can be adjoined to a head; only a maximal projection can be merged as a specifier. (See Carnie 1995, Nunes 1998, and Toyoshima 2000, 2001 for arguments against this condition.)

<sup>5</sup> Strikingly, in this case affixation and the ensuing morphological processes appear to involve elements that do not form a constituent. See section 5.1.2 for discussion.

<sup>6</sup> It should be noted that whereas adjunction of maximal projections is allowed in the minimalist framework for Merge, it is not usually proposed for Move (i.e., Agree + Merge). I will not be concerned with this minor difference between head movement and phrasal movement here, but my proposal resolves it without further assumptions.

Given that a moved maximal projection always c-commands its trace, property (f) does distinguish head movement from phrasal movement, and all attempts to remove this difference have relied upon changing the definition of c-command from the simple and strictly configurational “contained in the sister of” to something less so (e.g., to “is dominated by the first maximal projection that dominates”). In the minimalist framework, the c-command condition on movement is derived from the hypothesis that Merge must extend the tree (the *Extension Condition*, due to Chomsky 1995) and so cannot be changed easily. It is possible to suggest several alternatives to the Extension Condition that would allow head movement. One such proposal is featural cyclicity (Richards 1997); another, that the constraint is on where the attracting feature can search for the attractee rather than on the target position (Ken Hiraiwa, pers. comm.).

*2.1.1 Featural Cyclicity* It is not impossible that the Extension Condition is wrong. Richards (1997) suggests that it should be replaced by the notion of cyclicity (Chomsky 1995). This version of the constraint permits head movement and also allows the merger of a second specifier below the first one (*tucking-in*).

(3) *Cyclicity*

Strong features must be checked immediately upon being introduced into the derivation.

It should be noted, nonetheless, that (3) does not completely eliminate the requirement that phrasal movement target a c-commanding position. Consider LF movement (in particular, covert *wh*-movement), triggered by weak features and therefore not subject to (3). Since covert *wh*-movement is also constrained by c-command, all covert movement would have to be regarded either as triggered by mechanisms other than feature checking or as pre-Spell-Out movement.

An additional result of this approach is that there can be only one strong feature per head, with unclear consequences (though see Nash and Rouveret 1997 for just such a proposal in a different framework).

*2.1.2 Attraction to the Root* An alternative proposal would be to constrain the relation between the attractor and the attractee instead of imposing conditions on the relation between links of a chain. One possibility is to require the target of movement to be in the domain of the probe (Ken Hiraiwa, pers. comm.); another, to stipulate that only the topmost head of the tree can attract. Under this approach, nothing is said about the relation between a moved constituent and its trace, and as a result, both head movement and tucking-in will be allowed.

The problem is that this proposal makes chains formed by movement radically different from binding chains, and of course the Uniformity Condition on Chains (property (g)) is still required. We have gained little in terms of the number of speculations necessary to account for the differences between head movement and phrasal movement.

## 2.2 Summary

Chomsky (2000) proposes that head movement occurs on the phonological branch of the derivation, after Spell-Out. As a result, head movement is not reflected at LF and thus is no longer expected to have semantic effects (property (e)). Since PF movement occurs in the tree that is

already constructed, it is not required to extend the target, either (property (f)), and so head movement may not obey the c-command condition.

Apart from the fact that this proposal was never explicitly spelled out (though see Boeckx and Stjepanović 2001, Flagg 2002, and Sauerland and Elbourne 2002 for some discussion), several theoretical and methodological issues arise. Embick and Noyer (2001) argue against PF syntax from considerations of parsimony: postulating two modularly distinct syntactic systems departs from a minimalist model of the grammar. Zwart (2001) notes that the very existence of a phonological branch of the derivation is theoretically suspicious: assuming that the PF interface is the process of converting elements of the syntactic structure into strings of phonemes, any deviation from this basic definition must be argued for and its properties must be established. Furthermore, it can be shown (see the appendix) that Chomsky's proposal does not solve the problems it sets out to solve and creates some new ones.

To summarize, the difference between head movement and phrasal movement with respect to locality and constituency cannot be explained in the standard theory of head movement. In order to claim that both are triggered by the same mechanism (feature valuation), we need to explain why they differ. Our theory must also predict which type of movement happens in which environments.

### 2.3 *Head Movement Constraint*

I will begin with the assumption that there is no theoretical difference between head movement and phrasal movement. That is, I will assume that there is only one kind of movement: feature valuation followed by (Re)Merge.<sup>7</sup> Strong support for this claim comes from the fact that head movement and phrasal movement seem to be in complementary distribution (Pesetsky and Torrego 2001).

Travis (1984) shows that only the head of the sister of a head  $H^0$  can move to  $H^0$ .

#### (4) *Head Movement Constraint* (Travis 1984)

Head movement may not skip intermediate heads.

To the extent that such a claim can be theory-independent, there are no clear cases of moving a complement of  $X^0$  to a specifier of  $X^0$  (Comp-to-Spec movement).<sup>8</sup> As argued by Pesetsky and Torrego (2001), head movement is possible where phrasal movement is not, and vice versa.

#### (5) *Head Movement Generalization* (Pesetsky and Torrego 2001)

Suppose a head H attracts a feature of XP as part of a movement operation.

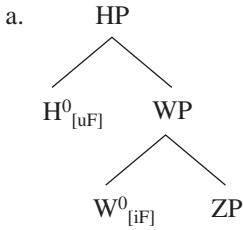
- a. If XP is the complement of H, copy the head of XP into the local domain of H.
- b. Otherwise, copy XP into the local domain of H.

<sup>7</sup> I have nothing to say about quantifier raising (QR)/wh-movement, but it is irrelevant here.

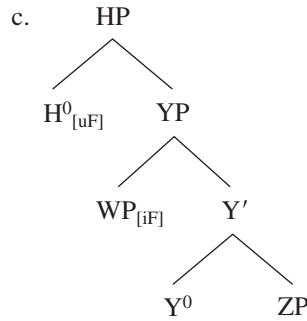
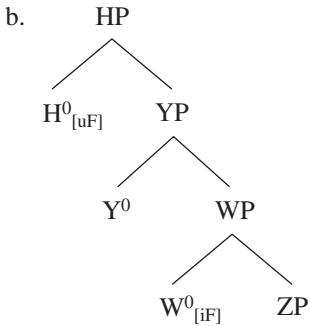
<sup>8</sup> See Cinque 2000, Julien 2000, Shlonsky 2002, and Sichel 2002, among others, for proposals using the cumulative ("snowballing") remnant and Comp-to-Spec movement to capture the head parameter.

In other words, in the configuration (6a), corresponding to the condition (5a), only  $W^0$ -to- $H^0$  head movement is possible; WP may not move. In the configurations (6b) and (6c), exemplifying (5b), only phrasal movement of WP is permitted.

(6) *Head movement only*



*Phrasal movement only*



If head movement and phrasal movement are two sides of one phenomenon (feature valuation followed by (Re)Merge), their complementary distribution, though expected, must still be explained.

2.4 *Locality*

Why does the Head Movement Constraint (property (c)) hold? To answer this question, we must address three others, in the following order: First, what kind of locality is involved? Second, how do we implement this? Third, why are there two types of movement?

2.4.1 *The Head of the Complement* C(ategorial)-selection shows the same kind of locality as head movement: a head can only c-select the head of its complement. To show that c-selection is required independently of the theory, consider the distribution of nouns versus adjectives and intransitive verbs. All three have the semantic type of a one-place predicate  $\langle\langle e, t \rangle\rangle$ ,<sup>9</sup> but while the first can appear with determiners (*the rain*), the latter two cannot (*\*the arrive, \*the cloudy*).

<sup>9</sup> I simplify a little here—though most nouns belong to this type, only *some* adjectives (strictly intersective, such as *French*) do. This fact does not change the argument, though.

Likewise, there is no semantic reason for lexical selection (on which see Pesetsky 1982, 1995)—that is, for why the English verb *depend* should appear with *on*, and its Hebrew equivalent, *t-l-h*, with *b-* ‘in’. Since we cannot relegate the matter to semantics (unless we assume that verb meanings are radically different in different languages), and since the motivation for c-selection comes from syntax, it has to be a syntactic phenomenon. While we still have no idea why c-selection should exist, it seems to be empirically necessary.

(7) *C-selection*

A head may select the syntactic category (and the lexical content) of the head of its complement.

The fact that the locality constraint on head movement restricts another syntactic process suggests that the head-of-the-complement locality is as basic as Attract Closest. It also offers support for the impression that head movement is not fully ad hoc, and its exclusion from the theory, suggested by Koopman and Szabolcsi (2000), Mahajan (2000, 2001), and others, should be motivated.

The question that concerns us now is *why* such locality conditions? In other words, what do c-selection and head movement have in common that they should be local in the same way? The simplest possible answer is that head movement is based on c-selection, just as phrasal movement is based on Agree.

The idea that c-selection is done by a feature, possibly via head movement, is not new. Svenonius (1994), Holmberg (2000a), and Julien (2000) use a special c-feature for the job. Following Julien (2000), I assume that c-features are (uninterpretable) counterparts of categorial features: [uN] for  $D^0$ , [uV] for  $T^0$ , and so on.<sup>10</sup>

*2.4.2 More on C-Features* Here, I will show that c-features behave like other features in that they can be interpretable, trigger phrasal movement, and are not limited to lexical category features.

If the c-feature of  $T^0$  is the uninterpretable counterpart of the categorial feature of the verb, is the *interpretable* V-feature ever overtly manifested on the verb (Noam Chomsky, pers. comm.)? Likewise, do nouns bear interpretable N-features? I believe that the answer is yes and that [iV], [iN], and so on, correspond to the ‘little x’ functional heads in the Distributed Morphology framework (Halle and Marantz 1993, 1994), that is, to verbalizing, nominalizing, and adjectivizing affixes. Combining the root with a category affix  $f^0$  ([iF]) results in a head of the appropriate lexical type ( $F^0$ ).

The next question is whether categorial features can trigger phrasal movement. Since there is, I believe, nothing exceptional about c-features, the answer should again be yes. Indeed, I will

<sup>10</sup> If there is more than one functional head in the extended projection, it may also be the lexical head of this projection that determines the c-feature of these functional heads. This means that  $D^0$ ,  $\text{Num}^0$ , etc., all share the [uN] feature, while  $T^0$ ,  $\text{Asp}^0$ , etc., are endowed with [uV]. Presumably,  $D^0$  selects a NumP because  $\text{Num}^0$  bears [uN] (cf. Julien 2000) or because  $N^0$  raises to  $\text{Num}^0$  covertly. The latter proposal would provide a means of formally expressing the notion of extended projection (Grimshaw 1991).



argue that A-movement of NPs is triggered by N-features (Chomsky 2001b) rather than by  $\phi$ -features (Chomsky 2001a).

That  $\phi$ -features cannot be involved is shown by the fact that it is impossible to find a  $\phi$ -feature that would be common to all kinds of DPs that can appear in subject position. Movement to Spec,IP cannot be triggered by [gender] (noun class, etc.) since there are languages that have no gender distinctions (e.g., English). [person] is excluded because nonpronominal xNPs would not bear this feature (on the standard assumption, due to Benveniste (1966), that 3rd person is lack of person specification and thus unmarked). [number] is equally impossible: (a) singular is almost never morphologically marked<sup>11</sup> and is therefore unlikely to correspond to a formal feature on a singular noun, (b) even if there were a [singular] feature for singular nouns, mass nouns clearly bear no interpretable number specification, and (c) there are languages without any number marking at all (e.g., Chinese). We conclude that  $\phi$ -features cannot be triggers for NP-movement. However, the categorial feature [N] may be such a trigger, since it is shared by all the functional heads in the noun phrase (see also Julien 2000).

Can head movement be triggered by a nonlexical (or a noncategorial) feature? There is no reason to reject this possibility, and it is definitely called for by such analyses as Pesetsky and Torrego's (2001) account of T<sup>0</sup>-to-C<sup>0</sup> movement in Germanic, which suggests [uT] as the trigger for head movement. Given that C<sup>0</sup> c-selects for T<sup>0</sup> (e.g., *for* requires a nonfinite T), there is no reason why [uT] should not be a trigger for head movement as well.

We conclude that head movement is similar to phrasal movement with respect to its trigger (c-selection). The issue now is whether we can explain why c-selection has such strict locality conditions.

*2.4.3 Why the Head of the Complement?* The locality condition on what a head can c-select for and what can, therefore, be targeted by head movement consists of two parts: only the complement (i.e., the sister of the head) is considered as a possible search domain, and then only its head (and not, e.g., the head of its specifier) can be categorially or lexically constrained.

The first part can be derived from the Extension Condition (Chomsky 1995).

(8) *Extension Condition*

Merge should be effected at the root.

The Extension Condition derives the c-command condition on movement: if at every stage of the derivation, an item can only (re)merge at the root (which is defined, at this stage, as the node that dominates all other nodes), then a remerged item will necessarily c-command its trace.<sup>12</sup> This means that the complement is privileged not only for head movement, but also for phrasal movement, and this constraint is due to the fact that nothing higher is present at this point in the derivation. But why is the *head* of the complement special for head movement and c-selection?

<sup>11</sup> See Corbett 2000 and Harbour 2003 for discussion of Kiowa, where plural appears to be the unmarked option.

<sup>12</sup> Hornstein (1999) and Nunes (1995) suggest that the Extension Condition could be maintained without necessarily yielding c-command if movement can occur between independently constructed subtrees (sideward movement). This does not help us with head movement, unless it is also treated as occurring between two trees (Bobaljik and Brown 1997).

To answer this question, we need to turn to the notion of projection. It is a standard assumption that when two trees are merged, it is necessary to determine which of them projects, that is, which one determines the syntactic behavior of the resulting tree (Chomsky 1994; but see also Collins 2001 for a proposal eliminating labels from narrow syntax). Therefore, whenever two trees are merged, syntax must be able to assess the featural composition of the topmost head of each and determine which one projects. Consequently, when a tree XP is merged with another tree YP, both  $X^0$  and  $Y^0$  can be accessed by the computational mechanism.<sup>13</sup>

Suppose we are merging a head  $X^0$  and a nontrivial tree YP, and  $X^0$  bears the uninterpretable categorial feature [uY]. C-Select establishes an asymmetric relation between the two heads, which may or may not result in movement.<sup>14</sup> Let us assume that Agree and C-Select (as well as other search procedures, if they exist) are constrained by the following condition, resembling the bracket erasure convention in *SPE* (Chomsky and Halle 1968) and the Phase Impenetrability Condition (Chomsky 2001b):

(9) *Transparence Condition*

A head ceases to be accessible once another head starts to project.

The role of the Transparence Condition is to ensure that the only head that is syntactically accessible is the projecting head. This means that a syntactic relation can be established between heads of two independent trees at Merge (when both are still projecting), but no later.

Further support for the existence of relations between heads comes from lexical selection (l-selection): a head can lexically select the head of its complement (Pesetsky 1982, 1995). While c-selection can be reformulated in terms of labels (maximal projections) rather than heads, l-selection cannot. Suppose now that l-selection is c-selection for several features, which are in fact bundled to compose a particular head—say, a preposition.<sup>15</sup> This means that l-selection and c-selection constitute the same phenomenon, which gives us an additional reason to believe that syntax can access and relate heads.

Speculative as the Transparence Condition is, its other consequence is to reduce the number of items in active memory. Indeed, at the point when two trees are merged, the only information required is the featural composition of the topmost heads, which would allow both c-selection and l-selection (on the assumption that they are the same thing). Once it has been established which head projects, there is no longer any need to keep track of the nonprojecting head separately from its projection. If C-Select is a kind of a search procedure in the constructed tree (like Agree, and unlike Select, which happens in the numeration), then once C-Select has taken place (and

<sup>13</sup> Note that the merger of a head  $H^0$  with a nonhead XP cannot be assumed to always result in the projection of an HP. One simple counterexample is merger of a subject pronoun (which at least in some cases is simultaneously a  $D^0$  and a DP) with a VP. The result is clearly a VP and not a DP.

<sup>14</sup> It could be imagined that when C-Select fails (as it does when a specifier is merged), another operation could take place, which would correspond to specifier-head agreement. A categorial feature can also be involved in this and other instances of Agree, but they would never involve the first Merge operation and thus will be distinct from C-Select.

<sup>15</sup> If this view is correct, we predict that only heads from closed lexical classes and possibly grammatically specified subclasses of open classes (e.g., animate vs. inanimate nouns) can be lexically selected for.

until a new search is initiated), the syntactic operation (Re)Merge can access the c-selected head (and movement of this head is possible).

The Transparency Condition has an additional interesting consequence. The ordering of various types of movement becomes nonrandom. Head movement must happen immediately after merger of a new projecting head, and therefore before any kind of phrasal movement. We will see how welcome this result is in section 4.

*2.4.4 Why Is There Phrasal Movement?* The Transparency Condition, in combination with another empirically required mechanism, pied-piping, is fully compatible with the necessity of phrasal movement. Under the assumptions that (a) the minimal syntactically visible element is a feature (Chomsky 1995) and (b) the minimal element that can be merged in syntax is a head (see section 5 for a discussion and possible implementation of this assumption), overt movement (which is what we are concerned with here) necessarily includes some degree of pied-piping.

Pied-piping is informally understood as the operation that moves the constituent containing the attracted element, if the attracted element cannot itself be moved. To restate this in slightly more precise terms, pied-piping is a recursive procedure called upon whenever (Re)Merge is inapplicable to the syntactic unit *X* it is being applied to (as would be the case if (Re)Merge were being applied to a feature), and returning the next-smallest movable element dominating *X* as the next candidate.<sup>16</sup> The extent of pied-piping is determined by what the minimal *movable* element containing *X* is (e.g., nonmaximal projections that are not heads cannot ever be moved), and it interacts with the reason why a particular constituent may be unable to move.

I therefore claim that just as head movement is pied-piping when applied to a feature, so phrasal movement is the next step: pied-piping when applied to a head. Because of the Transparency Condition, a head ceases to be movable once the next head up starts to project, or, for all practical purposes, in configurations where the Head Movement Generalization would have predicted phrasal movement.

To summarize, I have argued that the difference in locality between head movement and phrasal movement results from the different triggering mechanisms: whereas phrasal movement is based on Agree, head movement is triggered by the independently motivated mechanism of C-Select. C-selection determines which of the two merging constituents projects, and therefore is extremely local. I have argued that the head of a maximal projection is accessible as long as it is not merged with its own selector (the Transparency Condition). Once a head ceases to be accessible, the next-smallest constituent, the maximal projection containing the attracted head, is pied-piped. The Transparency Condition thus derives the complementary locality conditions on head movement and phrasal movement.

<sup>16</sup> The exact definition of pied-piping is not clear at this point. The scale of elements I have defined it for is pied-pipe (feature) = head (feature), and pied-pipe (head) = phrase (head). Empirically, pied-pipe (phrase) returns a larger phrase (e.g., the PP containing the attracted DP or the DP containing the attracted *wh*-word), but it is unclear how this result is obtained. Likewise, optional pied-piping of prepositions versus preposition stranding is not expected if the pied-piping results from inaccessibility of the attracted element.

I must leave both questions open here; but see Heck 2004.

## 2.5 Summary

I have proposed the beginnings of a unified theory of movement, whose goal is to account for head movement and phrasal movement in the same vein. I have suggested that both are triggered by the need to value an uninterpretable feature. As the minimal syntactic item is the head (no more than a stipulation at this point), (Re)Merge minimally targets the head containing the relevant active feature; all phrasal movement is due to generalized pied-piping applying when the head is inaccessible, for whatever reason.

The locality conditions on head movement and c-selection are the same because C-Select is the trigger for head movement. The existence of such locality constraints is due to the Transparency Condition, linking the syntactic accessibility of a head to its ability to project.

A head c-selecting the head of its sister projects. By the Transparency Condition, once the new c-selecting head is determined, the nonprojecting head ceases to be accessible (cf. “bracket erasure”). However, C-Select, being a search procedure, has by then found this nonprojecting head, which can then be accessed (remerged). In other words, though the Transparency Condition has rendered the internal structure of a phrase opaque, we have retained access to the head we want to remerge. The result is that head movement can now be viewed as feature-triggered.

The next step is to consider the peculiarities of head movement. Of these, lack of c-command, which is an otherwise fundamental constraint on syntactic and semantic operations, is probably the most important. It strongly suggests that head movement has been misanalyzed. Why should it be the case that phrasal movement targets a specifier, while head movement adjoins to the head?

I will now propose a novel analysis, which eliminates the usual problems associated with head movement. I will do so by contending that head movement does obey c-command, just like phrasal movement, and therefore targets the specifier of the attracting head (which is the specifier of the root at this point in the derivation, by the Extension Condition). Then I will argue for an additional morphological operation intervening after the movement of the head and converting the structure into the familiar pattern of two adjoined heads.

## 3 The Landing Site

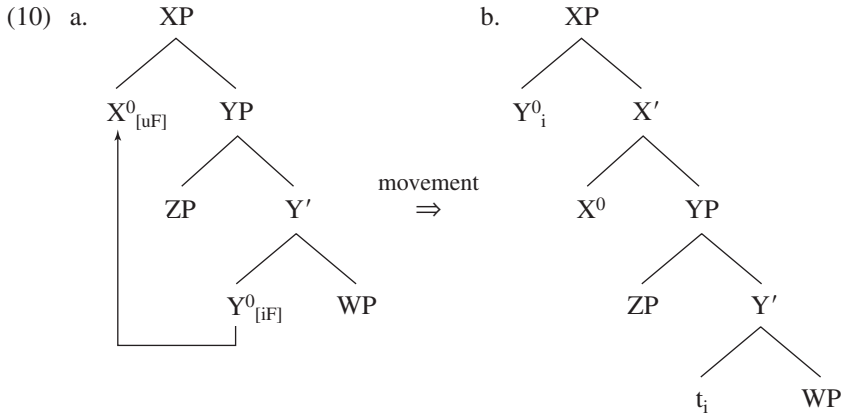
If head movement and phrasal movement are the same thing, why does phrasal movement target the specifier (i.e., the moved constituent is merged at the root), while head movement does not?

As discussed in section 2.1, one clear fact about head movement is that the probe and the target act as a single constituent after movement. The standard approaches to head movement have formalized this by assuming that the landing site of head movement is not the root, but the selecting head. Regrettably, this approach further multiplies the differences between the two types of movement.

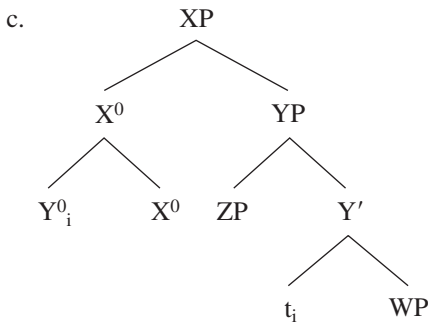
Instead of changing the definitions of c-command, the Extension Condition, and so on, I propose a different interpretation: while the movement operation itself is the same for head movement and phrasal movement (adjunction to the root), what happens *after* the movement is different. I suggest that the probe and the target of head movement are subject to a special head-

merging operation, which I will call *m-merger*. I propose that m-merger is an operation of the morphological component and is therefore separate from the movement operation itself.

To see this more concretely, consider the structures in (10). The first operation that takes place in (10a) is movement, and like all movement, it targets the root of the tree, resulting in (10b). In other words, I assume that head movement, just like phrasal movement, targets a specifier of the attracting head. This assumption in itself is not new (e.g., Toyoshima 2000, 2001).



The m-merger operation applies to (10b) to yield (10c).



Though the head adjunction structure in (10c) is a result of something other than movement, at first glance, the end result is the same as in the standardly assumed head movement theory. However, now all movement proper extends the tree and targets a c-commanding position.<sup>17</sup>

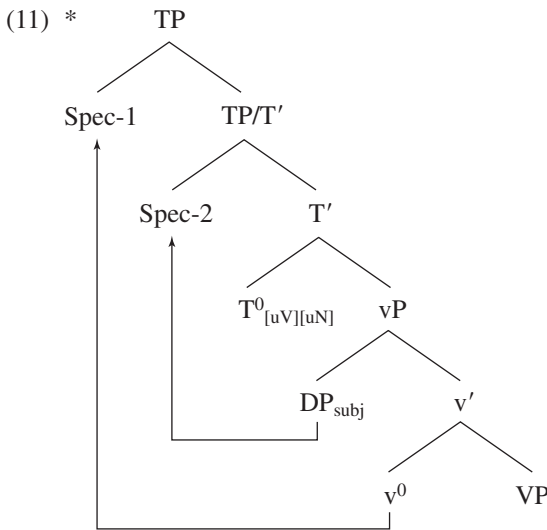
Separation between the movement operation and the head-merging operation leads to the expectation that we should find one without the other—head movement without m-merger and

<sup>17</sup> It should be noted that representationally, (10c) appears to suffer from the usual head movement problem: the moved head does not c-command its trace. I will show in section 5 how the apparent problem is resolved by certain plausible assumptions about m-merger.

m-merger without head movement. I will deal with these predictions in section 5, where I will also discuss reasons for assigning m-merger to morphology, rather than syntax.

#### 4 Ordering

Once we have unified head movement and phrasal movement, the question of their ordering becomes paramount. If all movement is feature-triggered, how do we ensure that head movement occurs before the first specifier is merged? In other words, if a head has two uninterpretable features, say,  $[uN]$  and  $[uV]$ , and a  $vP$  complement with a DP subject in the specifier, as in (11), what moves first—DP or  $v^0$ ? This question becomes particularly important in light of the proposal by Pesetsky and Torrego (2001) that a head (e.g.,  $T^0$  in (11)) is equidistant from the head and the specifier of its complement ( $v^0$  and  $DP_{subj}$ , respectively). As a result, neither  $v^0$  nor  $DP_{subj}$  has priority over the other with respect to which is attracted first, yielding the unattested optionality:  $DP_{subj}$  could theoretically move first, followed by the (re)merger of  $v^0$  as Spec,TP, without subsequent m-merger.<sup>18</sup>



Fortunately, the structure (11) is ruled out by independent considerations. When Merge combines  $T^0$  and  $vP$ , the necessity to determine which head projects triggers C-Select. This means that c-selection is obligatorily the first operation to take place in the newly created tree. Since C-Select is the same thing as Agree, once a C-Select relation is established between two heads,

<sup>18</sup> At this point, it is still an unresolved question whether head movement without subsequent m-merger is possible (see section 5.1.3). If the answer is no, then any attempt to subvert the “natural” ordering of movement operations will fail.

the possibility of (overt) head movement ensues until a new search procedure is initiated.<sup>19</sup> Head movement therefore must precede phrasal movement.<sup>20</sup>

In the next section, I will examine the operation of m-merger in detail and determine what its properties are.

## 5 The Nature of M-Merger

I have redefined head movement as a combination of three consecutive operations: C-Select, Merge, and m-merger.

Multiple questions arise. What does m-merger do? Why does it apply only to heads? How does it know that derived heads are also heads? How does syntax in general, for that matter? On the empirical side, does head movement occur without m-merger, and m-merger without head movement?

I will first show that m-merger is required for reasons independent of the head movement issue. I will use crosslinguistic data to provide evidence for the existence of m-merger as an operation separate from movement. Then I will address the issue of what it is exactly that m-merger does and why it is an operation best viewed as part of the morphological component.

### 5.1 *M-Merger as an Independent Operation*

In this section, I will empirically motivate m-merger by demonstrating that it can occur independently of head movement. This will achieve the major goal of the project: to show that the problematic case of movement can be accounted for in the standard theory of movement as merger at the root (and therefore obeying c-command) combined with an independently required operation, which can take place either with phrasal movement or without any movement at all. The separation of head movement into two components will also leave the theory enough flexibility that we can adjust it in accordance with the data, since the jury is still out on the question of the Head Movement Constraint (section 5.1.3).

First, I will provide an example of phrasal movement feeding m-merger (when the moved element is simultaneously a head and a maximal projection). Then I will examine cases of m-

<sup>19</sup> I do not discuss covert head movement here. Julien (2000) suggests that head movement is overt if the higher head bears, in addition to the c-feature, a strong categorial feature that attracts the head of its complement. Covert head movement is then equivalent to agreement in the shared feature. Such agreement could be used to ensure that a higher head, searching for the same categorial feature (e.g.,  $D^0$  requiring an  $N^0$ ), can take as a complement a functional projection (e.g., NumP) taking that categorial head as a complement. On my proposal, it can be argued that head movement is overt only if followed by obligatory m-merger.

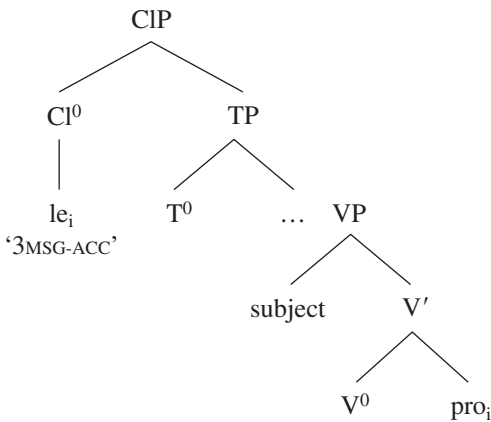
<sup>20</sup> As a cursory examination of tree (11) shows,  $T^0$  bears two uninterpretable features: [uN] and [uV]. From the syntactic standpoint alone, it can be imagined that an xNP rather than a VP is merged as the complement to  $T^0$ , with its nominal head triggering C-Select and eventually head movement. The vP can then be merged as the specifier of TP. Fortunately, this particular derivation can be ruled out by semantics, and hopefully all other derivations of this kind can be too.

merger without movement (i.e., affixation in situ). Finally, I will briefly investigate possible implementations of head movement without m-merger (also known as excorporation).

*5.1.1 M-Merger with Phrasal Movement: Clitics* As is well known, unstressed Romance pronominal direct and indirect objects (as well as some other pronouns) usually cliticize to the tensed verb, and the resultant item may move to  $C^0$  as a unit. While it is not altogether clear at what stage this happens (i.e., whether proclitics adjoin to  $T^0$  or whether they are adjoined to the main verb while it is still in some lower position), here and below I will assume that proclitics are adjoined to the main verb only after it has moved to Spec,TP and m-merged with  $T^0$ . The argument would not change if some lower position is involved.<sup>21</sup>

Two kinds of analyses of clitics have been proposed. According to one analysis (e.g., Jaeggli 1982, Borer 1983, Sportiche 1983), clitics are assumed to be functional heads ( $Cl^0$ ) in the extended VP, project their own independent projections (CIP), and bind the pro in the argument position.

(12) *French clitics (simplified)*



The major problem with this analysis is that it postulates heads in the extended VP projection that have nonverbal semantics and are suspiciously similar to DPs (as witnessed for example by the homophony of 3rd person accusative clitics with articles in French mentioned by Zribi-Hertz (to appear), among others). The existence of head-XP chains, required for the binding relation to hold between the clitic and the pro, is also theoretically questionable.

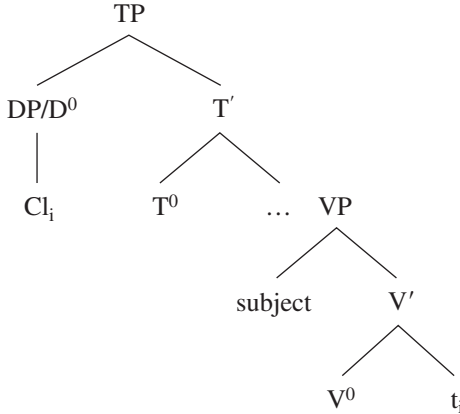
The alternative, proposed by Kayne (1975), Sportiche (1989), and others, which I will adopt here, is that clitics are pronominal DPs moved from argument positions. Since they are simultaneously heads and maximal projections, they can move as maximal projections with respect to locality, but adjoin as heads.

<sup>21</sup> Given the sketchy nature of the description, I will not address the issue of enclitics here.

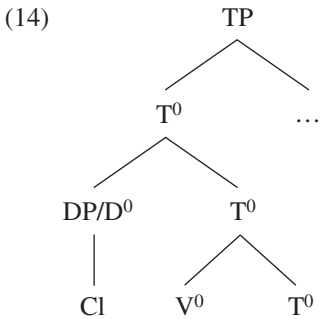


The latter analysis fits naturally into the approach I am proposing. Indeed, suppose clitics move and adjoin to the root of the tree as specifiers. This kind of movement is allowed for maximal projections under any assumptions, and we now have the configuration in (13).

(13) *Romance clitics (simplified)*



I have not yet defined m-merger properly, but what matters is that a clitic is syntactically a head as well as a maximal projection, and therefore m-merger can apply. Since nothing intervenes between the two heads, the clitic in Spec,TP m-merges with the derived V<sup>0</sup> + T<sup>0</sup> head, resulting in a standard cliticization structure as shown in (14).



Standard analyses of head movement as head adjunction, especially if burdened with the Uniformity Condition on Chains, have nothing to say about Romance clitics. What is needed to derive these structures is some operation that adjoins a maximal projection to a head (under the analyses due to Kayne (1975) and Sportiche (1989)). This happens only if this maximal projection is also a head and if the two heads are in a *head adjacency configuration*, that is, a configuration where nothing structurally intervenes between them. In other words, we need m-merger.

On the basis of classical French, Kayne (1991) and Poletto and Pollock (2004) argue that movement of clitics to the left of the verb need not be accompanied by the formation of a complex

head (in present terms, m-merger), as shown by the possibility of separating the clitics from the verb by certain adverbial phrases such as *mal* ‘badly’.

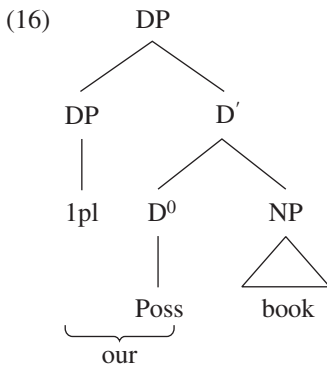
- (15) Nous faisons une vie si réglée qu’il n’est quasi pas possible de se  
 we do a life so orderly that + it NEG + is almost NEG possible of REFL  
*mal* porter.  
 badly bear  
 ‘We lead such an orderly life that it is almost impossible to be in poor health.’  
 (from Mme de Sévigné’s *Lettres*)

Thus, a further advantage of viewing m-merger as a phenomenon separate from movement is that it also allows us to capture the diachronic transition from what looks like scrambling to cliticization.

We conclude that syntactic cliticization provides independent evidence for the existence of m-merger, evidence that is not theoretically dependent on the view of head movement as a combination of movement and m-merger advocated here.<sup>22</sup>

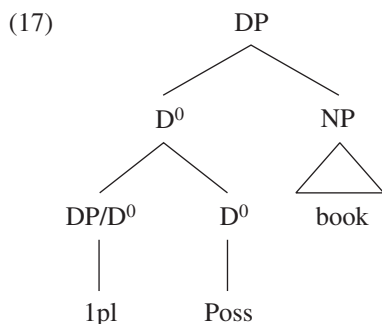
*5.1.2 M-Merger without Movement* If m-merger applies in the head adjacency configuration, we must ask whether it can apply when this configuration arises as a result of Merge from the numeration (External Merge, or just Merge) rather than as a result of movement (Internal Merge, or (Re)Merge).

One possible candidate for m-merger without prior movement is provided by Saxon genitives, where it can solve the apparent problem noted in section 2.1. Under the assumption that the possessor DP is located in Spec,DP and the genitive ‘s is a phrasal affix occupying D<sup>0</sup> (Abney 1987), suppletive lexical insertion can apparently target two morphemes that do not form a constituent, as illustrated in (16).



<sup>22</sup> Perhaps the same analysis of cliticization as phrasal movement followed by m-merger can be used for some of the second-position clitics discussed by Wackernagel (1892), Anderson (1993, to appear), Legendre (1998), and Pancheva (2002), among many others (see also Beukema and Den Dikken 2000).

In the approach proposed here, however, it is sufficient to observe that a pronominal possessor DP is simultaneously a maximal projection (DP) and a head ( $D^0$ ), and that the Poss morpheme in  $D^0$  and the 1pl possessor are two heads in the very local head adjacency configuration permitting m-merger. The structure in (16) can therefore be transformed into (17).



The m-merged  $D^0$ - $D^0$  head can serve as a target for the Distributed Morphology operation of Fusion (Halle and Marantz 1993, 1994), which creates a single syntactic node for lexical insertion, and therefore for suppletion.<sup>23</sup>

<sup>23</sup> As far as I can see, the term *fusion* is used in Distributed Morphology to describe two phenomena: (a) presyntactic bundling of features into one syntactic head (henceforth, *presyntactic bundling*), and (2) the effect of two features inside a single syntactic head being expressed by a single exponent (henceforth, *Fusion*). Although presyntactic bundling is generally accompanied by Fusion, such is not necessarily the case. Consider definite articles in a language like French.

- (i) *la*  
the-F.SG  
*le*  
the-M.SG  
*les*  
the-PL

On the one hand, the interpretable feature [definite] has to have merged with the uninterpretable gender and number features presyntactically, resulting in a single  $D^0$  head. On the other hand, the resulting  $D^0$  is clearly realized by two morphemes: the [definite] exponent *l-* and the gender/number exponents *-a*, *-e*, and *-es*, which are independently attested in possessives and demonstratives.

- |                |                    |
|----------------|--------------------|
| (ii) <i>ma</i> | (iii) <i>cette</i> |
| my-F.SG        | this-F.SG          |
| <i>mon</i>     | <i>ce/cet</i>      |
| my-M.SG        | this-M.SG          |
| <i>mes</i>     | <i>ces</i>         |
| my-PL          | this-PL            |

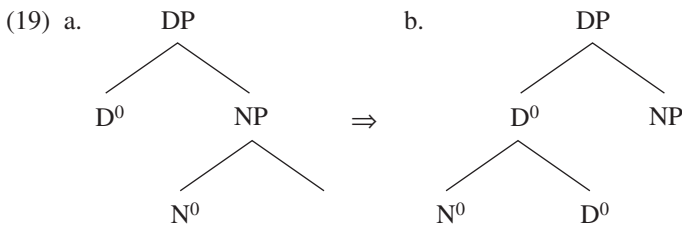
We conclude that presyntactic bundling and Fusion should be separated. While m-merger can feed Fusion (as in the suppletion case discussed here), there is much to be said in favor of viewing presyntactic bundling and m-merger as one and the same operation occurring in different environments (see below). Such an assumption can result in a nontrivial and highly useful notion of head-internal feature hierarchy, which crucially relies on Merge applying at all levels of linguistic derivation (cf. feature hierarchy in Head-Driven Phrase Structure Grammar and the discussion in Harley and Ritter 2002).

Even if the previous case could be argued to depend on prior phrasal movement of the possessor to Spec,DP (Giorgi and Longobardi 1991), such is not the case for definiteness marking in Danish, which Hankamer and Mikkelsen (2002) show to not be derivable by movement. We will now see that Danish definiteness marking provides an environment where m-merger seems to fill a clear need.

As (18a–c) show, in Danish a definite suffix is used with a bare noun; when the noun is modified, a prenominal definite article must be used.

- (18) a. hest-en  
           horse-DEF  
           ‘the horse’  
       b. \*den hest  
           DEF horse  
       c. den \*(røde) hest  
           DEF red horse  
           ‘the red horse’

Hankamer and Mikkelsen (2002) argue that  $N^0$ -to- $D^0$  raising approaches (Delsing 1993, Embick and Noyer 2001) cannot explain these data, because (a) modification blocks the use of the definite suffix, and (b) the choice of a particular head noun, such as *studerende* ‘student’, may rule out the definite suffix. Neither of these phenomena is expected on a head movement theory, because modifiers, such as APs and relative clauses, should not block head movement, and the selection of a particular word should not influence the syntax of the NP whose head it is. To account for these facts, Hankamer and Mikkelsen propose that suffixed definite forms like *hesten* ‘the horse’ are derived by a lexical rule that combines a noun with the definite suffix to yield a determiner ( $D^0$ ). I adapt this analysis to the present framework by observing that if this rule were to apply in syntax, it would be nothing other than m-merger, as shown in (19).



The only difference between Hankamer and Mikkelsen’s D-rule of definite noun formation and m-merger is that the D-rule derives definite nouns in the lexicon, while m-merger merely collapses two syntactic heads into one in the morphological component. (See section 5.2.2 for the reasoning behind relegating m-merger to morphology rather than syntax.) Since modification disrupts the very local head adjacency relation between  $D^0$  and  $N^0$  required for m-merger to take place, modified nouns cannot m-merge with the definite suffix, and the prenominal definite article

must appear. As Hankamer and Mikkelsen rule out movement in the derivation of the structure in (19b), Danish definiteness marking provides an example of m-merger applying in the absence of prior movement and feeding affixation.

M-merger is also preferable to Hankamer and Mikkelsen's D-rule because it provides a natural way of ruling out unattested variants like (18b). While Hankamer and Mikkelsen rely on some sort of syntactic blocking of the analytic form by the synthetic one, the analysis proposed here excludes examples like (18b) if m-merger is obligatory in this environment—a totally standard assumption about applicability of a morphosyntactic rule. On the other hand, the fact that the definite suffix is impossible with some nouns can be captured by assuming that they are simply not subject to m-merger, exceptions being not at all unusual in the morphological component. Since suffixation cannot apply without prior m-merger (the two relevant heads do not form a syntactic constituent), the prenominal definite article is used with these nouns.<sup>24</sup>

A question to ask at this point is whether m-merger without prior movement can be reliably distinguished from phonological cliticization. Several diagnostics can be offered now, but the list is far from complete. (a) If the complex element under consideration subsequently moves to a higher head, then it has to be a head itself. (b) Likewise, suppletion only targets heads (see the discussion of Saxon possessives above), and so complex sequences that can be suppleted cannot have been derived by phonological cliticization. (c) Finally, m-merger but not phonological cliticization can be disrupted by phonologically nonintervening modification (as is the case with definite article suffixation in Danish, which is blocked by relative clauses).

I also hypothesize that m-merger should not be conditioned by phonological deaccenting (e.g., English auxiliaries do not m-merge with their subject) or by broader context (e.g., m-merger does not occur when the French masculine article *le* contracts with the preposition *à* (to form *au*), because if the next word begins with a vowel, the article cliticizes onto this word). I leave other possible diagnostics for future research.

**5.1.3 Head Movement without M-Merger** Can head movement occur without m-merger? Evidence is shaky. If head movement without m-merger were possible, we would not have had any evidence for the Head Movement Constraint. In the approach proposed here, a head raised to a specifier and not m-merged could theoretically continue to move to the specifier of the next head (and maybe m-merge there), which would yield a result surface-equivalent to excorporation. Therefore, evidence for the Head Movement Constraint can be treated as an argument that all movement resulting from c-selection should be followed by m-merger. In this section, I briefly address purported evidence against the Head Movement Constraint.

<sup>24</sup> This approach cannot straightforwardly account for other Scandinavian languages, where the prenominal definite article also appears with modified nouns only, but the suffixal one is preserved in all cases (see, e.g., Julien 2002a). While the behavior of the prenominal definite article can be explained by the proposed theory, something else needs to be said about the suffixal definite article.

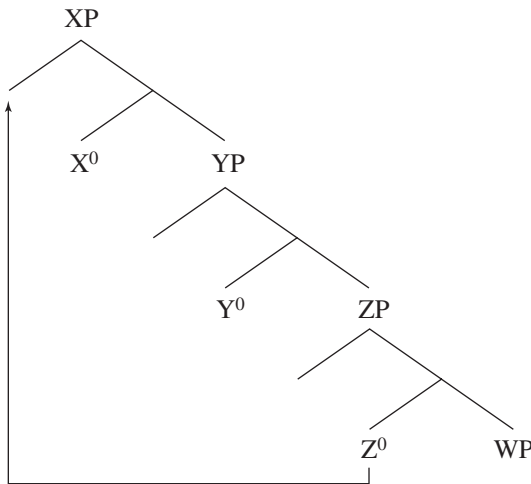
Lema and Rivero (1990) propose that the Head Movement Constraint can be violated. In such cases, a head appears to adjoin to another head, skipping over one or more intervening heads. Two major varieties of such movement have been examined: (a) movement of a verb or an auxiliary over intervening clausal negation or verbal particles, and (b) movement of the lexical verb to the left edge of the clause over the tensed auxiliary (so-called *long head movement*).

If head movement can skip intervening heads, the apparent locality of head movement observed in Travis 1984 must be due to some sort of coincidence. Two ways of handling this can be established: (a) to give up the Transparency Condition (9), or (b) to allow head movement without m-merger.

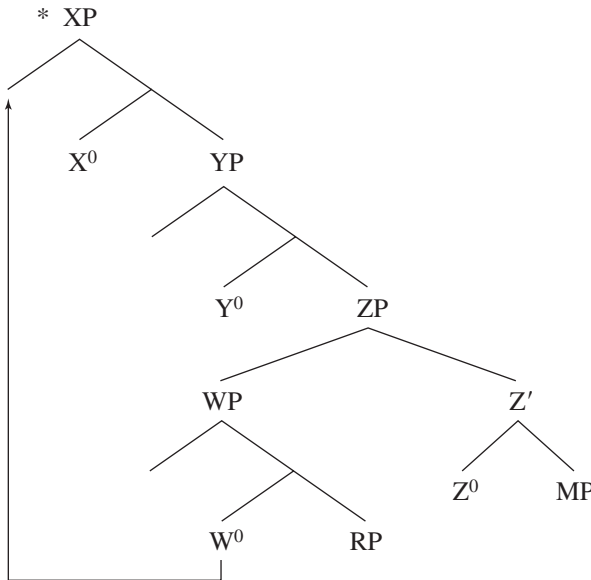
The first thing that should be noted is that the Transparency Condition distinguishes the environments for phrasal movement and head movement. If it were proved to be wrong, the idea that head movement is based on c-selection would be seriously undermined and a new algorithm for determining when head movement is chosen over phrasal movement would have to be found.

Moreover, simply giving up the Transparency Condition would mean that head movement is not constrained. This is an undesirable consequence, because even if violating Relativized Minimality (Rizzi 1990) by skipping over an intervening head as in (20a) is allowed (the two purported exceptions to the Head Movement Constraint), extracting a head out of a specifier as in (20b) is not (unlike extracting a maximal projection out of a specifier), so head movement must be constrained.

(20) a. *Long head movement, violation of Relativized Minimality*



b. *Head movement out of a specifier, not attested*



This means that we would need to find another way of reintroducing the Head Movement Constraint into the theory and define the heads that can be moved in a recursive way (the head of the complement, the head of the complement of the complement, etc.). Although keeping or abandoning the Transparency Condition will not affect the separation of head movement into movement and m-merger, the number of assumptions necessary to constrain head movement will not be diminished.

The second way of dealing with the two purported exceptions to the Head Movement Constraint is to assume that head movement is not necessarily followed by m-merger. Then a moved head could ‘‘excorporate’’ from its target and raise higher, creating the impression of long head movement, much as successive-cyclic *wh*-movement creates the impression of unbounded dependencies.

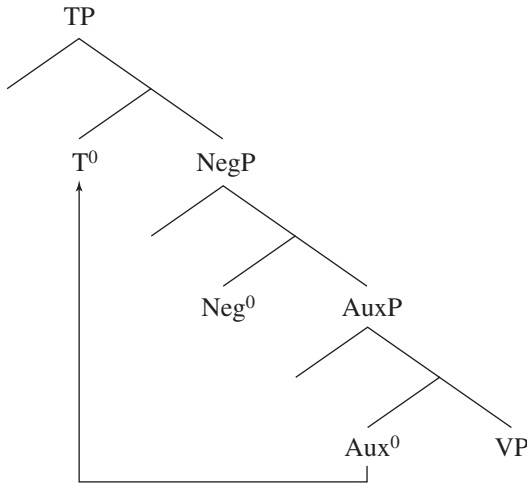
Fortunately, a third option is possible for some of the cases, allowing us to claim that the Head Movement Constraint is not violated there, contrary to appearances. One such case is the raising of the main auxiliary across negation in English.

- (21) a. Shouldn't Miranda sail a yacht across the Atlantic?
- b. Should Miranda not sail a yacht across the Atlantic?

(21a) demonstrates that negation can be a head in English, while (21b) shows that it does not have to raise along with the main auxiliary to C<sup>0</sup>. The problem that these examples pose for the

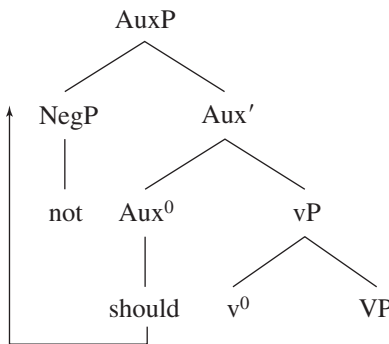
Head Movement Constraint is that head movement to  $T^0$  seems to skip over  $Neg^0$ , which intervenes between  $T^0$  and the auxiliary's base-generation site.<sup>25</sup>

(22) *Standardly assumed structure for English negation*



The approach proposed here allows for a possible solution that does not require acceptance or rejection of the Head Movement Constraint. Suppose that Germanic negation is a specifier of some IP layer rather than the head of its own maximal projection  $NegP$ , as in (23a). That this is a plausible hypothesis is indicated by languages such as Icelandic, Faroese, and Old Scandinavian, where negation is a phrasal adverb (see Holmberg, to appear, and references cited there). Then, being a head as well as a maximal projection, English negation may m-merge with the head in whose specifier it appears and therefore with the main verb that raises to this head, as in (23b). When m-merged,  $NegP/Neg^0$  will be pronounced as *n't*.<sup>26</sup>

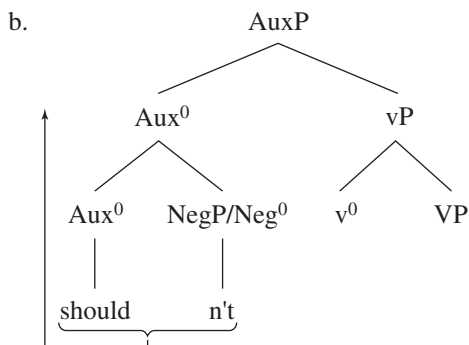
(23) a.



<sup>25</sup> The assumption that  $NegP$  is merged above  $TP$  does not resolve the problem, which would then be necessarily replicated with respect to  $T^0$ -to- $C^0$  movement (subject-aux inversion in English and V2 in Germanic languages).

<sup>26</sup> It should be noted that if  $Neg^0/NegP$  is assumed to be higher than  $T^0$ , the Mirror Principle (Baker 1985) is satisfied without any need to assume a higher functional head for the verb to move into.





A similar phenomenon is that of separable prefixes/particles in German and Dutch, which are left behind when the lexical verb moves. If these prefixes/particles are heads taking VP as a complement, their “stranding” is evidence against the Head Movement Constraint. However, if the approach outlined above can be applied to separable prefixes/particles (i.e., if they are simultaneously maximal projections and heads that can m-merge with the verbal root under some circumstances), their behavior will not provide evidence against the Head Movement Constraint.

A different phenomenon is fronting to the sentence-initial position in some languages, which can target both heads and maximal projections (long head movement). In Irish, the main verb (a head) competes with other predicates (XPs) for this position (e.g., McCloskey 1979, Chung and McCloskey 1987, Carnie 1995). The same is true for stylistic fronting in Icelandic, Faroese, and Old Scandinavian, where in clauses with a subject gap the subject position is targeted by both XPs and X<sup>0</sup> categories such as participles, adjectives, and particles (see Holmberg 2000b and references cited there). Other languages claimed to have long head movement include Albanian, Romanian, Breton (Borsley, Rivero, and Stevens 1996), Mainland Scandinavian (Toyoshima 2001), Modern Greek, Bulgarian (Lema and Rivero 1990, Bošković 1997, Toyoshima 2001; though see Embick and Izvorski 1995 for a counterargument), and Yiddish (Toyoshima 2001). Many of the cases where head movement alternates with phrasal movement have been argued to involve remnant movement,<sup>27</sup> but I cannot address the matter in depth here. If after careful consideration it turns out that there are indeed cases where phrasal movement and head movement target the same landing site, the complementarity of head movement and phrasal movement cannot be maintained.

To summarize, separating head movement into movement and m-merger allows us to accommodate Germanic negation and separable prefixes without recourse to long head movement. If we retain the assumption that movement of a head is subject to the Transparency Condition and stipulate that m-merger of a moved head is not obligatory, a moved head can “excporate” from its target and raise further. Such optionality of m-merger after movement of a head would permit us to incorporate long head movement into the theory. The Head Movement Constraint

<sup>27</sup> Toyoshima (2001) argues against the competing VP-to-Spec,CP remnant movement analysis by showing that the phenomenon is possible in languages (e.g., Mainland Scandinavian) where there is no scrambling (and thus arguments cannot be extracted out of VP prior to its movement to Spec,CP). Further research is required.

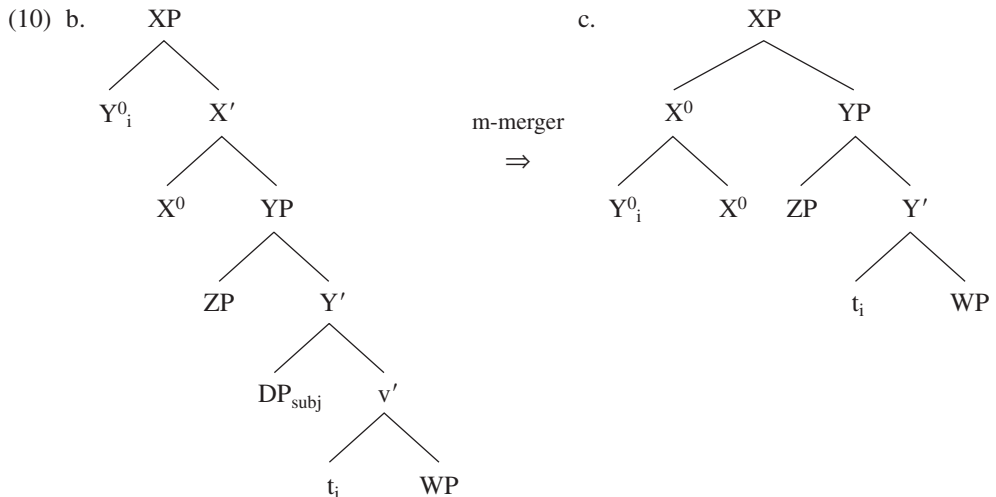
then must be a mere coincidence. Naturally, from the theoretical standpoint it would be more desirable if moving a head overtly was completely motivated by morphology: since a head moves in order to m-merge, it always does.

At this point, I cannot offer an educated argument for whether long head movement is or is not a counterexample to the Head Movement Constraint, so the issue remains.

### 5.2 What Is M-Merger?

This section deals with the question raised by Carnie 1995: How does syntax know if a node is a head, especially with derived heads? The answer to this question is very simply the answer to the question of what m-merger is. In other words, what does m-merger do that makes its result a head?

The absolutely minimal assumption required about m-merger is that it converts the structure in (10b) to the one in (10c) (repeated here). The derived head in (10c) can then be attracted by the next merged head in the following cycle of the derivation.



In other words, the input to m-merger is two heads in a particular (specifier-head) configuration and its output is one head. The major consequence of m-merger that we would like to derive from independent factors is the fact that excorporation out of the derived head is impossible. One way of doing so is by assuming that m-merger is an operation of the morphological component and as a result the internal structure of a derived head is syntactically opaque.

Suppose the conjecture that syntax cannot move features (Chomsky 1995) is correct and due to the fact that feature bundling (i.e., creation of heads) takes place in the morphological component. If m-merger and presyntactic bundling are the same operation occurring at different points in the derivation, then feature bundles created by m-merger (i.e., derived heads) will not be transparent to syntax either. We will then derive the impossibility of excorporation from a more general principle disallowing feature movement.

*5.2.1 Merge and Spell-Out* If m-merger is an operation of the morphological component, it must involve Spell-Out as a subcomponent; that is, the head created by m-merger is syntactically opaque (Eddy Ruys, pers. comm.). Syntactic opacity resulting from Spell-Out means that the relevant constituent remains accessible to syntax in the sense that it can be moved, but its internal structure is frozen and none of its subconstituents can be extracted. However, in order to preserve the intuition that a head created by m-merger (just like a simplex head inserted from the numeration) forms part of the input to the next syntactic cycle (merger of a specifier or merger of the next head), we need to assume a strongly cyclic view of syntax, where each newly merged node is a phase (Merge and Spell-Out).<sup>28</sup>

If m-merger involves (partial) Spell-Out, excorporation out of a derived head is disallowed, thus yielding its syntactic atomicity (Carnie 1995). A more general result is that feature extraction out of a derived head is impossible as well (though nothing is said about simplex heads, on which see section 5.2.2).

Under this view, a head created by m-merger is a syntactic phase (cf. Marantz 2001, Marvin 2002). Although a phase is spelled out (syntactically opaque), its lexical content and surface position are underdetermined. A parallel may be drawn with vP-fronting.

(24) Shut the gate the gatekeeper certainly did [~~shut the gate~~].

Once the vP phase is completed, the VP is spelled out. However, this does not mean that its phonology is fixed in situ, since the vP containing that VP can still be fronted, as in (24). In a way parallel to m-merger, Spell-Out fixes the featural content of the head, but not its lexical content (i.e., minimally, whether it is overt or covert) or its surface position.

To summarize, the assumption that m-merger includes (partial) Spell-Out allows us to derive the syntactic atomicity of derived heads. Unless this strongly cyclic view of syntax (Merge and Spell-Out) has undesirable consequences elsewhere (which is still something to check), it permits us to derive some properties of head movement from the assumption that m-merger takes place in a different module.<sup>29</sup>

*5.2.2 M-Merger as a Morphological Operation* We have defined heads as syntactically opaque feature bundles in (1). To ensure that derived heads are also treated as syntactically opaque feature bundles, we must define m-merger as the operation that creates feature bundles, as opposed to the syntactic Merge, which creates more complex structures. But such a definition means that m-merger does in syntax what also has to be done presyntactically, and that m-merger and presyntac-

<sup>28</sup> One possible implementation of this is to have partial PF and LF Spell-Out at each Merge under the assumption that nodes with no uninterpretable features can be removed from the narrow syntax, though their eventual position and lexical content would still be only partially fixed (cf. Fox and Pesetsky, to appear). This implies a certain analogy to cyclic versus postcyclic phonology (e.g., Chomsky and Halle 1968, Pesetsky 1979, Halle and Vergnaud 1987) along the lines suggested by Noyer (1997). The process resulting in the Transparency Condition could be a component of this partial Spell-Out.

<sup>29</sup> A further advantage to this view of syntax is that it leads us to expect both semantic and phonological derivations to follow Merge, resulting respectively in compositionality and cyclicity.

tic bundling can be viewed as the same basic operation, defined for different environments (see also footnote 23). But if *m*-merger is a morphological operation, transition to the morphological component would require (partial) Spell-Out and therefore yield the results just discussed in section 5.2.1. It has, furthermore, additional advantages.

If *m*-merger is a morphological mechanism, it does not have to be subject to such syntactic conditions as *c*-command/the Extension Condition, which removes one of the problems discussed in section 2.1. On the other hand, because *m*-merger occurs in the morphological component of the grammar, it is unsurprising that it should create an item whose internal structure is syntactically nontransparent (i.e., a head): if the morphological module is independent of syntax, the feature bundles it creates need not be analyzable by syntax. If a bundled feature is not extractable (per Chomsky's (1995) conjecture that minimal movable elements are heads), lack of excorporation follows (since *m*-merged heads are simply feature bundles).<sup>30</sup> Furthermore, the *c*-command relation established between a moved head and its trace persists after *m*-merger: in syntax, a feature bundle represents a single node and it *does* *c*-command the trace of the moved head. Thus, the tree in (10c) fully complies with the *c*-command requirement even from the representational point of view.<sup>31</sup>

Finally, if *m*-merger is a morphological operation, we obtain a straightforward answer to the question of why it only applies to heads (and by extension, why head movement only targets heads). Since maximal projections are created in syntax, they cannot be targeted by morphology (unless, of course, they are simultaneously heads as well) because morphology operates on feature bundles, which maximal projections are not.

To reiterate, defining *m*-merger as a morphological operation automatically blocks excorporation and yields the earlier stipulation that movement cannot move features. Movement becomes generalized pied-piping all the way down.

Furthermore, when we consider what *m*-merger does, it becomes clear that it and presyntactic bundling have much in common with the syntactic operation Merge. It seems tempting to define Merge as a general operation of the grammar, occurring in more than one module (cf. Embick and Noyer 2001). We can then treat *m*-merger and *s*-merger as its morphological and syntactic applications, respectively, thus making the central assumption of Distributed Morphology, Hierarchical Structure All the Way Down (Harley and Noyer 1999), follow from the fact that the same operation is used in different modules—that is, from the organization of the grammar.<sup>32</sup>

<sup>30</sup> The relation between a feature and the head containing it is best analyzed by the following analogy: Let's imagine that we want a white surface to write upon. We look around and see a cube, one of whose faces is white. The only way for us to access this white surface is to pick up the whole cube.

<sup>31</sup> An interesting question is whether *m*-merger is not just a special case of Morphological Merger (Marantz 1984). At first glance, *m*-merger appears to be more restricted, since it only applies to heads, is strictly cyclic, and operates under a much stronger notion of adjacency because it allows no intervention. However, locality conditions on an operation like Morphological Merger may change as a function of the stage of the derivation where the operation takes place (see Embick and Noyer 2001 for a discussion).

<sup>32</sup> As suggested above, identification of presyntactic feature bundling with *m*-merger is also a promising direction to pursue. Of course, then we can ask if there exists a phonological side of Merge, a *p*-merger.

Crucially, though from the morphological point of view there are levels of hierarchy within a derived head, from the syntactic perspective there is only one item: a feature bundle. Whether the bundling has occurred in the lexicon (presyntactic bundling) or as a result of m-merger, no internal hierarchy level within the head can be syntactically accessed, though features forming a (derived) head are visible to syntax. In other words, there is no need to explain how the interpretable V-feature ([iV]) in the derived  $T^0$ - $v^0$ - $V^0$  complex is visible to narrow syntax—it is simply visible in the same way it is visible in  $v^0$ . Once m-merger is defined as an iterative operation (as it must be), any number of heads can be involved.

I conclude that the necessity of treating derived heads as heads makes the existence of m-merger an obligatory feature of the linguistic theory. If we adopt some version of the strongly cyclic ‘Merge and Spell-Out’ view of syntactic derivation, the syntactic opacity due to m-merger can be derived from the stipulation that it does not apply in narrow syntax. This stipulation is however not essential to the view of m-merger as the second operation involved in head movement.

## 6 Conclusion

The unified theory of movement formulated above claims that all movement consists of Search-F (feature valuation) + Merge. Search-F can be Agree or C-Select (and possibly others; Scope, for QR/*wh*-movement, is one candidate). On the other hand, Merge is a basic operation of syntax, which takes two syntactic trees (a single head is a trivial tree) and merges them at the root (and thus obeys the Extension Condition). Head movement is required in any syntactic framework operating on the assumption that movement is Search-F + Merge and taking heads to be minimal syntactic elements.

Search-F targets features, which cannot be manipulated by Merge (s-merger, to be precise). The generalized pied-piping operation ensures that when an item cannot undergo Merge, the smallest item containing it is targeted instead. Accessibility is directly related to projection: once the C-Select relation is established, the nonprojecting category is no longer transparent. This is why locality conditions on c-selection and head movement are the same.

When movement places two feature bundles (heads) into a head-adjacency configuration, m-merger must (or may) occur. M-merger is an independently justified morphological operation, possibly equivalent to Halle and Marantz’s (1993, 1994) presyntactic bundling, that takes two (possibly trivial) feature bundles and returns one. Independent motivation for m-merger comes from the fact that it may apply in the absence of prior movement (e.g., with the Danish definite suffix) or with phrasal movement (as with Romance clitics). Because m-merger belongs to the morphological component of the grammar, the inner structure of the item that it creates (i.e., a head) is syntactically opaque. We thus explain, via the modular nature of the grammar, the lack of excorporation in head movement and Chomsky’s (1995) stipulation that features cannot be moved: the inner structure of a head is not syntactically accessible.

To summarize, we have derived the phenomenon of head movement from the interaction of four processes: feature valuation (C-Select), generalized pied-piping, Merge, and m-merger. This proposal permits us to return head movement to the domain of the narrow syntax.

The theory of the grammar that emerges out of the unified theory of movement suggests that there are several search relations that can be established in the syntax: C-Select, Agree, and so on. Each of them can be followed by (Re)Merge. In turn, (Re)Merge is an all-pervasive operation of the grammar, manifesting itself both in syntax (s-merger) and in morphology (m-merger).

The major theoretical innovations proposed here are a unified approach to phrasal movement and head movement, the relation between head movement and c-selection, and m-merger.

### **Appendix: Head Movement as Phonological Movement**

In this appendix, I will discuss the latest attempt to tackle the problems that the special properties of head movement discussed in section 2.1.2 pose for linguistic theory. Chomsky (2000) proposes that head movement (again, with the possible exception of incorporation) occurs on the phonological branch of the derivation, after Spell-Out. As a result, head movement is not reflected at LF and thus no longer expected to have semantic effects (property (e)). Since LF movement occurs in the tree that is already constructed, it is not expected to extend the target either (property (f)), and so head movement may escape c-command.

Apart from the pretheoretical problems with the proposal (see Embick and Noyer 2001, Zwart 2001), it gives rise to certain suspicious coincidences. Chomsky (2001:38) argues that ‘‘overt V-to-T raising, T-to-C raising, and N-to-D raising are phonological properties, conditioned by the phonetically affixal character of the inflectional categories.’’ This is followed by ‘‘Considerations of LF uniformity might lead us to suspect that an LF-interpretive process brings together D-N and C-T-V . . . to form wordlike LF ‘supercategories’ in all languages, not only those where such processes are visible.’’

To paraphrase: on the one hand, the C-T-V and D-N complexes are assumed to be phonologically associated by head movement; but on the other hand, the same complexes are also assumed to be associated by an LF process. Importantly, these two processes are in no way connected.

What Chomsky is creating here is a very interesting loophole, provided by the assumption that syntactic or similar-to-syntactic operations can occur on both branches of the derivation after Spell-Out. In the absence of a clear definition of the PF branch, what looks like a single pre-Spell-Out movement operation may in fact turn out to be an Agree operation at LF happening in parallel with an overt movement operation on the phonological branch. In other words, every case of pre-Spell-Out movement can potentially be reanalyzed as a combination of independent operations on the two branches of the computation, for phrasal movement as well as for head movement—a far from desirable development in the theory.

In the rest of this appendix, I will provide more concrete evidence against analyzing head movement as a phonological phenomenon. My objections question the source of this restriction and its consequences.

#### *A1 Exclusion from Narrow Syntax*

In the minimalist framework, properties of operations should follow from interface conditions. Under the assumption that all movement is a combination of Agree and (Re)Merge, the postulated

“phonological” nature of head movement can be due to the properties of either Agree or (Re)-Merge. In other words, if we want to exclude head movement from the narrow syntax, either Agree or (Re)Merge should prevent heads from being able to move.

The search operation Agree targets features and thus cannot constrain head movement to not apply before Spell-Out. Moreover, Agree is known to take place both before and after Spell-Out, and so an operation based on Agree cannot be forced to apply only after Spell-Out.

Likewise, there is no immediately obvious reason why (Re)Merge should fail to target heads. In fact, since in order to create a syntactic tree it is necessary to merge heads (terminal nodes) with something else, it should also be possible to remerge them.

### *A2 Behavior on the Phonological Branch*

Suppose there exists a reason why (Re)Merge cannot target structures smaller than maximal projections. Will the assumption that head movement occurs on the phonological branch of the computation allow us to account for its special properties, such as the impossibility of excorporation (a), stricter locality (c), and so on? I believe the answer is no, especially since there are no independently established properties of post-Spell-Out operations taking place on the phonological branch.

One problem that is not removed by postulating that head movement applies after Spell-Out is how it differs from phrasal movement. Chomsky (2001a,b) contends that certain types of phrasal movement can also occur on the phonological branch of the computation. If both head movement and phrasal movement can occur after Spell-Out, both the choice of the moved element (a head or a phrase) and the difference in the locality constraints, as well as the nature of these constraints, remain unexplained.

Likewise, one of the main features of the minimalist framework is that syntactic movement is feature-driven (Agree). Relegating head movement to the phonological branch does not remove the question of the trigger. On the other hand, if phonological head movement is an exception in not being feature-driven, the question arises what triggers phonological movement.

Finally, if head movement is obligatorily a post-Spell-Out operation, the phonological branch of the computation must contain an operation absent from pre-Spell-Out syntax. It then becomes an open question to what extent the phonological branch of the computation is syntactic.

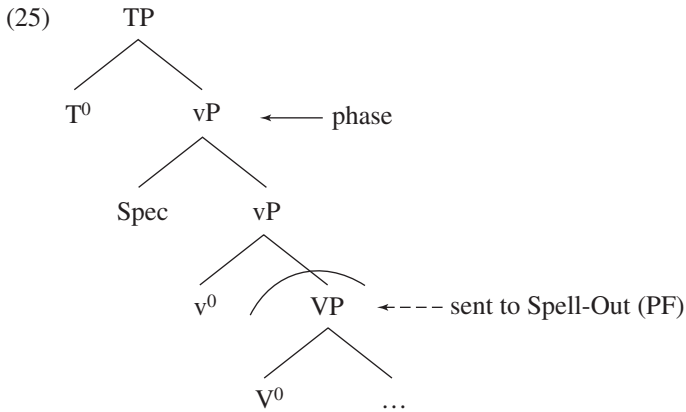
I conclude that postponing head movement until after Spell-Out is not likely to solve the problems that head movement creates. In the next section, I will show that the Phase Impenetrability Condition (Chomsky 2000) makes it impossible for head movement to occur after Spell-Out.

### *A3 Head Movement and the Phase Impenetrability Condition*

Chomsky (2000) describes phases as minimal domains that contain a  $C^0$  or a  $v^0$  (i.e., phases are propositional). Functionally, they are chunks of syntactic derivation, which are spelled out in a cyclic fashion (i.e., a phase is syntactically impenetrable from outside). To ensure the possibility of extracting out of a phase, Chomsky assumes that when a phase HP is sent to Spell-Out, what is actually removed from the computation is the sister of  $H^0$ . Both  $H^0$  and the specifier(s) of HP remain until the next phase.



Now consider verb raising in a structure like (25).



According to the definition by enumeration given above, a vP is a phase, which means that what is spelled out is the sister of  $v^0$  (the VP), with the edge of the phase ( $v^0$  and Spec,vP) left until the next phase.

Now consider what happens in French, where the main verb raises and adjoins to  $T^0$  (Pollock 1989). Given that (at least) vP intervenes between TP and VP,  $V^0$ -to- $v^0$  movement is obligatory prior to verb movement to  $T^0$ ; if it did not take place, the Head Movement Constraint would be violated. This derivational step is unavoidable, whether we consider  $v^0$  to be the verbalizing affix (Marantz 1997, Chomsky 2000) or the functional head introducing the external argument (Kratzer, to appear), since the phonological features of a verbal stem must originate at least as low as  $V^0$ . The only question is, when does  $V^0$ -to- $v^0$  movement occur?<sup>33</sup>

Since head movement is assumed to occur on the phonological branch of the computation,  $V^0$ -to- $v^0$  movement must take place after Spell-Out. At Spell-Out, the VP is sent off to PF, while the vP is not. Therefore, the edge of the current phase,  $v^0$ , is not accessible for phonological movement, because it is still in the pre-Spell-Out stage of the computation!

Suppose that  $V^0$ -to- $v^0$  movement takes place when the next (CP) phase is sent to PF. Then the phonological computation must deal with two (or more) phases at once, which defeats the original (declared) purpose of introducing phases, which is to reduce the computational burden. If such movement is available in the phonological component, the computational burden of dealing with more than one phase after Spell-Out cannot be the same as before Spell-Out, unless some additional assumptions are made.

Importantly, our conclusion from  $V^0$ -to- $v^0$  movement in French should not be that head movement always happens before Spell-Out. Rather, what I have shown is that if a head must

<sup>33</sup> Chomsky (2000) suggests that  $V^0$ -to- $v^0$  movement is obligatory in all languages. Under his view, this movement is motivated by the necessity of turning a category-free root in  $V^0$  into a syntactic/morphological verb (cf. Marantz 1997). This assumption also seems difficult to square with the proposal that head movement, including  $V^0$ -to- $v^0$  movement, takes place on the phonological branch of the computation.



move across a phase boundary, it must do so in the narrow syntax. In other words, this argument is not against head movement at PF; it is an argument against head movement at PF *only*.

Suppose, however, that ‘movement’ operations at PF are not syntactic and therefore not constrained by the same costs as pre-Spell-Out movement. There would then be no reason to believe that two phases cannot be handled at once by the phonological component of the grammar.

The weakness of this argument is that it shifts the burden of explanation from narrow syntax to the phonological branch of the derivation, without defining the properties of the PF branch (Zwart 2001). In the next section, I will show that head movement does not behave as if it were a phonological phenomenon.

#### A4 Head Movement and Phonology

If head movement is phonological, we expect it to be sensitive to or be triggered by some phonological properties of either the attracting head or the moving head. These expectations are not realized.

German verb-second (V2) is head movement to a position that is not phonologically realized, as is  $V^0$ -to- $v^0$  raising (which Chomsky assumes to be universal). Therefore, the ability of a head to attract another head is independent of its phonological realization: a null head may be an affix (in the sense that it triggers overt movement to itself) or may not be one: thus,  $C^0$  is null in English and in German, but only German is V2 in matrix clauses. Likewise, head movement to filled heads (*adjunction*, as opposed to *substitution*) is also possible (e.g., verb movement to tense), showing that phonological realization of a head does not influence its ability to trigger head movement.

Conversely, there exist heads, such as Romance definite articles, that are clearly phonological clitics and yet do not trigger head movement to themselves. This shows that having a phonological dependency is not equivalent to being an affix, which is what we would have expected if head movement were phonological.

An alternative possibility is that it is the phonological properties of the *attracted* head that determine whether it is raised. This approach seems ad hoc in the case of verb movement. A claim that verbal stems are phonological affixes in French (obligatory  $v^0$ -to- $T^0$  movement) but not in English (nonauxiliary verbs stay in situ) means postulating the very property we are trying to explain. Moreover, assigning such a property to the attractee predicts that there should only be two types of verb raising:  $V^0$ -to- $C^0$  movement, as in German, and verb-in-situ, as in English; if it is the properties of the verb that determine whether it raises, then it raises either all the way or not at all. An obvious counterexample to this claim is French, where the main verb is assumed to raise to  $T^0$ , but not any further.

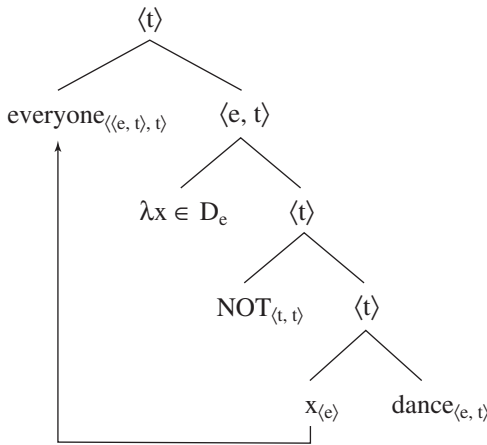
So if the phonological properties of either the attracting or the attracted heads play no role in head movement, it becomes unlikely that head movement is phonological in nature. In the next section, I will show that the (apparent) absence of LF effects of head movement (property (e)), which Chomsky uses as an argument in favor of his hypothesis, is due to an independent factor and leads to further and unexplored consequences in other cases of movement without obvious LF effects, that is, reconstructing movement.

### A5 Other Effects of Head Movement

A strong argument for relegating head movement to the realm of post-Spell-Out operations is the apparent absence of LF effects: raised verbs in French and in-situ verbs in English have the same meaning.

Chomsky suggests that head movement must be a post-Spell-Out operation as a result of semantic economy. He attributes the absence of semantic effects for head movement to a design feature of the language faculty: ‘‘The interpretive burden is reduced if, say, verbs are interpreted the same way whether they remain in situ or raise to T or C’’ (Chomsky 2001b:37). This is a valid reason, and this statement is trivially true for any lexical item in the formal semantics of (e.g.) Heim and Kratzer (1998): the meaning of an item does not change with position. Looking at (26), consider how scope is achieved in this framework via syntactic movement.

#### (26) Subject scope (simplified)



The meaning of the generalized quantifier *everyone* is the same, whether it stays in situ or moves.

$$(27) \llbracket \text{everyone} \rrbracket = \lambda f \in D_{\langle e, t \rangle} . \forall x \in D_e . f(x) = 1$$

When a generalized quantifier moves, it may leave behind a trace of the type  $\langle e \rangle$ , which is abstracted over in the landing site of the quantifier (26). The meaning of all the parts remains the same, but the order of their combination is different: as a result, the generalized quantifier now takes scope over negation. In other words, it can be trivially true that the meaning of an item would not change with position and yet movement can change the meaning of the totality.

Conversely, if Chomsky’s statement were taken to mean that movement accompanied by any change of meaning should be restricted from applying to some lexical items or even simply to heads, the line becomes very difficult to draw, since so many items are simultaneously a head and a maximal projection (*what* is one example). We therefore cannot rule out pre-Spell-Out head movement on the basis of semantic economy.

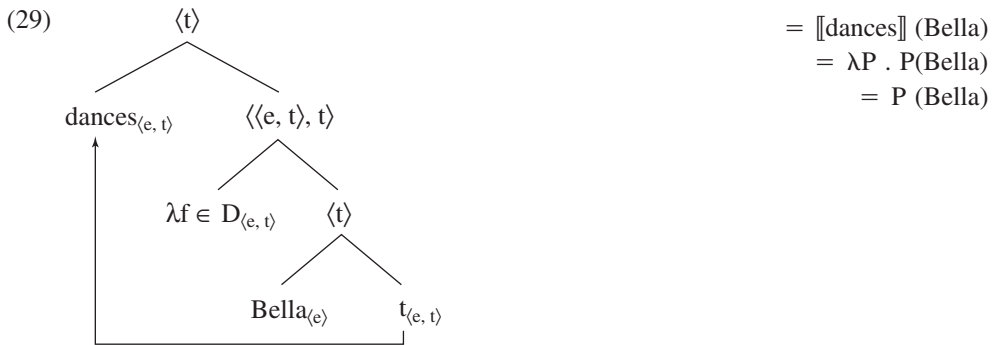
Furthermore, lack of semantic effects cannot be a diagnostic for post-Spell-Out movement, since then all reconstructing movement should be treated as phonological (as suggested in Sauer-

land and Elbourne 2002 for phrasal movement), which would be incompatible with other assumptions that Chomsky makes. Consider once again the movement of the subject over negation in (26). The surface scope is usually assumed to be due to EPP-triggered movement to Spec,TP, and the inverse scope is achieved via reconstruction.

(28) Everyone didn't dance.  $\forall > \neg$  (surface scope),  $\neg > \forall$  (reconstructed)

EPP-triggered movement must be analyzed as pre-Spell-Out, since EPP is an uninterpretable feature and would cause the derivation to crash at PF. Nevertheless, the subject may reconstruct under negation, which means that its movement to Spec,TP should be viewed as phonological! I conclude that, keeping all assumptions constant, lack of semantic effects says nothing about when movement occurs.

Why is it the case, then, that head movement has no LF effects? The reason lies in the nature of the items moved by head movement. It is easy to see that most of them are predicates of some sort (verbs, nouns, most affixes, etc.). Consider now how the moved predicate *dances* in (29) might be interpreted, leaving aside for the moment whether it moves by head movement or not.



Assuming the same system as above, a moved predicate leaves behind a trace. This trace should be able to semantically combine with the subject and therefore must be interpreted as a variable of the same type as the moved verb:  $\langle e, t \rangle$ . This variable is then abstracted over and the whole tree is combined with *dances*, and the end result is as if the verb had never moved. In other words, whether we assume that predicates must reconstruct (the standard description of the effect, based on the empirical generalization) or allow them to be interpreted in their final position, the outcome is the same: predicate movement is not reflected at LF. Most important, this fact is independent of the nature of the movement: we have used exactly the same mechanism as the one we have used for DP-movement. Given that most heads (and in particular verbs) are predicates, the lack of semantic effects of head movement is now unsurprising.

Other items claimed to be displaced by head movement without any semantic effects are kinds (if noun incorporation as defined by Baker (1988) is indeed head movement). Movement of DPs denoting individuals, of which kinds are a special case, has no LF effect for the same reason: they have the same semantic type  $\langle e \rangle$  as their traces, so it is unsurprising that head movement will have no semantic effect in this case either.

We conclude that we should not even *expect* any semantic effects to arise from movement of nonquantificational heads. The next question to ask is whether there are heads with a meaning similar to *everyone*. To answer this question, we can examine the behavior of quantified heads, such as modals.

- (30) a. Yolanda can't leave. Neg > Mod  
 b. Yolanda shouldn't leave. Mod > Neg

The relative positions of the modal and negation are the same in (30a) and (30b): the modal is higher than the negation. However, while the existential modal *can* is interpreted in the scope of negation, the universal modal *should* must instead take scope over it.<sup>34</sup> Leaving aside the question of why universal and existential modals must take scope the way they do, how are the two scopes obtained?

Negation has the semantic type  $\langle t, t \rangle$ , and therefore its trace should have the same type in order to be interpretable. Covert movement of negation will thus have no semantic effect, just like movement of predicates (see also Lechner 2005 for an empirical argument against such Neg-shift). On the other hand, a modal like *should* or *can* quantifies over possible worlds (type  $\langle \langle s, t \rangle, t \rangle$ ) with a trace of the type  $\langle s \rangle$ , which means that its movement may change its scope (just as in (28)). In both examples in (30), the modal is higher than negation on the surface, but only in (30b) does the interpretation correspond to surface scope: in (30a), the modal is interpreted below negation. Since movement of negation has no semantic effect, the inverse scope in (28) (quantifier under negation) can only be obtained by the quantifier reconstructing to its original position below negation, and likewise for the modal in (30a). The only difference between (28) and (30) is that in (28), both surface and inverse scope were available for the same item (*everyone*); in (30), *can* must reconstruct, while *should* cannot.<sup>35</sup>

If heads can reconstruct, they are predicted to be able to undergo QR covertly as well. I still have to find an example of this (though see Benedicto 1997), but a careful examination of other quantified heads should produce the required results.

For the sake of completeness, I should note here that several syntactic theories have suggested that head movement may have an effect on the syntax of the moved item. Thus, Larson (1988) suggested that V-movement in VP shells is required for  $\theta$ -role assignment, Koopman (1984) proposed that V-movement is required for Case assignment, and Chomsky (2000; building on

<sup>34</sup> It should be noted that the difference in interpretation between universal (*must*) and existential (*can*) modals is not restricted to deontic modals. Given that epistemic modals appear higher in the structure than deontic ones, the QR/reconstruction head movement cannot be supplanted by assuming merger in different positions and syntactically different negations (a head vs. a specifier), as proposed by Holmberg (2002). On the other hand, it is not impossible that the divergent behavior of universal and existential modals with respect to negation is due to some sort of pragmatic *strengthening*, as has been argued for Neg-raising by Horn (1989) and for the scope of superlatives by Heycock (2003). In this case, of course, we have no argument for the semantic effects of head movement based on modals.

<sup>35</sup> Lechner (2005) uses the interaction between modals and scope splitting with negative universals (*not every student*) to argue that overt head movement of a modal (from  $T^0$  to  $\text{Agr}_S^0$ ) permits it to take scope over the quantified subject, and therefore head movement can affect interpretation. See also Lechner 2001, 2004 for an argument that some constraints on interpretation have an effect on applicability of head movement.

Marantz 1997) proposed that  $V^0$ -to- $v^0$  movement is needed to assign a syntactic category to the root. Likewise, Harley (2004; building on Chomsky 1993, Bobaljik 1995, Thráinsson 2000, etc.) argues that since Scandinavian object shift is dependent on head movement, head movement must be a syntactic operation, and the same is true for transitive expletive constructions, which can only appear in languages with  $V^0$ -to- $I^0$  movement as well as V2 (Vikner 1990). Unfortunately, these arguments very much depend on the specific theories they are realized in and thus cannot be considered decisive in any way.

#### A6 Summary

In this appendix, I have examined the assumption that head movement takes place not in the narrow syntax but rather on the phonological branch of the computation: after Spell-Out. This approach gives rise to a number of problems, not the least important of which is the fact that no properties of the postulated PF branch have been definitely established.

I have suggested that quantified heads, whose movement can have an effect on their interpretation, interact with other scope-bearing elements and that this interaction can be explained if head movement happens before Spell-Out. On the other hand, relegating head movement to the phonological component makes any movement across the phase boundary impossible. Finally, undergoing overt head movement appears to be totally unrelated to the phonological properties of either of the heads under consideration, further supporting the notion of head movement as a fully syntactic (i.e., potentially pre-Spell-Out) operation.

To conclude, relegating head movement to the phonological component of the grammar creates new problems and does not in any immediate way resolve the old ones. We still have to explain the differences between head movement and phrasal movement: the impossibility of excorporation (a), stricter locality (c), exemption from the Extension Condition (f), and the complementary distribution of adjunction to a head and movement into a specifier (g).

#### References

- Abney, Steven. 1987. The English noun phrase in its sentential aspect. Doctoral dissertation, MIT, Cambridge, Mass.
- Anderson, Stephen R. 1993. Wackernagel's revenge: Clitics, morphology and the syntax of second position. *Language* 69:68–98.
- Anderson, Stephen R. To appear. Second position clitics in Tagalog. In *The nature of the word: Essays in honor of Paul Kiparsky*, ed. by Sharon Inkelas and Kristin Hanson. Cambridge, Mass.: MIT Press.
- Baker, Mark. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16:373–416.
- Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Benedicto, Elena E. 1997. V-movement and its interpretational effects. Paper presented at GLOW 20, Rabat, Morocco, March 19–21, 1997.
- Benveniste, Emile. 1966. Structure de relations de personne dans les verbes. In *Problèmes de linguistique générale*, ed. by Emile Benveniste, 225–236. Paris: Gallimard.
- Beukema, Frits, and Marcel den Dikken. 2000. *Clitic phenomena in European languages*. Amsterdam: John Benjamins.

- Bobaljik, Jonathan David. 1995. Morphosyntax: The syntax of verbal inflection. Doctoral dissertation, MIT, Cambridge, Mass.
- Bobaljik, Jonathan David, and Samuel Brown. 1997. Interarboreal operations: Head movement and the Extension Requirement. *Linguistic Inquiry* 28:345–356.
- Boeckx, Cedric, and Sandra Stjepanović. 2001. Head-ing toward PF. *Linguistic Inquiry* 32:345–355.
- Borer, Hagit. 1983. *Parametric syntax*. Dordrecht: Foris.
- Borsley, Robert D., Maria-Luisa Rivero, and Janig Stevens. 1996. Long head movement in Breton. In *The syntax of the Celtic languages: A comparative perspective*, ed. by Robert D. Borsley and Ian Roberts, 53–74. Cambridge: Cambridge University Press.
- Bošković, Željko. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, Mass.: MIT Press.
- Carnie, Andrew. 1995. Non-verbal predication and head movement. Doctoral dissertation, MIT, Cambridge, Mass.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1994. Bare phrase structure. MIT Occasional Papers in Linguistics 5. Cambridge, Mass.: MIT, Department of Linguistics and Philosophy, MITWPL.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2001a. Beyond explanatory adequacy. MIT Occasional Papers in Linguistics 20. Cambridge, Mass.: MIT, Department of Linguistics and Philosophy, MITWPL.
- Chomsky, Noam. 2001b. Derivation by phase. In *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 1–52. Cambridge, Mass.: MIT Press.
- Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Chung, Sandra, and James McCloskey. 1987. Government, barriers, and small clauses in Modern Irish. *Linguistic Inquiry* 18:173–237.
- Cinque, Guglielmo. 2000. On Greenberg's Universal 20 and the Semitic DP. In *University of Venice working papers in linguistics 10*, 45–61. Availability information at [http://venus.unive.it/lingdida/modules.php?op=modload&name=EZCMS&file=index&print=1&page\\_id=48](http://venus.unive.it/lingdida/modules.php?op=modload&name=EZCMS&file=index&print=1&page_id=48).
- Collins, Chris. 2001. Eliminating labels. MIT Occasional Papers in Linguistics 20. Cambridge, Mass.: MIT, Department of Linguistics and Philosophy, MITWPL.
- Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.
- Delsing, Lars-Olof. 1993. The internal structure of noun phrases in the Scandinavian languages. Doctoral dissertation, University of Lund.
- Embick, David, and Roumyana Izvorski. 1995. On long head movement in Bulgarian. In *ESCOL '94*, ed. by Janet M. Fuller, Ho Han, and David Parkinson, 104–115. Ithaca, N.Y.: Cornell University, CLC Publications.
- Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32:555–598.
- Flagg, Elissa. 2002. Interface issues in the English imperative. Doctoral dissertation, MIT, Cambridge, Mass.
- Fox, Danny, and David Pesetsky. To appear. Cyclic linearization of syntactic structure. *Theoretical Linguistics*.
- Giorgi, Alessandra, and Giuseppe Longobardi. 1991. *The syntax of noun phrases: Configuration, parameters and empty categories*. Cambridge: Cambridge University Press.
- Grimshaw, Jane. 1991. Extended projection. Ms., Brandeis University, Waltham, Mass.
- Hale, Ken, and Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, Mass.: MIT Press.

- Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from Building 20: Essays on linguistics in honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, Mass.: MIT Press.
- Halle, Morris, and Alec Marantz. 1994. Some key features of Distributed Morphology. In *Papers on phonology and morphology*, ed. by Andrew Carnie and Heidi Harley, with Tony Bures, 275–288. MIT Working Papers in Linguistics 21. Cambridge, Mass.: MIT, Department of Linguistics and Philosophy, MITWPL.
- Halle, Morris, and Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, Mass.: MIT Press.
- Hankamer, Jorge, and Line Hove Mikkelsen. 2002. A morphological analysis of definite nouns in Danish. Ms., University of California, Santa Cruz.
- Harbour, Daniel. 2003. Elements of number theory. Doctoral dissertation, MIT, Cambridge, Mass.
- Harley, Heidi. 2004. Merge, conflation, and head movement: The First Sister Principle revisited. In *NELS 34*, ed. by Keir Moulton and Matthew Wolf. Amherst: University of Massachusetts, GLSA.
- Harley, Heidi, and Rolf Noyer. 1999. Distributed Morphology. *GLOT International* 4:3–9.
- Harley, Heidi, and Elisabeth Ritter. 2002. Structuring the bundle: A universal morphosyntactic feature geometry. In *Pronouns: Grammar and representation*, ed. by Heike Wiese and Simon J. Horst, 23–39. Amsterdam: John Benjamins.
- Heck, Fabian. 2004. A theory of pied-piping. Doctoral dissertation, Universität Tübingen.
- Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.
- Heycock, Caroline. 2003. On the interaction of adjectival modifiers and relative clauses. Ms., University of Edinburgh.
- Holmberg, Anders. 2000a. OV order in Finnish. In *The derivation of VO and OV*, ed. by Peter Svenonius, 123–152. Amsterdam: John Benjamins.
- Holmberg, Anders. 2000b. Scandinavian stylistic fronting: How any category can become an expletive. *Linguistic Inquiry* 31:445–483.
- Holmberg, Anders. 2002. Questions, polarity, and verb movement in Finnic and Germanic. Paper presented at Scandinavian Conference of Linguistics 19, Tromsø, January 10–12, 2002.
- Holmberg, Anders. To appear. Stylistic fronting. In *SynCom (The syntax companion)*, ed. by Henk van Riemsdijk and Martin Everaert. Oxford: Blackwell.
- Horn, Larry. 1989. *A natural history of negation*. Chicago: University of Chicago Press.
- Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30:69–96.
- Jaeggli, Osvaldo. 1982. *Topics in Romance syntax*. Dordrecht: Foris.
- Johnson, David, and Shalom Lappin. 1997. A critique of the Minimalist Program. *Linguistics and Philosophy* 20:273–333.
- Julien, Marit. 2000. Syntactic heads and word formation: A study of verbal inflection. Doctoral dissertation, University of Tromsø.
- Julien, Marit. 2002a. Determiners and word order in Scandinavian nominal phrases. *Studia Linguistica* 56: 265–315.
- Julien, Marit. 2002b. Inflectional morphemes as syntactic heads. In *Morphology 2000: Selected papers from the 9th Morphology Meeting, Vienna*, ed. by Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer, and Maria D. Voeikova, 175–184. Amsterdam: John Benjamins.
- Kayne, Richard S. 1975. *French syntax: The transformational cycle*. Cambridge, Mass.: MIT Press.
- Kayne, Richard S. 1991. Romance clitics, verb movement, and PRO. *Linguistic Inquiry* 22:647–686.
- Koopman, Hilda. 1984. *The syntax of verbs*. Dordrecht: Foris.
- Koopman, Hilda, and Anna Szabolcsi. 2000. *Verbal complexes*. Cambridge, Mass.: MIT Press.
- Kratzer, Angelika. To appear. *The event argument and the semantics of verbs*. Cambridge, Mass.: MIT Press.
- Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19:381–405.



- Lechner, Winfried. 2001. Reduced and phrasal comparatives. *Natural Language & Linguistic Theory* 19: 683–735.
- Lechner, Winfried. 2004. *Ellipsis in comparatives*. Berlin: Mouton de Gruyter.
- Lechner, Winfried. 2005. Interpretive effects of head movement. Ms., Universität Tübingen.
- Legendre, Géraldine. 1998. Second position clitics in a verb-second language: Conflict resolution in Macedonian. In *ESCOL '97*, ed. by Jennifer Austin and Aaron Lawson, 139–149. Ithaca, N.Y.: Cornell University, CLC Publications.
- Lema, José, and Maria-Luisa Rivero. 1990. Long head-movement: ECP vs. HMC. In *NELS 20*, ed. by Juli Carter, Rose-Marie Déchaine, Bill Philip, and Tim Sherer, 333–347. Amherst: University of Massachusetts, GLSA.
- Mahajan, Anoop. 2000. Eliminating head movement. *GLOW Newsletter* 44:44–45.
- Mahajan, Anoop. 2001. Word order and remnant VP movement. Ms., UCLA, Los Angeles, Calif.
- Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge, Mass.: MIT Press.
- Marantz, Alec. 1993. A late note on late insertion. In *Explorations in generative grammar: A festschrift for Dong-Whee Yang*, ed. by Young-Suk Kim, Byung-Choon Lee, Kyoung-Jae Lee, Hyun-Kwon Yang, and Jong-Yurl Yoon, 396–413. Seoul: Hankuk.
- Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *Proceedings of the 21st Annual Penn Linguistics Colloquium*, ed. by Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark, and Alexander Williams, 201–225. Pennsylvania Working Papers in Linguistics 4.2. Philadelphia: University of Pennsylvania, Penn Linguistics Club.
- Marantz, Alec. 2001. Words. Paper presented at the 20th West Coast Conference on Formal Linguistics, University of Southern California, February 23–25, 2001.
- Marvin, Tatjana. 2002. Topics in the stress and syntax of words. Doctoral dissertation, MIT, Cambridge, Mass.
- McCloskey, James. 1979. *Transformational syntax and model theoretic semantics: A case study in Modern Irish*. Dordrecht: Reidel.
- Nash, Léa, and Alain Rouveret. 1997. Proxy categories in phrase structure theory. In *NELS 27*, ed. by Kiyomi Kusumoto, 287–304. Amherst: University of Massachusetts, GLSA.
- Nilsen, Øystein. 2003. Eliminating positions. Doctoral dissertation, Utrecht University.
- Noyer, Rolf. 1997. *Features, positions and affixes in autonomous morphological structure*. New York: Garland.
- Nunes, Jairo. 1995. The copy theory of movement and linearization of chains in the Minimalist Program. Doctoral dissertation, University of Maryland, College Park.
- Nunes, Jairo. 1998. Bare X-bar theory and structures formed by movement. *Linguistic Inquiry* 29:160–167.
- Pancheva, Roumyana. 2002. The rise and fall of second-position clitics. Ms., University of Southern California, Los Angeles.
- Parrott, Jeffrey. 2001. A morphological theory of head movement. Ms., Georgetown University, Washington, D.C.
- Pesetsky, David. 1979. Russian morphology and lexical theory. Ms., MIT, Cambridge, Mass. Available at <http://web.mit.edu/linguistics/www/pesetsky/rusmorph.pdf>.
- Pesetsky, David. 1982. Paths and categories. Doctoral dissertation, MIT, Cambridge, Mass.
- Pesetsky, David. 1995. *Zero syntax*. Cambridge, Mass.: MIT Press.
- Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 355–426. Cambridge, Mass.: MIT Press.
- Poletto, Cecilia, and Jean-Yves Pollock. 2004. On the left periphery of some Romance *wh*-questions. In *The structure of CP and IP. The cartography of syntactic structures, vol. 2*, ed. by Luigi Rizzi, 251–296. New York: Oxford University Press.
- Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20:365–424.



- Richards, Norvin. 1997. Featural cyclicity and (anti-)superiority. Paper presented at the 21st Penn Linguistics Colloquium, University of Pennsylvania, February 22–23, 1997.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Rosen, Sarah Thomas. 1989. Two types of noun incorporation: A lexical analysis. *Language* 65:294–317.
- Sauerland, Uli, and Paul Elbourne. 2002. Total reconstruction, PF movement, and derivational order. *Linguistic Inquiry* 33:283–319.
- Shlonsky, Ur. 2002. The form of Semitic noun phrases. Ms., University of Geneva.
- Sichel, Ivy. 2002. Phrasal movement in Hebrew adjectives and possessives. In *Dimensions of movement: From features to remnants*, ed. by Artemis Alexiadou, Elena Anagnostopoulou, Sjeff Barbiers, and Hans-Martin Gärtner, 297–339. Amsterdam: John Benjamins.
- Sportiche, Dominique. 1983. Structural invariance and symmetry in syntax. Doctoral dissertation, MIT, Cambridge, Mass.
- Sportiche, Dominique. 1989. Le mouvement syntaxique: Contraintes et paramètres. *Langages* 95:35–80.
- Svenonius, Peter. 1994. C-selection as feature checking. *Studia Linguistica* 48:133–155.
- Thráinsson, Höskuldur. 2000. Object shift and scrambling. In *The handbook of contemporary syntactic theory*, ed. by Mark Baltin and Chris Collins, 148–202. Oxford: Blackwell.
- Toyoshima, Takashi. 2000. Heading for their own places. In *Proceedings of the 9th Student Conference in Linguistics (SCIL 9)*, ed. by Ljuba Veselinova, Susan Robinson, and Lamont Antieau, 93–108. MIT Working Papers in Linguistics 36. Cambridge, Mass.: MIT, Department of Linguistics and Philosophy, MITWPL.
- Toyoshima, Takashi. 2001. Head-to-Spec movement. In *The minimalist parameter: Selected papers from the Open Linguistics Forum, Ottawa, 12–23 March 1997*, ed. by Galina M. Alexandrova and Olga Arnaudova, 115–136. Amsterdam: John Benjamins.
- Travis, Lisa. 1984. Parameters and effects of word order variation. Doctoral dissertation, MIT, Cambridge, Mass.
- Vikner, Sten. 1990. Verb movement and the licensing of NP positions in the Germanic languages. Doctoral dissertation, University of Geneva.
- Wackernagel, Jacob. 1892. Über ein Gesetz der indo-germanischen Wortstellung. *Indogermanische Forschungen* 1:333–436.
- Zribi-Hertz, Anne. To appear. La syntaxe de l'article défini: De la structure à l'interprétation. In *Modèles syntaxiques*, ed. by Dan van Raemdonck. Gembloux, Belgium: De Boeck-Duculot.
- Zwart, C. Jan-Wouter. 2001. Syntactic and phonological verb-movement. *Syntax* 4:34–62.

UMR 7023

15, rue Catulienne

93200 Saint Denis

France

[matushan@univ-paris8.fr](mailto:matushan@univ-paris8.fr)

<http://alum.mit.edu/www/matushan/>