UCLA
UCLA Electronic Theses and Dissertations

Title
Head Movement in Narrow Syntax

Permalink
https://escholarship.org/uc/item/3fg4273b

Author
O'Flynn, Kathleen Chase

Publication Date
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Head Movement in Narrow Syntax

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Arts

in Linguistics

by

Kathleen Chase O'Flynn

2016

© Copyright by

Kathleen Chase O'Flynn

2016

ABSTRACT OF THE THESIS

Head Movement in Narrow Syntax

by

Kathleen Chase O'Flynn

Master of Arts in Linguistics

University of California, Los Angeles, 2016

Professor Hilda Koopman, Chair

The status of head movement has become controversial within current syntactic theory because its properties appear to be sufficiently dissimilar from those of phrasal movement that the two movement types must be governed by different mechanisms. Standard syntactic analyses within the minimalist framework seek to reduce this complexity either by relegating head movement to the phonological component of the grammar, or by reanalyzing purported cases of head movement as phrasal movement. In this thesis I propose a version of head movement in narrow syntax that is internally coherent, and I show that it is compatible with standard minimalist theories. Case studies of the syntax of Romance clitics demonstrate the potential utility of this approach, and I argue that the mechanisms proposed are likely to be computationally equivalent to better-studied versions of minimalist grammars (MGs) which do not include head movement. It remains an empirical question whether head movement is in fact required as a part of natural language grammars, but I conclude that there is no theory-internal reason why the possibility must be ruled out.

The thesis of Kathleen Chase O'Flynn is approved.

Edward L. Keenan

Pamela Munro

Hilda Koopman, Committee Chair

University of California, Los Angeles

2016

TABLE OF CONTENTS

1 Introduction **1**
1.1 Narrow syntax vs. the phonological component 4
1.2 Long Head Movement 6

2 Head movement as incorporation: Roberts (2010) **8**
2.1 Background 8
2.2 Agree as the sole mechanism of incorporation 10
2.2.1 Excorporation and pied-piping 15
2.2.2 Romance clitic climbing 17
2.2.3 Defining the phase edge 21
2.2.4 Losing the subset criterion 23
2.2.5 Phase impenetrability and locality 24

3 Head movement in minimalist grammars **26**
3.1 Minimalist grammars (Stabler 2001, 2010) 26
3.1.1 The shortest move constraint 29
3.2 Adding head movement 30
3.2.1 Persistent features 32
3.2.2 Defining the mover 33
3.2.3 Separation of HM and phrasal movement features 38
3.2.4 Head movement and the SMC 40
3.3 Romance clitic clusters 41
3.3.1 me le 45
3.3.2 le lui 47
3.3.3 ?lui le 50
3.4 Interim summary 54

4 Assessment **55**

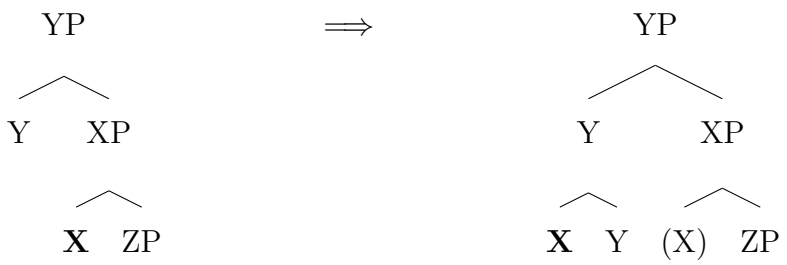
5 Conclusion **58**

6 References **60**

1 Introduction

Head movement is a seemingly simple operation which takes the head of one phrase and moves it by itself to adjoin to the head of a higher phrase. It is schematized below, where X moves as a head to Y:

(1)



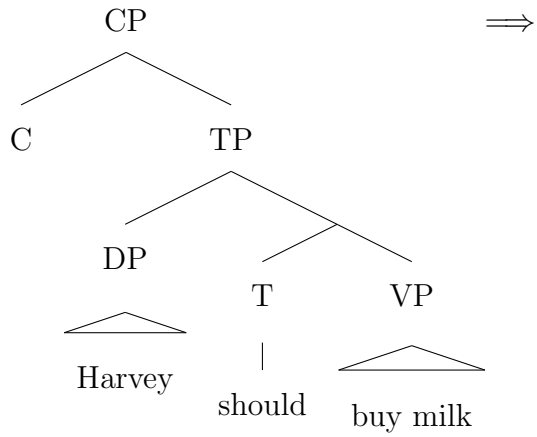
This kind of movement is familiar, and has long been standardly assumed to be the mechanism behind a large variety of syntactic phenomena, the core cases including French verb raising and English subject-auxiliary inversion, traditionally analyzed as V-to-T movement and T-to-C movement, respectively. For example:

(2) a. Harvey **should** buy milk.

b. **Should** Harvey __ buy milk?

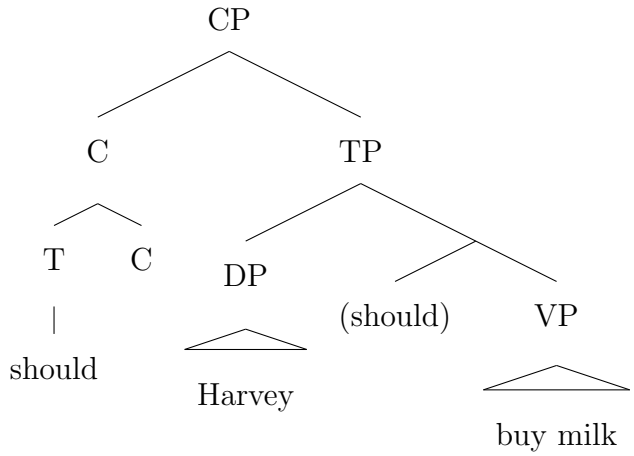
In (2b), the auxiliary *should* is a head, and it has clearly moved by itself, leaving behind its complement VP, *buy milk*. It is true that there may be ways to analyze it as phrasal movement instead, or even as something entirely other than movement. But in structural terms, the simplest analysis is to say that this is head movement, since the displacement of the head *should* is the only operation that we can directly observe. In head movement terms, then, the derivation of (2b) would look something like the following:

(3) a.



$\Longrightarrow$

b.



Despite the apparent simplicity of this approach and the long history of head movement analyses within the literature, the status of head movement has become controversial within current syntactic theory, and particularly within the minimalist research program initiated by Chomsky (1995, 2001). The argument is that head movement is too dissimilar from phrasal movement, having many properties that are incompatible with our understanding of the basic operations of the narrow syntax. Chomsky (2001) points out that, unlike any other narrow syntactic rule, head movement produces an adjunction structure where the moved head does not c-command its trace. Furthermore, head movement is countercyclic and subject to quite different locality restrictions than phrasal movement. In addition, the existing mechanisms of the theory provide no basis for the choice between a movement of a head and movement of an XP, since both must be triggered by the same kind of features.

Finally, and perhaps most importantly for Chomsky, the semantic effects of head movement are "slight or non-existent" in comparison with those that systematically hold for phrasal movement (38). That is, verbs are not interpreted differently in French, where they raise to T, than they are in English, where they remain in situ.

These kinds of challenges have prompted Chomsky and others to propose a variety of alternatives to narrow syntactic head movement, as traditionally understood. The solution offered in Chomsky (2001) is to relegate head movement to the phonological component of the grammar, as opposed to the narrow syntax. Other approaches include the reanalysis of certain claimed head movement phenomena as instead involving remnant XP-movement (Koopman & Szabolcsi 2000), or morphological solutions along the lines of Halle and Marantz (1993).

However, recent work by Matushansky (2006) and Roberts (2010) has called into question some of the common objections to narrow syntactic head movement, and proposed new ways of reconciling head and phrasal movement within the theoretical architecture of minimalist syntax. I believe that, before we choose to abandon head movement for one of the alternative approaches mentioned above, these proposals at least deserve some careful consideration. Thus I spend the bulk of this paper working out the formal implications of Roberts's proposal and my own variant of it, with an eye towards evaluating the overall plausibility of narrow syntactic head movement within minimalist grammar. But before diving into the specifics of Roberts's (2010) proposal, let us first review some of the arguments for why head movement should be included in the narrow syntax after all.

1.1 Narrow syntax vs. the phonological component

Chomsky (2001, 37-38) argues that most head movement processes[1] occur in the phonological component of the grammar, following Spellout, rather than in the narrow syntax. The foremost benefit of this proposal is that it can explain the lack of semantic effects of head movement, since in principle no change that occurs on the phonological branch of the derivation can affect LF. According to Chomsky, it can also explain why it is a head and not an XP that moves, since the movement is phonologically conditioned instead of being triggered by syntactic features. This makes it possible to retain an assumption that syntactic features only trigger phrasal movement, thus simplifying the syntactic component of the computation. Finally, phonological movement is not necessarily subject to the same constraints as syntactic movement, so any difficulties in accounting for the unusual structural properties of head movement – including its locality restrictions and the lack of c-command between head and trace – can be avoided by assuming that head movement occurs at PF, instead of in the syntax.

Matushansky (2006, 98-105) takes issue with all of these points, arguing first that, because there currently exists no serious proposal about the properties of the PF branch, simply relegating head movement to PF is not sufficient to explain most of its troubling properties. Rather, it merely pushes the same problems into a different part of the grammar. In addition, it is unclear why the narrow syntactic operations, (Re)Merge and Agree, should be unable to target heads, since these operations target syntactic features and features reside on heads. Indeed, it must be possible for (Re)Merge to target heads at some point, because otherwise no head could ever enter the derivation in the first place. Thus PF head movement does not solve the problem of why it is a head that moves in some cases and an XP in others. Nor is it clear that head movement is in any way sensitive to the phonological features of either

[1] With the possible exception of incorporation in the sense of Baker (1988).

the attracting head or the moving head. For instance, movement to a null head is possible, as in German V2, and by contrast, Romance definite articles fail to trigger movement to themselves, despite the fact that they are clitics and thus phonologically dependent on a host. This is not what we would predict if phonological features were relevant as triggers of head movement. In short, PF head movement does not actually address many of the challenges it has been claimed to solve. Indeed, Matushansky (2006, 100) points out that it may cause a new difficulty as well: post-Spellout movement of a head across a phase boundary will run afoul of the Phase Impenetrability Condition (PIC) because Spellout will make the moving head inaccessible to any attracting head within the higher phase. Thus, it is not clear that PF head movement is the right explanation for the unusual structural properties of head movement, or for its lack of semantic effects.

Matushansky (2006, 102-3) makes an alternative proposal about the lack of semantic effects: specifically, she argues that we should not expect to find such effects for head movement because what is moving in each case is just a lexical item, usually a predicate. In the terms of Heim and Kratzer (1998), the meaning of any lexical item will not change just because it moves, and any predicate that moves will leave behind a trace of the same semantic type, which can combine with other elements in the same way as the original predicate would. Thus movement of a predicate will not alter the interpretation of the totality, and so it is not surprising that head movement lacks semantic effects.

Roberts (2010) goes further than Matushansky, and points out one case where head movement appears to have a semantic effect after all: English subject-auxiliary inversion, which is traditionally analyzed as head movement, can license negative polarity items. As evidence, Roberts (2010, 10) cites these data, originally from McCloskey (1996, 89):

(4) a. *Which one of them does anybody like?

b. Which one of them doesn't anybody like?

See Roberts (2010, 10-12) for full argumentation that the negation, *-n't*, cannot cliticize to

C, but must rather cliticize to T and then move with the auxiliary to C, where it can c-command the NPI, *anybody*. If this argument is correct, and NPI-licensing is considered to be an LF effect, then movement of the negative auxiliary *doesn't* must occur in the narrow syntax, and not on the phonological branch following Spellout.

Altogether, I believe the arguments reviewed in this section offer sufficient motivation to at least reconsider the status of narrow syntactic head movement within minimalist theory. In the next section, I also reconsider one of the purported troubling properties of head movement: its strict locality restrictions.

1.2 Long Head Movement

Early understandings of head movement tended to assume it was subject to some version of the Head Movement Constraint (HMC), first proposed in Travis (1984):

(5) Head movement may not skip intermediate heads.[2]

Most phenomena that have traditionally been analyzed as involving head movement, such as V-to-T movement and T-to-C movement, do indeed seem to adhere to this constraint However, there is a substantial body of evidence suggesting that the HMC may not hold in all cases. Long head movement (LHM) seems to occur in Breton (Borsley, Rivero and Stephens 1996), as well as a number of Balkan, Slavic and Old Romance languages (Rivero 1991). A Slovak example:

(6) *Spytal* **som** sa ci **si** *napísal* list.
asked have+1SG REFL if have+2SG written letter
I asked if you wrote the letter. (Rivero 1991, ex. 2)

[2]Travis's (1984) version of the HMC was formulated in terms of governance. Many updated versions of the constraint have been proposed. This one comes from Matushansky (2006,74).

Here, the matrix V *spytal* precedes its auxiliary, *som*. Taking the embedded clause to represent the base order, where the V *napísal* follows the auxiliary, it is clear that *spytal* has moved to its observed position, crossing over the auxiliary to do so. If this is an instance of head movement, then it does not obey the HMC, since the auxiliary must be assumed to be a head. See Rivero (1991) for arguments that this is in fact long head movement, as opposed to: VP-preposing or movement to Spec,CP; stylistic fronting to Spec,IP; adjunction of V^o to IP; or short head movement.

Certainly, LHM analyses remain controversial. Matushansky (2006), for instance, considers the evidence for it "shaky" (89), and suggests that, with careful consideration, alternative analyses involving phrasal movement rather than long head movement may be found for many purported cases. But if we accept that cases like (6) do indeed involve long head movement, then any theory of head movement must offer sufficient freedom to allow such movement, while still being restrictive enough to capture the data that led to the original proposal of the HMC. There are various suggestions in the literature for how to achieve this. For example, Koopman & Sportiche (1986, 361) argue that the correct locality restrictions on long head movement essentially fall out from the ECP.

Roberts (2010) is a more recent attempt to formulate a theory of narrow syntactic head movement that is in line with standard minimalist assumptions and that also allows for the possibility of long head movement. Let us turn now to a careful consideration of Roberts' system.

2 Head movement as incorporation: Roberts (2010)

2.1 Background

The core of Roberts (2010) proposal is that head movement should be treated as similarly to phrasal movement as possible, and that it should fall out naturally from mechanisms that already exist within minimalist theory: specifically, from Agree. The theoretical assumptions underlying this idea are familiar, borrowed from Chomsky (2001), but it is worth reviewing them briefly to see how they lead into Roberts' proposal.

First of all, a *lexical item* is a set of features. Some of these features are *formal*, meaning they are relevant and available to syntactic operations, and some, such as the phonological features that code the pronunciation of the lexical item, are not. A *feature* is an ordered pair of attribute and value, $\langle Att, Val \rangle$. Val may be undefined, represented by __, as in $\langle Att, __ \rangle$, and features with an undefined value are called *uninterpretable*. Such features need to be checked (i.e. have their values filled in) by syntactic operations, such as Agree:

(7) An *Agree* relation "holds between terms α and β, where α has interpretable inflectional features and β has uninterpretable features" (Chomsky 2001, 3).

There are structural restrictions on where Agree may apply: one of the terms, called the *probe*, must c-command the other, called the *goal*. If there is more than one potential goal (i.e. more than one lower head that possesses the features being probed), only the closest goal may Agree; no intervening goals are allowed. When these conditions are met, *Match* applies:

(8) *Match*: given a well-formed Agree relation of which α and β are the terms (i.e. Probe or Goal) where α's feature matrix contains $\langle Att_i, __ \rangle$ and β's contains $\langle Att_i, val \rangle$, for some feature Att_i, copy *val* into __ in α's feature matrix. (Roberts 2010, 60, ex. 29)

According to Roberts, head movement is incorporation, and "incorporation is a way of

satisfying Agree that gives the effect of movement" (61). This movement effect occurs in exactly one case: when the goal of an Agree relation is defective. A goal G is *defective* with respect to its probe P iff G's formal features are a proper subset of P's. To see why this follows naturally from the definition of Match above, consider the trigger and outcome of an Agree relation where the goal is defective with respect to the probe:

(9) a. *Trigger for Agree*

Probe: $[\langle Att_1, __\rangle, \langle Att_2, __\rangle, \ldots]$ Goal: $[\langle Att_1\colon F\rangle, \langle Att_2\colon G\rangle]$

b. *Outcome of Agree*

Probe: $[\langle Att_1\colon F\rangle, \langle Att_2\colon G\rangle, \ldots]$ Goal: $([\langle Att_1\colon F\rangle, \langle Att_2\colon G\rangle])$

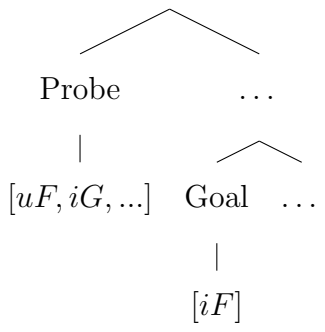
Because the features of the goal are a proper subset of those of the probe, copying them exhausts the content of the goal, effectively creating a copy of the entire goal at the position of the probe. In Chomsky (2001), phrasal movement too is assumed to work by creating a chain of copies, all but the highest of which will later be deleted via chain reduction (Nunes 2004). Roberts' proposal is that what results from the Agreement of a defective goal is indistinguishable from a chain of copies created by movement, and chain reduction may apply to it in just the same way.

Thus what we have here is a theory of head movement that appears to have a number of desirable properties. First, it is in line with standard minimalist assumptions about syntax. Also, it explains why it is a head that moves instead of an XP, since only defective goals can incorporate. Furthermore, it imposes locality restrictions on head movement (specifically, the locality restrictions that already apply to Agree), but it is not as strict as the Head Movement Constraint, thus leaving open the possibility of long head movement, as suggested by the constructions in Slavic and other languages, discussed in §1.2. However, on closer inspection, this theory suffers from several internal inconsistencies, which suggest that ultimately it may not be the best solution to the ongoing controversy surrounding head movement. I spend the rest of this section exploring these inconsistencies.

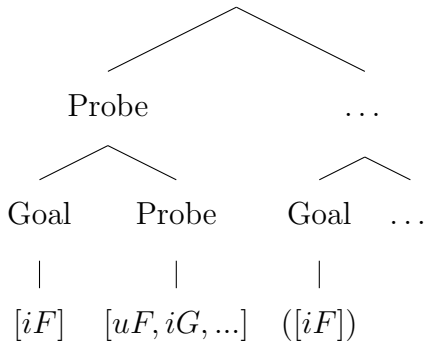
2.2 Agree as the sole mechanism of incorporation

Schematically, Roberts (2010) consistently represents the result of incorporation as a classic head adjunction structure, with the moved head left-adjoined to the probe. Starting with the structure in (10), he derives the structure in (11), where [uF] is a shorthand for an unvalued feature $\langle Att_F, __ \rangle$, [iF] abbreviates the valued feature $\langle Att_F, Val \rangle$, and parentheses indicate that a particular feature bundle will not be pronounced in that position (though it remains to be seen how exactly it is determined what should be pronounced where):

(10) *Trigger for incorporation:*

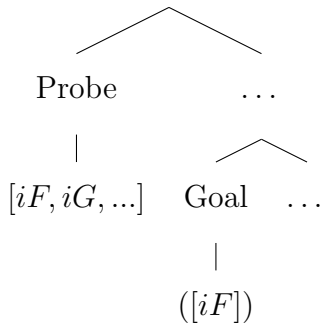


(11) *Result of incorporation (as head adjunction):*



However, if we are really to assume that Agree, or more specifically Match, is the only operation at play here, then the outcome is in fact represented by (12), rather than (11). In (12), just Val has been copied from goal to probe, essentially filling in the probe's existing feature matrix and turning [uF] to [iF]:

(12) *Result of incorporation (as a flat feature set):*



As it currently stands, this structure is lacking in one important respect: it says nothing about how the phonological material associated with the goal reaches its observed position on the probe. (Presumably the phonetic features of the goal are not a proper subset of those of the probe, and so are not copied along with its formal features.) It appears that there are two options for analysis here. First, it might be that a separate movement operation is triggered by the Agreement of the defective goal. This seems to be what Roberts actually intends, judging by the structures he draws, but in fact it abandons the attractive idea that incorporation should fall out naturally from the existing mechanism of Agree alone. That idea is worth pursuing at least somewhat further, so let us consider the alternative option.

The other possibility is that phonetic material is inserted late, though if this is the case, then the exact details of Spellout become fuzzy. Several questions arise. First, how does the grammar know which features correspond to the incorporee, and which to the host? This information will be necessary for purposes of linearization, since the incorporee must precede the host. Also, when the goal is not defective, but Match still copies some of its features to the probe, what prevents those features from being spelled out on the probe?

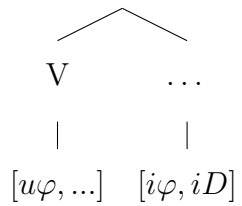
One example of this sort of case is strong object pronouns in Romance, which, unlike object clitics, do not incorporate with the verb. An Italian example:

(13) a. Gianni **li/*loro** stima. (Roberts 2010, 47)
John 3p.clitic/3p.strong esteems
'John esteems them.'

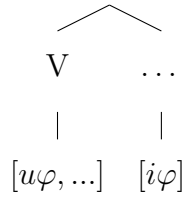
b. Gianni stima ***li/loro**.
John esteems 3p.clitic/3p.strong

According to Roberts (2010), the difference between (13a) and (13b) is due to the presence of D-features on the strong pronoun (*loro*) but not the clitic (*li*). In both cases, presumably, the verb has uninterpretable φ-features and Matches with its object, but only the clitic will incorporate, since its features will be a proper subset of the verb, whereas the verb has no D-features and thus the pronoun must remain in place. That is, Match applies in both (14a) and (14b):

(14) a. *stima + loro:*

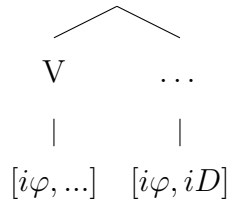


b. *stima + li:*

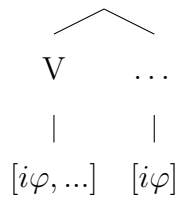


In both cases, the result of Match will be that the object's φ-features are copied into the verb's feature matrix, changing [uφ] to [iφ]:

(15) a. *stima loro:*



b. *li stima:*



So, when the time comes to insert the phonetic material, how does the grammar know that the interpretable φ-features on the verb should be pronounced as a clitic, *li*, in (15b) but not in (15a)? As the proposal stands, it is unclear why (16) should be ungrammatical:

(16) *Gianni **li** stima **loro**.
John 3p.clitic esteems 3p.strong

Thus it is also unclear how head movement can be accounted for with the structure in (12). However, there is a simple solution to the difficulties discussed above, which allows us to retain for the moment the idea that Agree is the only operation involved in incorporation. This involves marking the features with indices to show where they came from. It is already standardly assumed (Chomsky 1995) that lexical items are indexed in the numeration, so that two instances of "run", for instance, could be distinguished as run_i and run_j. Because lexical items are feature sets, it is a minor extension of the theory to say that individual features also have indices, and the index is the same for all features of a lexical item. That is, features are actually ordered triples: $\langle Att, Val, Index \rangle$

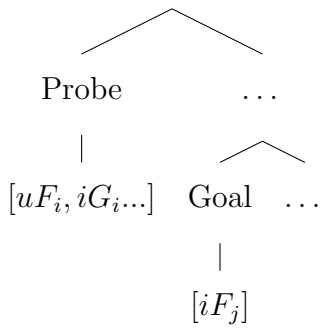
Thus the statement of the Match function needs to be revised. Previously, it stated that when the probe and goal have matching attributes, any undefined value in one will be replaced by a defined value from the other. To this must now be added the additional requirement that when a value is copied from the goal to the probe, its index will be copied as well, replacing the original index for that feature in the probe:

(17) *Match (revised)*: given a well-formed Agree relation of which α and β are the terms (i.e. Probe or Goal) where α's feature matrix contains $\langle A, __, j \rangle$ and β's contains

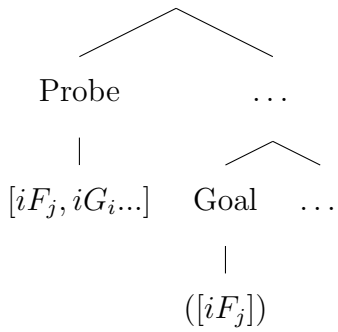
$\langle A, V, k \rangle$, for some feature A, copy V into __ in α's feature matrix, and replace j with k.

With this extension, it becomes possible to keep track of which features have been copied, and where they have been copied from. Thus the structure in (10) should be updated to that in (18), and (12) should be replaced with (19). The notation remains the same as before, with iF and uF representing interpretable and uninterpretable F features, respectively, but with the addition that these features are marked with a subscript index such as i or j:

(18) *Trigger for incorporation:*



(19) *Result of incorporation:*



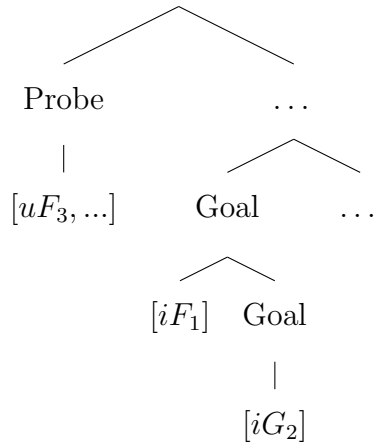
In (19), the probe's feature $[iF_j]$ is marked as having been copied from the goal, which will be necessary information for Spellout. However, it may not yet be sufficient information, as the next section will make clear.

2.2.1 Excorporation and pied-piping

Roberts' account allows for the possibility of excorporation, though only in very restricted circumstances. Specifically, Roberts follows Marantz (2001, 2006) in treating words/heads as phases, and argues that a head may excorporate only when it is on the left edge of its phase, i.e. of the larger derived head. However, if the Agreeing features of the goal are not at the phase edge, then there can be no excorporation. In such cases the entire derived head is pied-piped[3] to incorporate with the probe. Using a head-adjunction structure rather than a flat feature set, it is easy to determine whether the Agreeing features are in the phase edge or not. Thus in (20a), the interpretable F-features of the goal are at the left edge of the derived head, so they can excorporate, producing (20b):

(20) Excorporation:

a. *Trigger*

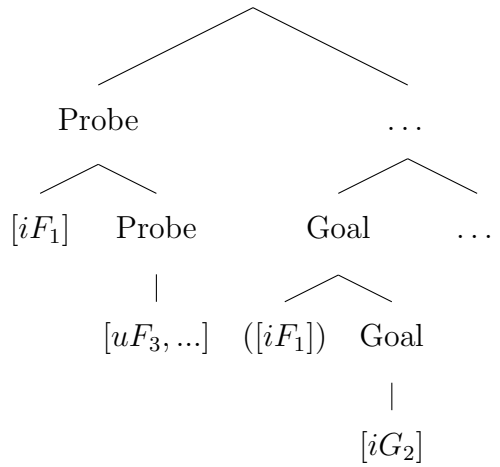


[3]The term *pied-piping*, when applied to phrasal movement, refers to cases like the following, where a wh-word that is embedded in a larger NP or PP moves, carrying the entire NP or PP with it:

(1) *To whom* did you give the money?

Roberts (2010) extends this concept to head movement, using it to describe cases where a head that is embedded inside a larger head complex moves and carries the whole head complex with it.

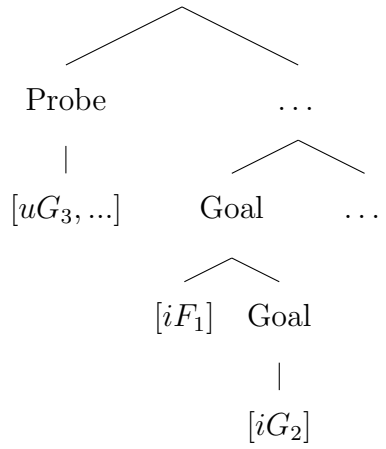
b. *Result*



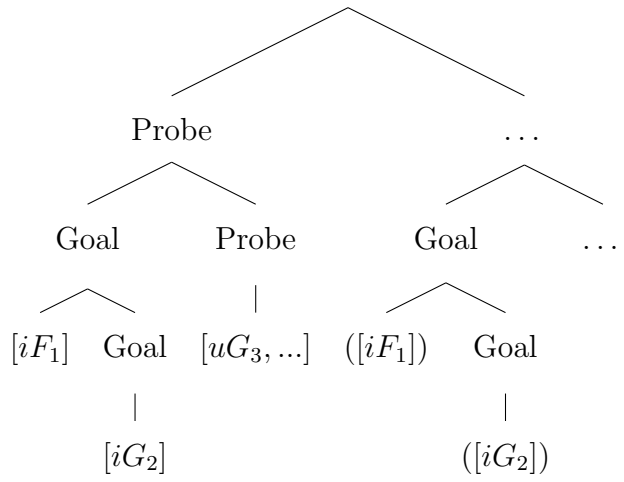
However, in (21a), the Agreeing G-features are not at the left edge of the goal, so excorporation is not possible, and the entire goal must be pied-piped, producing (21b):

(21) Pied-piping:

a. *Trigger*



b. *Result*



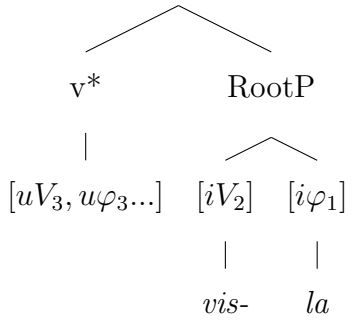
The distinction between excorporation and pied-piping is straightforward if we assume a head adjunction structure like that in (20) and (21), but it runs into a challenge if instead we assume that the only mechanism of head movement is Agree. This is because Agree will simply copy feature values and indices into an existing feature matrix, which is not inherently structured. That is, it is not immediately clear how to define a phase edge in terms of a flat feature set, and without the notion of an edge, we cannot determine when to excorporate, as opposed to pied-piping.

2.2.2 Romance clitic climbing

In order to illustrate some of the difficulties that arise when we try to use a flat feature set here, it may be useful to go through a concrete example in some detail. One phenomenon that Roberts analyzes as involving both pied-piping and excorporation is Romance clitic climbing in periphrastic tenses. In Standard Italian *L'ho vista*, 'I have seen her', the object pronoun *la*, which would cliticize to the main verb *vista* in a non-periphrastic tense, attaches to an auxiliary, *ho*, instead. The step-by-step derivation of this example is as follows, starting

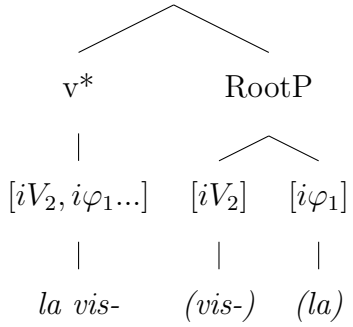
with the smallest substructure where Agreement must occur:[4]

(22)



Both the verb root[5] *vis-* and the clitic *la* are probed by v*, so Match applies and copies their values and indices into v*'s feature matrix[6]:

(23)



Next, the external argument (EA) and a higher v head (called Part, because it defines the root as a verbal participle) are merged:

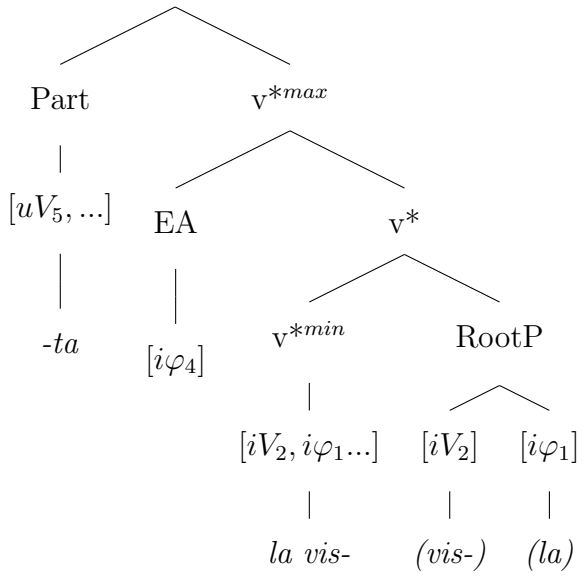
---

[4]I have included the phonetic material in the following trees for purposes of exposition, but we must assume as before that in fact this material will not be inserted until Spellout.

[5]In fact, Roberts (2010) follows Marantz (2001, 2006) in assuming that the root is acategorial, and thus is only marked as a verb after incorporation with little v, which has interpretable V-features. However, it is not clear to me precisely what features the root is supposed to have prior to incorporation. Since it incorporates into little v and all incorporation is feature driven, it must have some interpretable feature that Matches an uninterpretable feature of little v. In the interest of being fully explicit, I have represented this mystery feature as a V-feature in the examples given here. See §3.2.2, where it becomes important that v and not V is marked as iV.

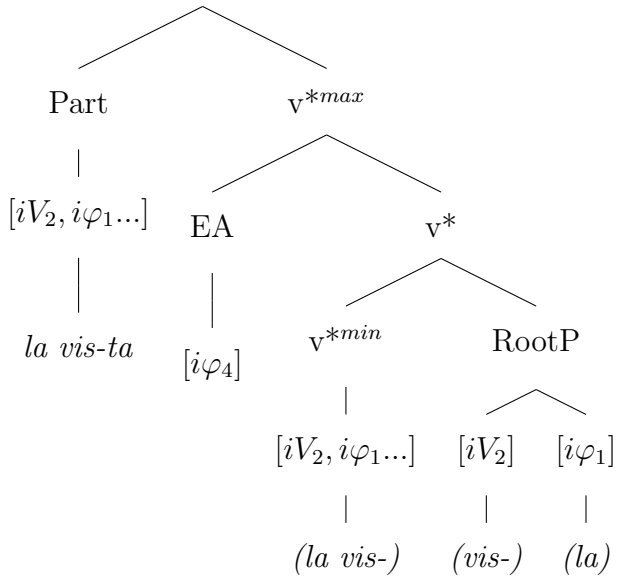
[6]According to Roberts, if two goals are equidistant from the probe, then the "less prominent" goal moves first. (See appendix for definition of *prominence*.) In this case the verb root is less prominent, so this accounts for the observed order *la* > *vis-*

(24)



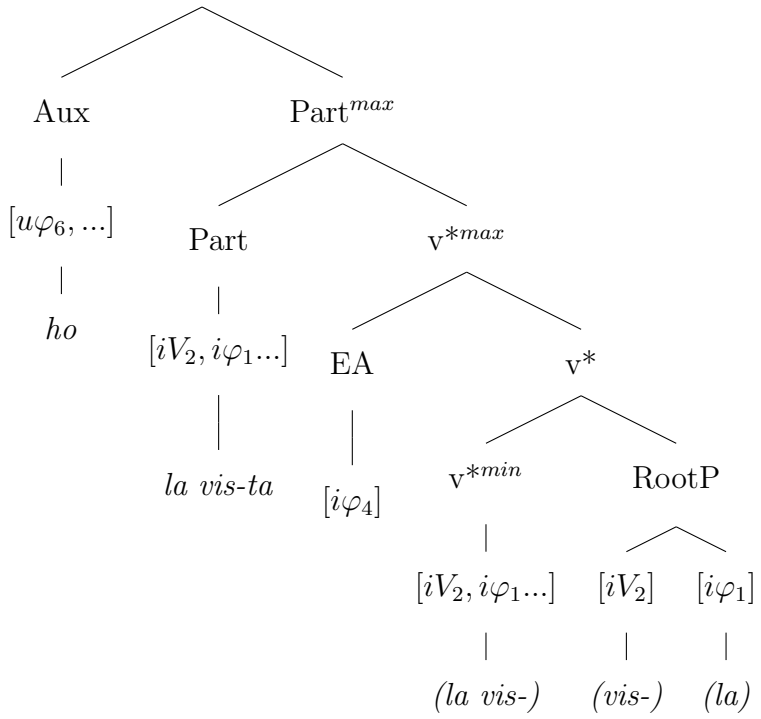
Notice that Part has V-features, but no φ-features. Thus when it probes v*, it can only Match $[iV_2]$, that is, the verb root *vis-*. However, *vis-* is not at the left edge of its phase (though we have not yet defined what exactly this means), so the entire derived v^{*min} head, consisting of $[iV_2]$ and $[i\varphi_1]$, must be pied-piped to incorporate with Part:

(25)



Then the auxiliary is merged. This too is a v head (labeled Aux for convenience), but unlike Part, it has φ-features, not V-features:

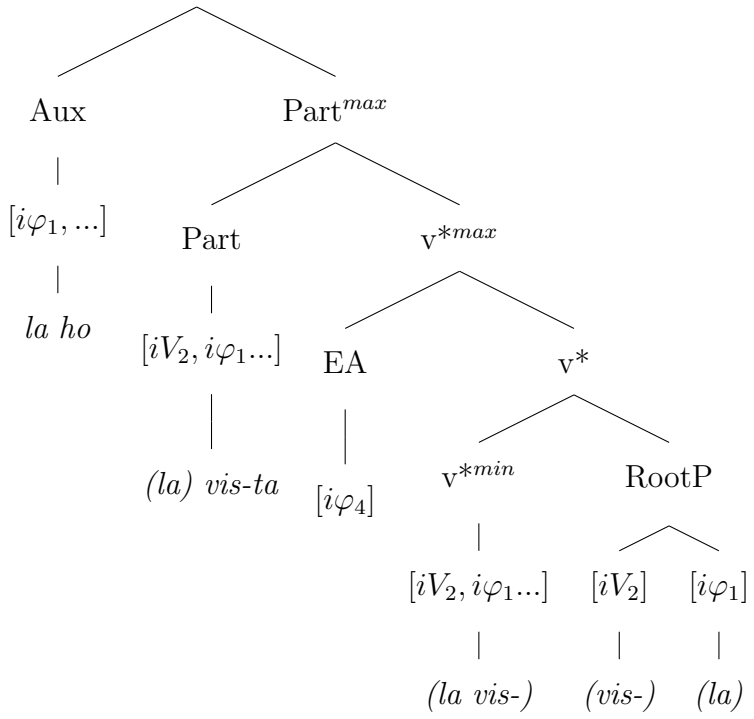
(26)



Aux then probes Part and Agrees with its φ-features[7]. However, in this case, instead of moving the whole derived Part head, the clitic *la* is able to excorporate and move alone to Aux because it is on the left of edge of its phase:

[7]Although the EA also has φ-features, the clitic inside Part counts as a closer goal at this point in the derivation. The EA will later be probed by T, and move to Spec,TP due to its EPP feature. (Roberts 2010, 78)

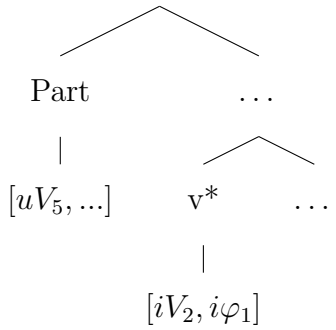
(27)



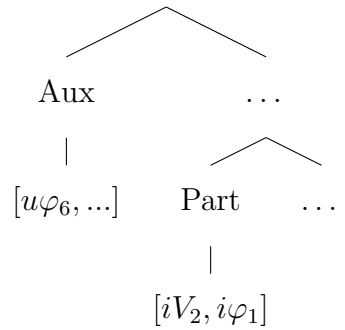
2.2.3 Defining the phase edge

The trouble with the derivation just presented is that both excorporation and pied-piping occur, but the structural configurations that lead to each appear identical. Lacking any further stipulation, there is no obvious reason that $[iV_2]$ should not be able to excorporate in (28), below, or that $[i\varphi_1]$ should not be able to pied-pipe $[iV_2]$ with it in (29):

(28)



(29)



This is not to say that we could not define the notion of phase edge on a structure of this sort. Rather, the point here is only that something extra must be said in order to do so. Specifically, what would be required is information about the order of operations: whichever set of identically-indexed features has been copied to the node most recently can be understood to be on the phase edge, just as in Roberts' head adjunction schema, where all heads are adjoined to the left, meaning that the most recently adjoined will necessarily be the leftmost.

Even remaining with the flat feature set account, where order of adjunction is not represented schematically, the required information is recoverable from the tree. Because we know that less prominent categories move first and that all copies are indexed, it is possible to determine where the copies came from and the relative prominence of their original positions. From there, we can figure out which incorporated most recently.

Similar calculations would be necessary in any case to determine the linear order of the head at Spellout. There, the category that has incorporated most recently must be linearized as the leftmost. Thus, in principle, there is no reason we could not find a way to structurally distinguish cases that should trigger excorporation from those that trigger pied-piping, though I will return to this point and discuss some further challenges in §2.2.5.

2.2.4 Losing the subset criterion

There is a larger problem here than just the necessity of defining the notion of phase edge. Roberts' central claim is that incorporation happens precisely when the goal is defective, in the sense that its formal features are a proper subset of the probe's. However, if we continue with a strict feature-set-based account of incorporation, then this subset relation will not hold in cases of excorporation and pied-piping.

To first take the case of excorporation: in (29), the probe is Aux, whose features include uninterpretable φ-features (presumably among others), but crucially no V-features. The goal is Part, which has both φ- and V-features. There is no subset relation between these feature sets, so Part is not defective with respect to Aux. The clitic *la*, of course, is defective relative to Aux, since it consists solely of φ-features. However, these features cannot constitute a goal independent of Part's other features. Thus there is no way we can say that an Agree relation holds between Aux and the clitic, though intuitively it seems that the Agreement should be between Aux and the clitic rather than Aux and Part.

The Match function, as presently defined, does not in fact require the goal to be defective. This is as it should be, since in many cases we want Agree to be able to apply without triggering movement, as with the strong pronouns in Romance, discussed in §2.2. The case of excorporation then is simply where Agree results in movement even though the goal is not actually defective. This contradicts the initial claim that head movement occurs only when the goal is defective, but it is still in line with the idea that what incorporates is a subset of the features of the probe. The only difference is that what incorporates does not constitute the entire goal.

Pied-piping poses a greater challenge. Just as with excorporation, the trigger for this sort of movement is not a defective goal. However, unlike excorporation, where at least what incorporates into the probe is a subset of the probe's features, pied-piping involves movement of the entire goal, including any features it has that are not also in the probe. A substantial

revision of the Match function will be required to capture these facts:

(30) *Match (second revision)*: given a well-formed Agree relation of which α and β are the terms, where α c-commands β, α's feature matrix contains $\langle A, __, j\rangle$ and β's contains $\langle A, V, k\rangle$, for some feature A:

i) If β is at the left edge of its phase, copy V into __ in α's feature matrix and replace j with k.

ii) If β is not at the left edge of its phase, copy V into __ in α's feature matrix, replace j with k, and copy all other features of the head containing β into α's feature matrix.

Part (ii) of this reformulation is tantamount to an admission that some mechanism in addition to Agree is required in order to account for head movement. There are two possible theoretical moves here: first, we could abandon the notion that everything should fall out naturally from Agree and propose some other kind of movement operation. Alternatively, we could retain the complicated version of Match given in (30). Initially, it might seem that these options are merely notational variants of each other, neither one quite as simple as Roberts' original insight, but neither obviously superior to the other. However there are concrete reasons to be skeptical of a Match function like that proposed in (30), as we will see in the next section.

2.2.5 Phase impenetrability and locality

The Match function must make use of a tree search algorithm in order to determine the order of incorporation. This is necessary for purposes of linearization at Spellout, as well as for determining whether a particular feature bundle is on the phase edge and thus whether its Agreement with the probe should trigger pied-piping of the rest of the head. Informally, for a derived head α, the algorithm will work by taking each index that appears in α and searching down through the rest of the tree for other instances of that index. If all lower

instances of an index $i \in \alpha$ are c-commanded by some instance of an index $j \in \alpha$, then the phonetic realization of the features indexed j will linearly precede that of those indexed i, and (assuming no $k \in \alpha$ that c-commands all instances of j) j will count as being on the left edge of the phase.

This search algorithm is non-local. It cannot simply consider e.g. the next head down, since what it really needs to do is compare to each other all lower heads that share indices with the derived head in question. In some cases, it may be necessary to search into a lower phase in order to determine c-command relations among the copies there, and this is problematic because it is standardly assumed that everything below a phase head becomes inaccessible as soon as the phase head's maximal projection is complete, so it should not be possible to search there.

This difficulty disappears if every head is spelled out/linearized as soon as it is complete (as would in any case be predicted if heads are phases) since Agree is subject to the PIC. But then this account becomes no different from one where incorporation is head adjunction triggered by Agree, and all relevant structure is built as you go along, with no necessity to search the tree (after the initial search by Agree itself). Thus I abandon any version of the flat feature set analysis at this point, and switch to using the head adjunction structures that are suggested by Roberts' diagrams of head movement structures, and by most of the large existing literature on head movement.

3 Head movement in minimalist grammars

Although some of the mechanisms that Roberts proposes as the trigger for head movement appear suspect, the basic idea that head movement should be treated as similarly as possible to phrasal movement is worth pursuing. In this section I propose a different system which closely follows the minimalist grammar framework of Stabler (2001, 2010), while striving to stay as close to the spirit of Roberts' proposal as possible.

3.1 Minimalist grammars (Stabler 2001, 2010)

First, a little background on the framework will be necessary. For Stabler (2001,2010), a lexical item consists of semantic and phonetic features, which are represented either by the conventional spelling of the word or by the empty string ϵ, separated from a sequence of syntactic features by a double colon:

(31) Phon :: feature1 feature2 ... featureN (Stabler 2010, 3)

These syntactic features are ordered: at any time, only the first one in the sequence will be visible to outside operations, but such operations can delete the first feature. Thus later features become available as the derivation proceeds. The syntactic features are of four types. First, there are the features involved in selection:

(32) Selectional features

a. *Categorial features*: `N, V, A, P, D, C, T,` etc.

b. *Selector features*: `=f`, for any categorial feature `f`

For instance, an expression with the selector feature `=N` as its first feature will select for one with the categorial feature `N`. Then there are the features involved in (phrasal) movement, which are separate from the selectional features:

(33) Movement features

a. *Licensor features*: `+wh, +focus, +case,` etc.

b. *Licensee features*: `-wh, -focus, -case,` etc.

An expression with first feature `+wh` will license movement of one with the licensee feature `-wh`.

A minimalist grammar (MG), is just a finite list of such lexical items. For example, the following lexicon is an MG which can produce the sentence "Who praises Marie?":

(34) `Marie::D`

`who::D -wh`

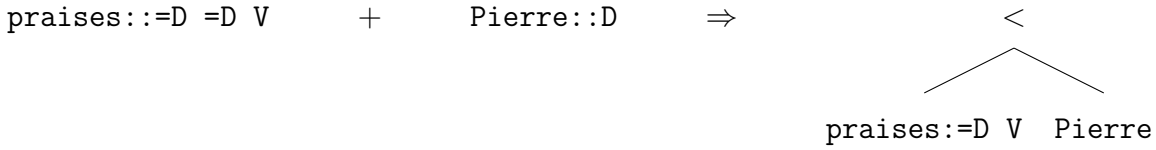
`praises::=D =D V`

ϵ`::=V T`

ϵ`::=T +wh C` (Stabler 2010, 4)

Two operations may apply to these lexical items: Merge and Move. Merge is triggered by the selectional features, and combines two trees into one (where lexical items are assumed to be one-node trees). There are two cases of Merge: first, when a lexical item selects another element, that element is attached to the right of the selector as its sister, and is called a complement, as in (35). If, instead of a lexical item, a derived expression is the selector, then the selected expression is attached on the left, and is called the specifier, as in (36). In either case, both the relevant selector feature `=f` and the categorial feature `f` will be deleted (i.e., checked) as a result of Merge.

(35) *A lexical item selects another expression:*



(36) *A derived expression selects another expression:*

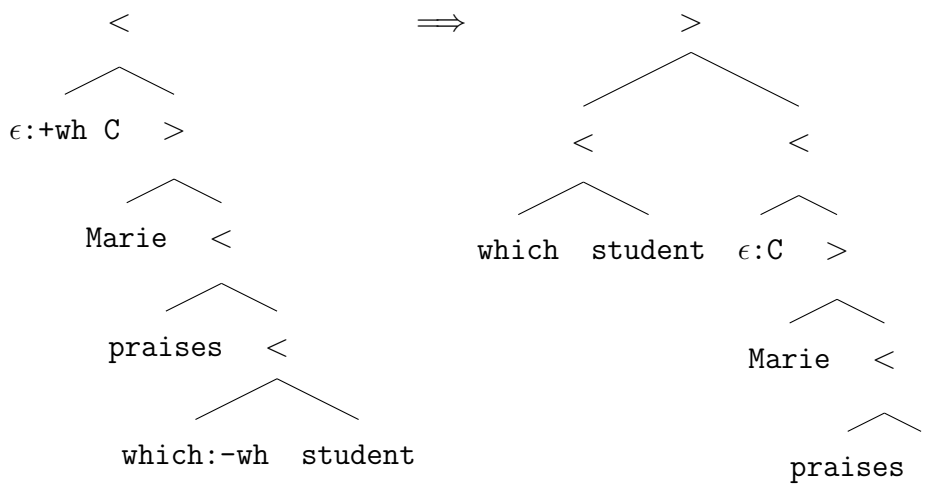
The order symbols $<$ and $>$ point towards the head of the subtrees they dominate. When two expressions are merged, the selector will always be (or contain) the head, so the head is *praises* in both (35) and (36).

Move operates on a single tree whose head's first syntactic feature is a licensor feature (e.g. `+x`). It searches through the tree for a head with a matching licensee feature (`-x`), and moves the maximal projection of that head into the specifier of the the licensor, leaving behind an empty subtree ϵ, and deleting both the licensor and licensee features.

(37) A *maximal projection* is a subtree that is "not properly included in any larger subtree that has the same head". (Stabler 2010, 3)

Thus in (38), the licensee feature `-wh` is on *which*, but it is the entire maximal projection, *which student*, that moves:

(38)



3.1.1 The shortest move constraint

Move is subject to the shortest move constraint (SMC), which requires that for each licensor, there can be only one potential mover. Otherwise the derivation will crash:

(39) SMC: If a tree has `+x` as the first feature of its head, then exactly one head in the tree has `-x` as its first feature. (Stabler 2010, 5)

This constraint is quite restrictive, and rules out several types of analyses that are common in the syntactic literature, where more than one potential mover may exist. For instance, Chomsky (1995, 184-5) suggests that either of two targets of movement may move, as long as they are equidistant from the probe, meaning that they share the same minimal domain. This idea is widespread, and has been used to explain phenomena as diverse as Chichewa V-incorporation (Chomsky 1995), Scandinavian object shift (Wu 2008), and DP-raising in Greek ditransitives (Anagnostopoulou 2003).

I make no claim about the overall validity of these analyses, but I do note that from a computational perspective, it is highly desirable to have some finite bound on the number of movers that the derivation needs to keep track of, and simply replacing the SMC with equidistance removes any such bound. Nevertheless, there are certain variations on the SMC which are compatible with the idea of equidistance, since equidistance only has to do with

the closeness of various movers to the probe, and not to the actual number of movers. For example, we could impose the SMC2 instead of the SMC:

(40) SMC2: If a tree has `+x` as the first feature of its head, then at least one and at most two heads in the tree have `-x` as their first feature.

This seems likely to capture the majority of analyses that rely on equidistance, but if necessary, it would also be possible to have an SMC3, which allows at most three movers, or an SMC4, which allows at most four, etc. There is a downside to such proposals, in that they require counting, but as long as there is some finite bound on the number of movers, the system remains computationally tractable. Another possibility is to assume that there is no bound on the number of potentially moving elements, so the system is intractable in the worst case (Salvati 2011), but ordinary, fluent language use only confronts simple parsing problems. For present purposes, however, I will assume that the version of the SMC given in (39) still holds.

3.2 Adding head movement

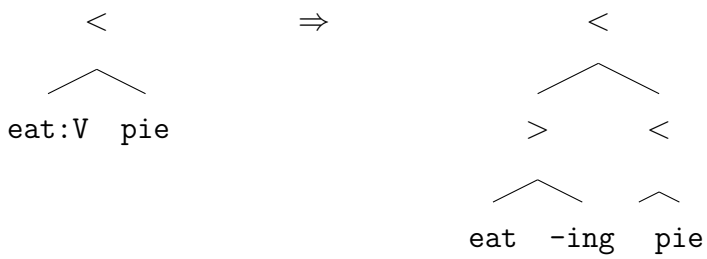
The MGs introduced above so far only allow for phrasal movement, not head movement. There is more than one way that head movement could be incorporated into the system. Stabler (2001), for example, does it by modifying the selectional features. The categorial and selector features introduced in (32) are retained, but two new feature types are added:

(41) a. *Right incorporators*: `f<=`, for any categorial feature `f`

b. *Left incorporators*: `=>f`, for any categorial feature `f`

The only difference between these incorporation features and the selector features (`=f`) introduced above is in where Merge places the expressions they select for. Specifically, a `f<=` feature will select for a subtree with `f` as its first feature, and it will right-adjoin the head

of that tree to its own head. A `=>f` feature will do the same, except the head of the selected tree will be left-adjoined instead of right-adjoined:

(42) `-ing::=>V` + 

Thus head movement is closely tied to selection in this system, and since selection is strictly local, one consequence is that head movement will also be strictly local, as predicted by the Head Movement Constraint. However, as noted in §1.2, there is evidence that the HMC is too strict, and must be relaxed at least enough to allow for long head movement, which has been argued to occur in Slavic, Old Romance, Breton, and other languages. Furthermore, the kind of excorporation effects that Roberts (2010) proposes are impossible, given the version of head movement in Stabler (2001). Therefore, I propose a different mechanism for head movement, retaining the framework of Stabler (2001, 2010), but allowing derivations of the sort given in Roberts (2010).

Instead of adding a new type of selectional features, as in Stabler (2001), I propose a new type of movement features, which will differ from phrasal movement features only in terms of what moves (a minimal rather than maximal projection), and the landing site for the movement (left-adjoined to the licensing head). These features will be notated as follows:

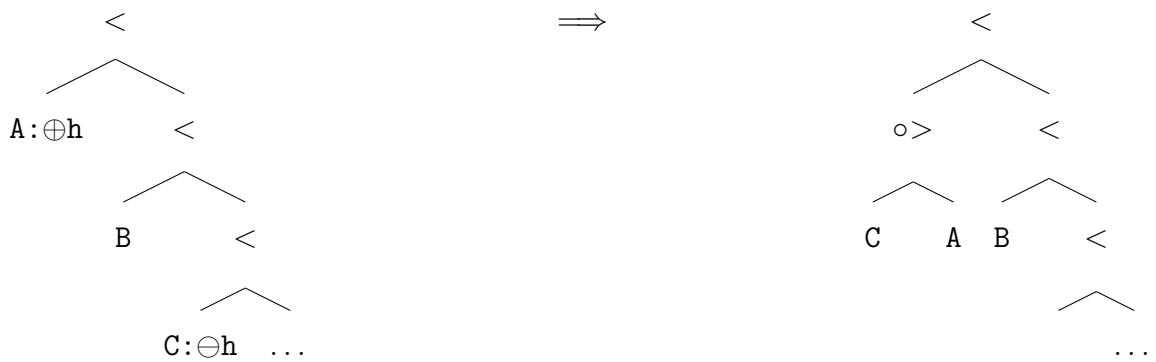
(43) a. *Head movement licensor features*: $\oplus$`f`

b. *Head movement licensee features*: $\ominus$`f`

Because we will be using this system to build complex heads and we still want to be able to distinguish minimal and maximal projections, it will be useful to also have notation indicating headedness within a head complex. Just as $<$ and $>$ point towards the head of a phrase, $< \circ$ and $\circ >$ will point towards the head within a larger head complex. That is, if a head α moves to incorporate with a head β, β will be considered the head of the newly created head complex $[_\beta\ \alpha\ \beta]$. It will become clear why we need such a notion of head-internal headedness in §3.2.2.

As with phrasal movement, other heads can intervene between the licensor and licensee of head movement, as long as these interveners are not also marked as licensees. This opens the door for long head movement, while still imposing some restrictions; head movement is phase-bounded, and only the closest licensee can move. Thus in (44), `A`, `B`, and `C` are all heads, but because `B` has no $\ominus$`h` feature, `C` is able to move past it:

(44)



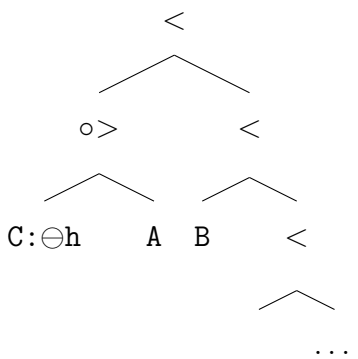
So far so good, but if we want to allow Roberts-style excorporation as well, a few complications must be added to the very simple idea sketched above:

3.2.1 Persistent features

First, because excorporation in this sense involves multiple movements of the same head, triggered by the same features of that head, those features cannot simply be deleted as part

of the Move function. If they were, then there could be no further movement. Therefore, persistent features will be required. However, it is only the head movement licensee feature (⊖`x`) that needs to persist; the head movement licensor (⊕`x`) can and should delete as usual, since excorporation does not require the same head to trigger multiple instances of the same kind of movement. Therefore the outcome of head movement in (44) should instead be as follows, where `C` retains its ⊖`h` feature, making it available for further head movement:

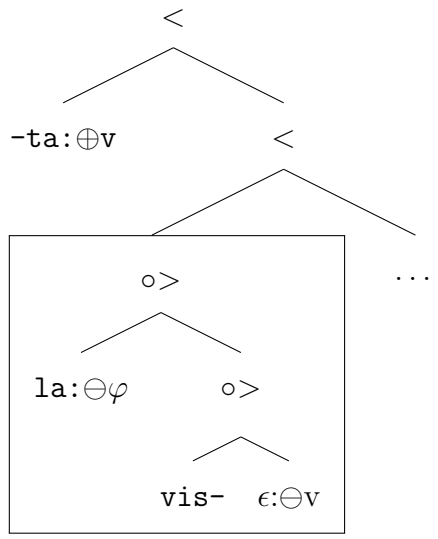
(45)



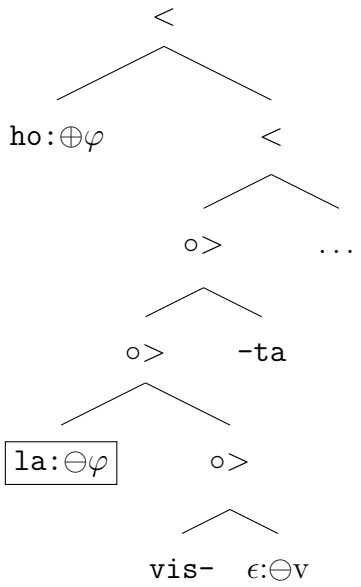
3.2.2 Defining the mover

A second complication lies in distinguishing between excorporation cases, where only the leftmost element of a complex head moves, and pied-piping cases, where the entire head moves. Returning to the Romance clitic climbing example from §2.2.2 (*L'ho vista* 'I saw her'), and updating it to the current formalism, the structural configurations that result in excorporation and pied-piping are as follows:

(46) a. *Trigger for pied-piping (of 'la vis-'):*



b. *Trigger for excorporation (of 'la'):*



I have already said that the thing that moves in head movement is a minimal projection, but I have not yet defined that notion precisely. Traditionally, "minimal projection" has been more or less synonymous with "head", but since we are dealing with heads within heads, this definition must be complicated a bit:

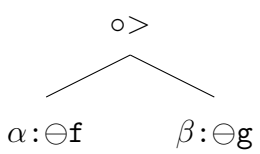
(47) a. The *minimal projection* of a subtree t is the maximal head complex that is the head of t.

b. A *maximal head complex* is a head complex that is not properly included in any larger head complex that has the same head.

Note that the term *head* is ambiguous between the meanings "head of a phrase" and "head within a head complex". In (47a), if the subtree t is a phrase, then the head of t is the head of a phrase, and if t is a head complex, then its head is the head within a head complex.

Given the definitions in (47), something close to Roberts' excorporation story falls out quite naturally. To see why, consider the following:

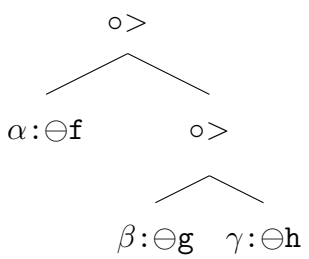
(48)



If an expression with first feature $\oplus\texttt{f}$ probes (48), it will license movement of the maximal head complex that has $\ominus\texttt{f}$ as its first feature. Since α is marked as $\ominus\texttt{f}$, it will move, but since it is not the head of the larger head complex $[_\beta\ \alpha\ \beta]$, it cannot carry β with it. Thus α will excorporate, exactly as Roberts suggests for this case. By contrast, a $\oplus\texttt{g}$ probe will license movement of the maximal head complex of the head with $\ominus\texttt{g}$ as its first feature, that is, the maximal head complex of β. Because β is properly included in a larger head complex $[_\beta\ \alpha\ \beta]$ which has the same head, it must be the entire head complex that moves. Thus movement of β will carry α with it, deriving Roberts-style pied-piping.

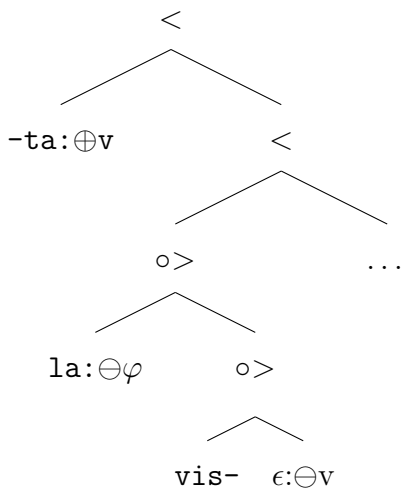
So far, this looks like a fairly elegant way to capture the excorporation/pied-piping distinction. However, it is worth checking a few more complex cases to ensure that it makes the predictions we want. For instance, the case where there are three or more heads in a right-branching structure, as in (49):

(49)



Here, α can excorporate, and movement of γ will pied-pipe the rest of the head with it, as desired. The questionable case is that of β, which can also excorporate, leaving α and γ behind, since it is not the head of any head complex larger than itself. Notice that the position of *vis-* in (46a), repeated below as (50), is structurally equivalent to that of β in (49) (although I have not marked *vis-* as a head movement licensee). This implies that it should be possible to extract *vis-* by itself, leaving the clitic behind, given the necessary feature specification for *vis-* and the probe:

(50)



Recall from footnote 5 in §2.2.2 that it is a bit uncertain what features the verb root, *vis-*, is supposed to have prior to incorporation with v, represented here as ϵ:⊖v. In the current framework, we must assume that since the root moves as a head, it must in fact have some kind of ⊖x feature, which will persist after movement, as do all head movement licensee features. That is, it should probably in fact be represented as vis-:⊖x in (50). So, in theory, it would be possible for there to be a higher ⊕x head that could license its later

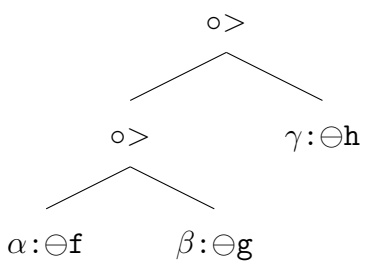
excorporation.

Whether it is desirable for a head like *vis-* to be able to excorporate is essentially an empirical question: are there syntactic phenomena where such a movement is the only (or at least the most elegant) way to derive the facts? This is not an easy question to answer. In this particular case, it seems undesirable, since Italian only allows *L'ho vista*, not **Vis l'ho -ta*, or anything of the sort. But any such individual case can be ruled out simply by saying that there is no lexical item involved in the derivation that has the necessary feature specification to trigger the undesirable excorporation. That is, there is no higher $\oplus$`x` head that can attract `vis-:`$\ominus$`x` in (50). This kind of solution sidesteps the question of whether the basic mechanism of excorporation should be allowed to apply here. I do not claim to have the empirical evidence necessary to decide this, and indeed, neither does Roberts. None of the example derivations in Roberts (2010) require such a movement, but none rule it out either.

Internally to the theory, it makes sense to allow this sort of excorporation because it matches up neatly with familiar ideas about derivation by phase (Chomsky 2001). In a phase like CP or vP, only the left edge of the phase is accessible to outside operations, where *left edge* is defined as the head and any specifiers it may have. Roberts (2010) argues that heads are phases too, so only things on their left edge should be extractable. Because all heads within a head complex are adjoined to the left, and heads never select other heads, the head of a head complex will always be the rightmost leaf. This suggests that anything other than the rightmost leaf is a specifier, and should therefore be able to excorporate, whereas movement of the head, that is, the rightmost leaf, will carry the rest of the head complex with it, deriving pied-piping.

Note that this means that complex specifiers can also excorporate:

(51)



In (51), a higher ⊕g probe will trigger movement of the $[_{\beta}\ \alpha\ \beta]$ head complex. This movement is simultaneously pied-piping (because β carries α with it) and excorporation (because a smaller head complex moves out of a larger one). Again, it is unclear whether there is an empirical need for this kind of movement, but it is worth pointing out that the theory predicts it.

3.2.3 Separation of HM and phrasal movement features

A further question regards the potential interaction of head movement and phrasal movement. Both kinds of movement are triggered by features that reside on heads, so what happens if a single head has both kinds of features? For instance, consider the following MG, where B has both head movement and phrasal movement features:

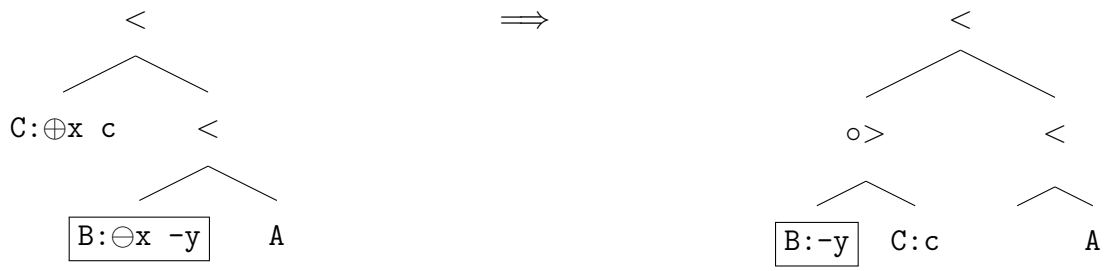
(52)
```
A::a
B::=a b ⊖x -y
C::=b ⊕x c
D::=c +y d
```

In (53a) we see head movement of B to C.[8] Note that B carries a phrasal movement feature (-y) with it, which triggers the movement in (53b). Because B is now head-adjoined to C,

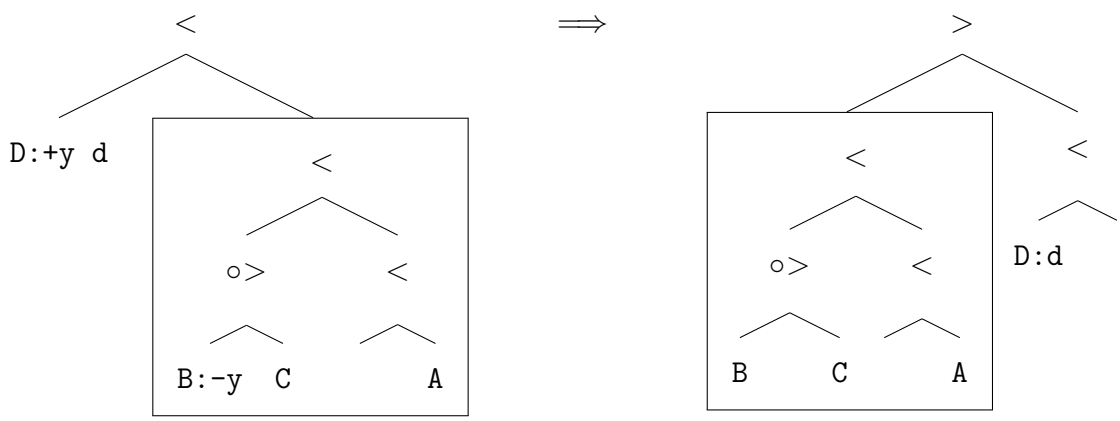
[8]For purposes of exposition, I have left out the persistent HM licensee features in these trees. If they persisted here, then they would be the first feature in the sequence, making any following features, like -y in (53a), inaccessible to later operations. This is problematic, but since I argue that the system already requires modification on other grounds, it is safe to gloss over this detail.

the whole head complex is marked as `-y`. Therefore it is the entire maximal projection of the $[_C$ `B C]` head, including `A`, that must then move into `D`'s specifier:

(53) a. *Head movement...*



b. *...followed by phrasal movement:*



This is not a linguistically plausible derivation. What we have here is a head moving into a higher constituent, and then causing the movement of that constituent, while the maximal projection of its original position remains in place. To the best of my knowledge, there is no attested syntactic phenomenon that would require such a movement. Thus we will want to rule out derivations of this sort.

One way to rule it out would be to stipulate that any single lexical item can only have one of the two types of licensee features. However, as we will see in §3.3, Roberts (2010) proposes an analysis of French clitic clusters that requires a lexical item with both types of features. Specifically, a V head-moves out of VP, leaving behind its object, and later the remnant VP fronts: a phrasal movement that can only have been triggered by a feature on the (now absent) V. In order to accommodate this kind of derivation, we need to be able to

leave some features behind after head movement. Thus I propose that head movement splits the licensee's feature sequence into two sets:

(54) a. The *projecting features* of a head α are those that are left after all $\ominus$`x` and phonetic features are deleted from α.

b. The *head features* of α are the result of deleting all projecting features from α.

What actually moves in head movement is only the head features of the minimal projection of the subtree with $\ominus$`x` as its first feature. The projecting features are left behind at the head's original position, so instead of (53a), we have (55):

(55)

The separation of projecting and head features guarantees that the later phrasal movement will move just the remnant phrase containing `A`, rather than the the maximal projection of the complex $[_C$ `B C`$]$ head.

3.2.4 Head movement and the SMC

The last remaining question is about the status of the shortest move constraint (SMC) with respect to the present proposal about head movement. Recall the definition of the SMC given in (39), repeated here:

(56) SMC: If a tree has `+x` as the first feature of its head, then exactly one head in the tree has `-x` as its first feature. (Stabler 2010, 5)

As written, this condition applies specifically to phrasal movement, but it would be desirable if we could extend it to head movement as well, for the same purpose as discussed in §3.1.1: to

limit computational complexity. For most phenomena that have traditionally been analyzed as involving head movement (e.g. V-to-T movement, T-to-C movement, etc.), there will be no problem with imposing the SMC. In fact, on any theory where head movement is subject to the HMC, it will trivially obey the SMC as well, since the HMC implies that there is only one position from which a head can move and thus that there is only ever one potential mover. However, in an approach like this one, where long head movement is possible, nothing automatically rules out having multiple potential movers.

It would be easy enough to extend the SMC to apply to head movement as well as phrasal movement. The question is whether the SMC is too restrictive to allow the sorts of derivations we want. Roberts (2010) contains at least one proposal for a derivation that does not obey the SMC, involving Romance clitic clusters: head movement of both direct and indirect object clitics is triggered by their φ-features, and at certain points two different clitics compete to move to the same head. As before, it would be possible to weaken the SMC to something like the SMC2, as in (40), where at most two movers are allowed, and this would permit the kind of derivation Roberts proposes for clitic clusters. But before we make that move, it will be worthwhile to consider this example in some detail, both because of its implications with respect to the SMC, and because it raises larger questions about this overall approach to head movement.

3.3 Romance clitic clusters

The status of the SMC in this version of minimalist grammar seems to hang on the proper analysis of object clitic clusters in Romance, as in the following French example:

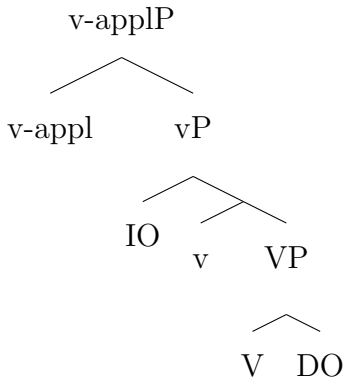
(57) Elle **me** **le** donnera.
3s.nom 1s.dat 3s.acc give-fut
"She will give it to me"

According to Roberts (2010), the clitics *me* and *le* reach their observed positions in (57)

through incorporation with distinct v heads. The incorporation is triggered by the clitics' status as defective goals: they both consist solely of φ-features. Translating this idea into the current formalism, we must say that they both have $\ominus\varphi$ features and the v-heads that they incorporate with have $\oplus\varphi$ features. Because head movement licensee features are persistent, both clitics will be active as potential goals simultaneously, in violation of the SMC. We could accommodate derivations of this sort by imposing something like the SMC2 instead, as discussed in §3.1.1. But I will argue that because the derivation Roberts proposes for these data is controversial and in fact contains at least one internal inconsistency, it does not constitute a sufficient reason to do away with the SMC.

Roberts (2010) assumes the following as the universal underlying structure for ditransitives, where "the applicative v selects a vP whose 'external argument' is the indirect object and whose complement is a standard VP containing a direct object" (136):

(58)

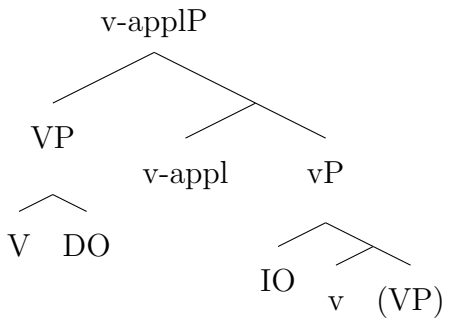


In French (and most other Romance languages), the usual surface realization of this structure places the direct object (*le livre*, in (59)) before the indirect object (*Marie*), which is contained in a PP headed by *à*:

(59) Je donnerai le livre à Marie.
1s.nom give-fut the book to Marie
"I will give the book to Marie."

According to Roberts, this DO > IO order is derived by phrasal movement, fronting the VP:

(60)



When the objects are realized as clitics, the story becomes more complicated, since clitics are subject to the Person-Case Constraint (PCC):

(61) PCC: In a dative construction where both the direct and indirect objects are represented by phonologically reduced or grammaticized elements such as clitics or agreement markers, then:

a. *Strong version:* The direct object has to be third person.

b. *Weak version:* If there is a third person, it has to be the direct object.

(Bonet 1991)

This is suggested by Bonet (1991) to be a linguistic universal, showing effects in many languages. However, Roberts (2010) focuses on its application to French and Italian, and specifically to its effects on the ordering of various clitic combinations in those languages. The French facts are as follows. The order is DO > IO when both objects are third person, but this order is ungrammatical if either is first/second person:

(62) DO > IO

a. Elle le lui donnera.
3sf.nom 3sm.acc 3s.dat give-fut
'She will give it to him.'

b. *Elle le me donnera.
3sf.nom 3sm.acc 1s.dat give-fut
'She will give it to me.'

If the IO is first/second person, it precedes the DO, though this order is ungrammatical when both are third person:

(63) IO > DO

a. Elle me le donnera.
3sf.nom 1s.dat 3sm.acc give-fut
'She will give it to me.'

b. *Elle lui le donnera.
3sf.nom 3s.dat 3sm.acc give-fut
'She will give it to him.'

It is impossible to have a first/second person DO clitic if the IO is a also a clitic, regardless of the order:

(64) a. *Elle me lui présentera.
3sf.nom 1s.acc 3s.dat introduce-fut
'She will introduce me to him.'

b. *Elle lui me présentera.
3sf.nom 3s.dat 1s.acc introduce-fut

Rather, in such cases, the IO must be realized in a PP:

(65) Elle me présentera à lui.
3sf.nom 1s.acc introduce-fut to 3sm.dat
'She will introduce me to him.'

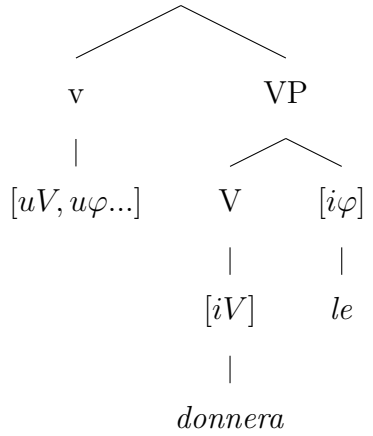
Roberts' proposal to account for these facts takes advantage of the following idea: third person can be considered the unmarked value for the person feature. That is, any instance of $\langle Pers, __ \rangle$ in the derivation can be filled in by default at the interfaces as $\langle Pers, 3 \rangle$, and indeed $\langle Pers, 3 \rangle$ is only ever specified in this way: third-person marking is not included in any lexical item when it first enters the derivation, whereas first and second person marking is. One might wonder then, what happens when a Probe with uninterpretable φ-features Agrees with a third person Goal (i.e. one with the feature $\langle Pers, __ \rangle$). According to Roberts,

vacuous Agree is possible in such a situation: the Goal and the Probe both have a Matching $\langle Pers, __ \rangle$ feature, so Agree is satisfied. The next question is whether vacuous Agree can trigger incorporation[9], and Roberts argues that it can, but the incorporation is optional in such cases. Thus, to return to the empirical facts under discussion, a first or second person clitic will be forced to incorporate, whereas a third person clitic will not be, though it optionally can. The trouble is in defining the circumstances under which a third person clitic will incorporate.

3.3.1 me le

Recall that when the indirect object is first person, the order of clitics is IO > DO, as in *Elle me le donnera*, which is derived as follows:

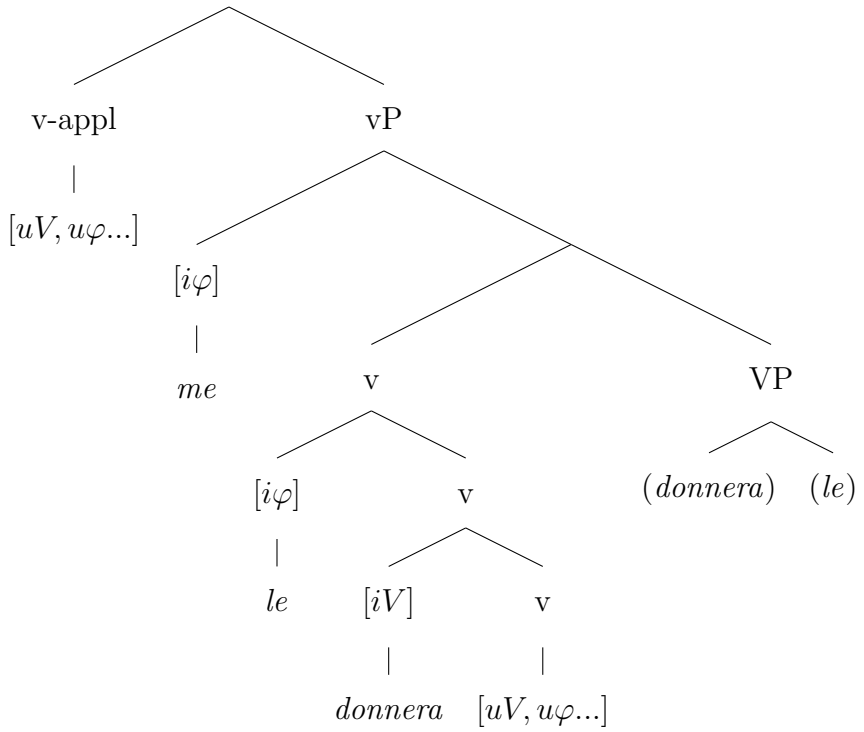
(66)



In (66), v Agrees with both V and the clitic *le*. V incorporates first, because it is less prominent. The clitic is third person, and therefore has the option of incorporating or not. In this case, it does incorporate, though why exactly it does is an open question. After that, the indirect object and an applicative v head are Merged:

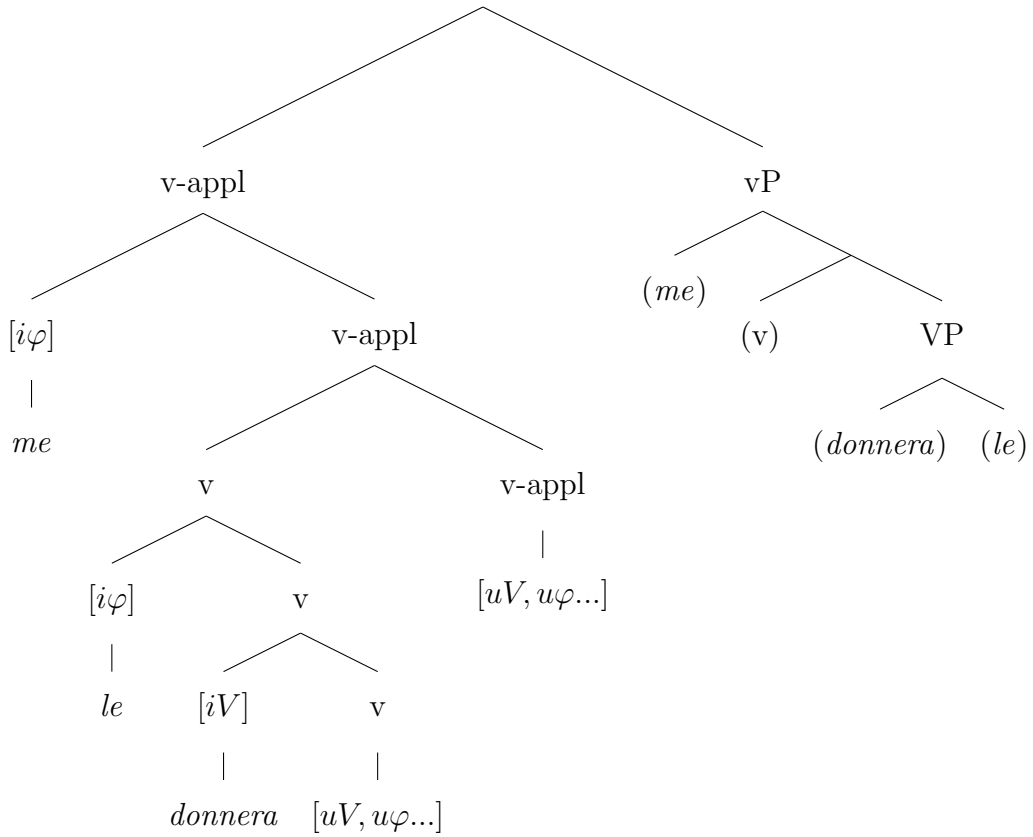
[9]Note that this question would be incoherent on the flat feature set account of these facts, since on that analysis Agree and incorporation are one and the same thing. There is no separate operation that can be triggered.

(67)



At this point, v-appl's V-features probe the V inside the derived v head, which then incorporates, pied-piping the entire v head with it because *donnera* is not on the phase edge. Then v-appl's φ-features probe the indirect object *me*, which is forced to incorporate because it is first-person and therefore its Agreement with v-appl is not vacuous. The result is (68):

(68)



This straightforwardly derives the order *me le* – that is, IO > DO – and though the derivation is clearly not complete, the head $[_{v-appl}$ IO $[_v$ DO V]] will travel as a unit for all further verb movement.

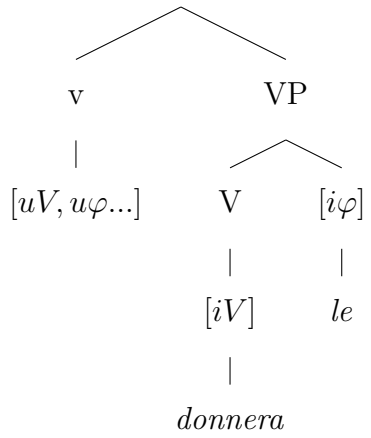
3.3.2 le lui

When both the objects are third person the observed order is DO > IO:

(69) Elle le lui donnera.

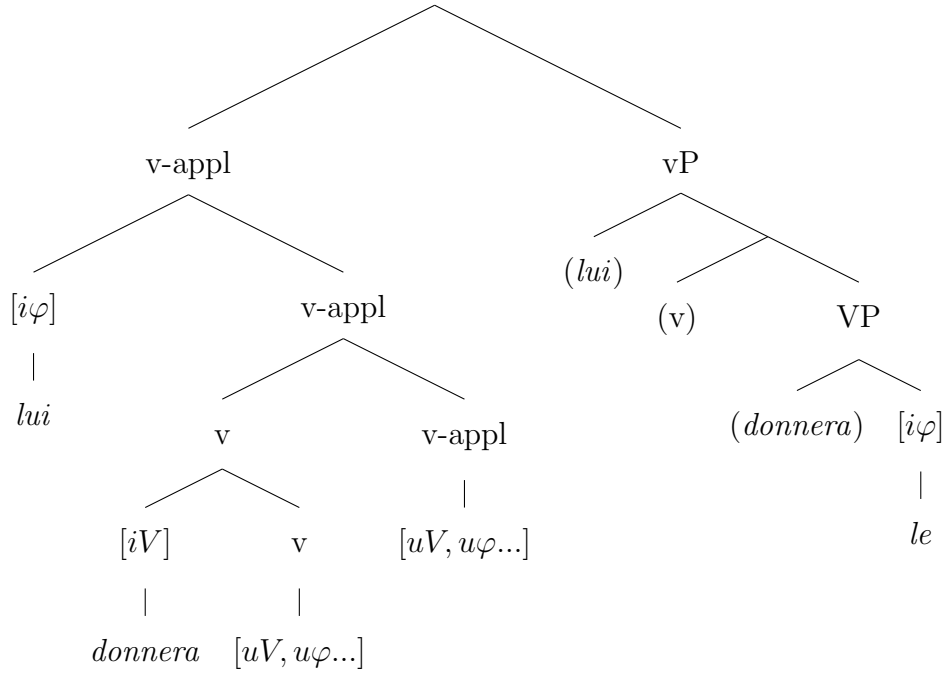
However, the derivation for (69) starts out the same way as above:

(70)



(70) is identical to (66) above, and just as before, v Agrees with both V and *le*, vacuously in the case of the clitic. But in this case, the vacuous Agreement does not trigger incorporation. Thus *le* remains in the VP while *donnera* incorporates with v, and then the indirect object and applicative v are Merged. As before, the v-appl head incorporates the complex v head and the indirect object, resulting in this structure:

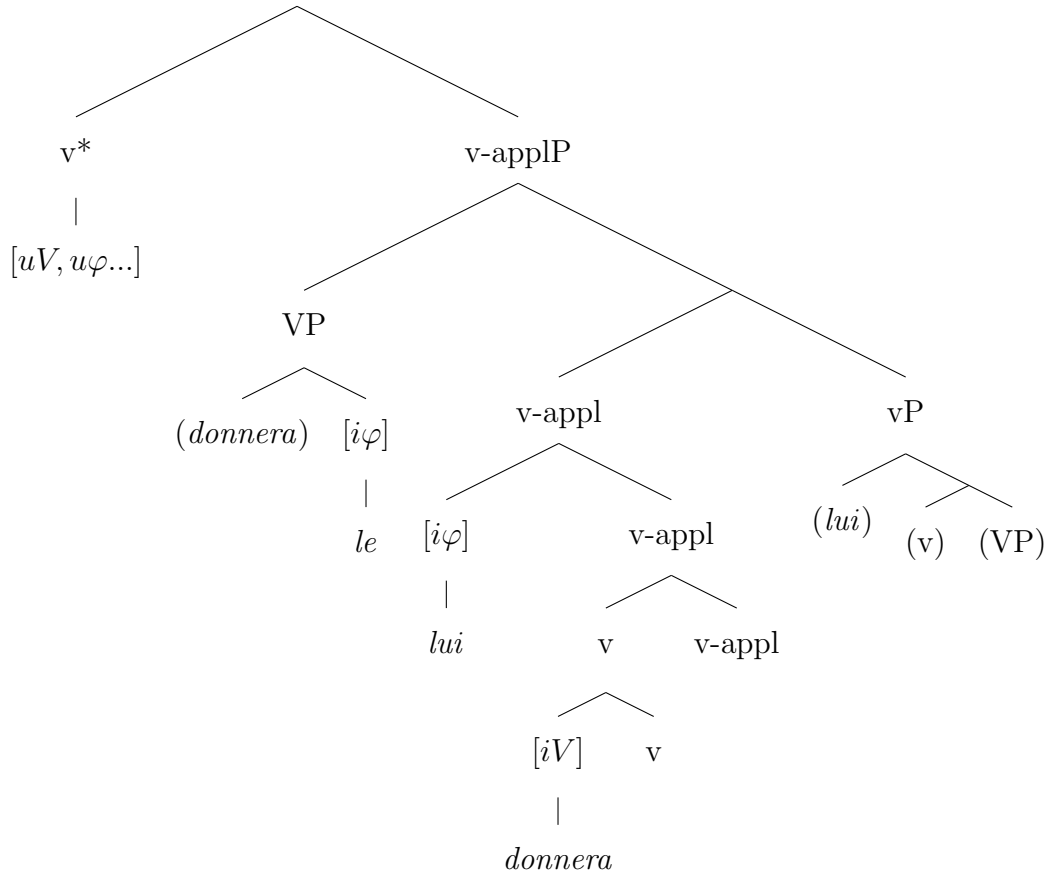
(71)



At this point, the VP – whose only current overt element is the DO clitic *le* – fronts, just as it would in a sentence involving full DP objects, such as *Je donnerai le livre à Marie.* This is phrasal movement triggered by an EPP feature on v-appl, not incorporation, and so

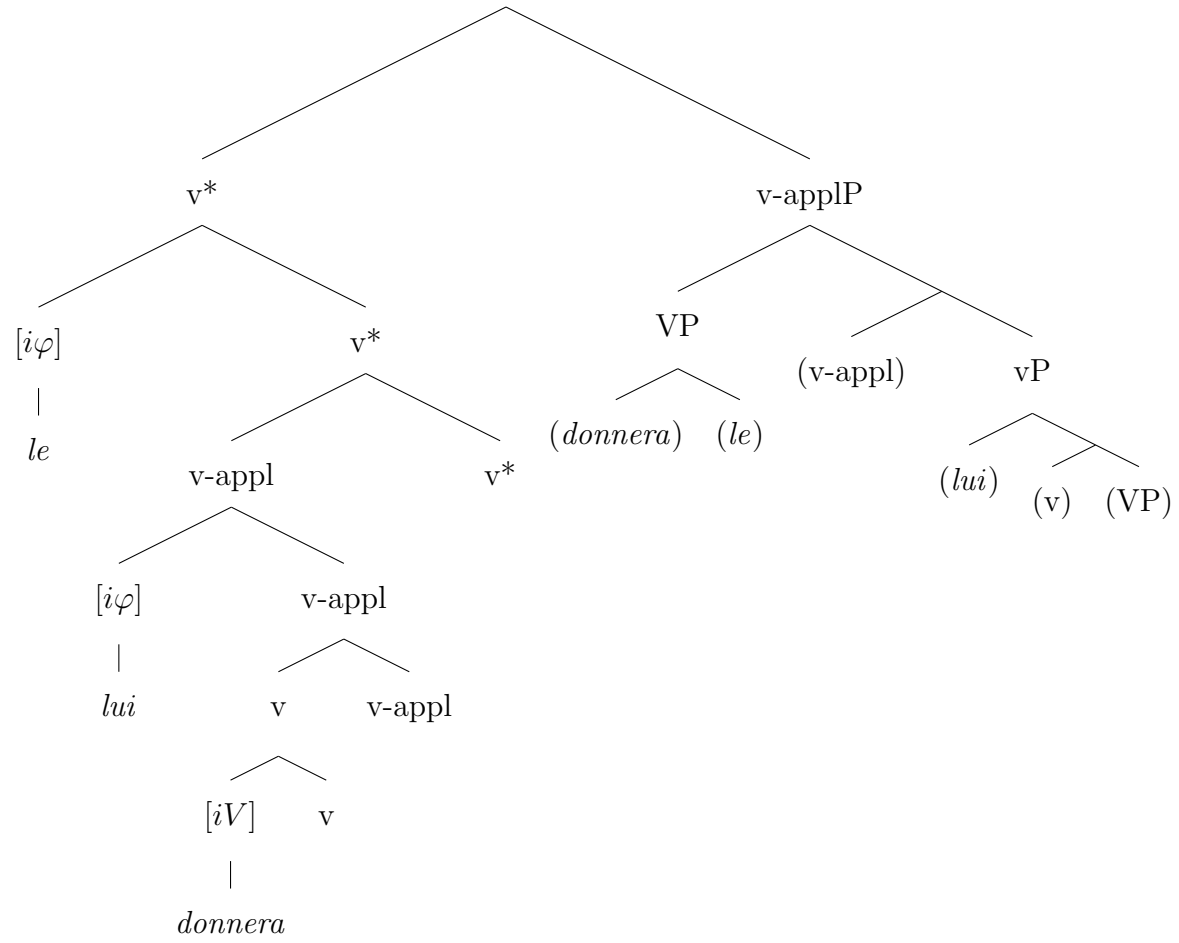
the landing site is Spec,v-applP. Afterwards, a v* head is Merged, and it, like v-appl and v, has uninterpretable V and φ features:

(72)



Next, v* probes the verb's V-features, triggering incorporation of the entire derived v-appl head, including the indirect object *lui*. However, *lui* does not satisfy v*'s φ-features. Thus v* also probes the direct object *le* in its position in the fronted VP, and it incorporates from there:

(73)



3.3.3 ?lui le

In the event that the direct and indirect object are both unmarked for person, the system leaves open two possibilities for the ordering of the clitics. In §3.3.2 we saw that the DO > IO order *le lui* is derived if vacuous Agree fails to trigger incorporation of the direct object to v, thus leaving *le* behind in the VP so that it may later incorporate with v* instead. But it is not clear exactly what prevents the DO from incorporating here, since it does incorporate when the IO is first/second person, as seen with *me le* in §3.3.1. In fact, at the relevant point in the derivation, a structure that triggers incorporation of a third person DO (such as (66)) may be identical to one that does not trigger incorporation (such as (70)). This predicts that (74a) should be just as grammatical as (74b), but this is not the case:

(74) a. *Elle lui le donnera

b. Elle le lui donnera.

Roberts in fact intends to leave both ordering possibilities open in order to account for the fact that the Romance languages differ on this point. French has DO > IO here, but other languages, such as Italian, have IO > DO: *glielo*. Thus, when Roberts says that incorporation is optional, he does not mean that speakers of a particular language vary in which one they choose; rather, he means that there is parametric variation between different languages, so French chooses nonincorporation where Italian chooses incorporation.

But French does not choose nonincorporation across the board in all cases where Agreement is vacuous. It only chooses nonincorporation of a third person DO in the specific case where the IO is also unmarked for person. Indeed, even in that case, the same DO clitic does incorporate later in the derivation, only its incorporation is with the higher head v* rather than with v. Roberts' account does not adequately explain the apparently variable triggering of incorporation following vacuous Agreement.

A benefit of my proposal from §3.2 is that it is capable of explicitly stating when to move the DO clitic and when not to. There is no optionality in the application of Move: it applies in every instance where the formal features involved in the derivation allow it to do so. Since all such features are lexically specified, all that is necessary in order to specify when exactly to move the DO clitic is to propose two different lexical entries for the null v head:[10]

(75) a. ϵ::=V $\oplus$x $\oplus\varphi$ v

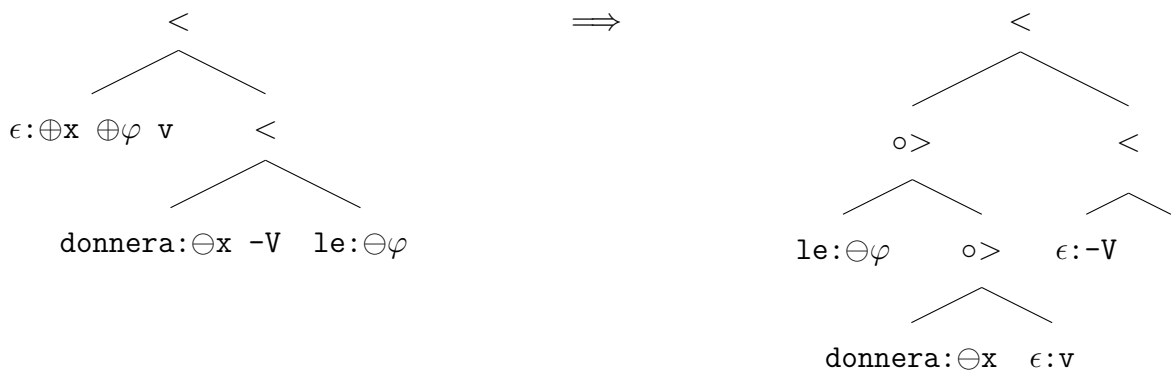
b. ϵ::=V $\oplus$x v

Both lexical entries in (75) specify that this particular null head selects for a V, triggers head movement of a $\ominus$x head, and can itself be selected as a v. The only difference is that (75a)

[10]The lexical entries given here are slightly simplified, for clarity. They should in fact also include a =D feature, which will allow v to select the IO as its specifier.

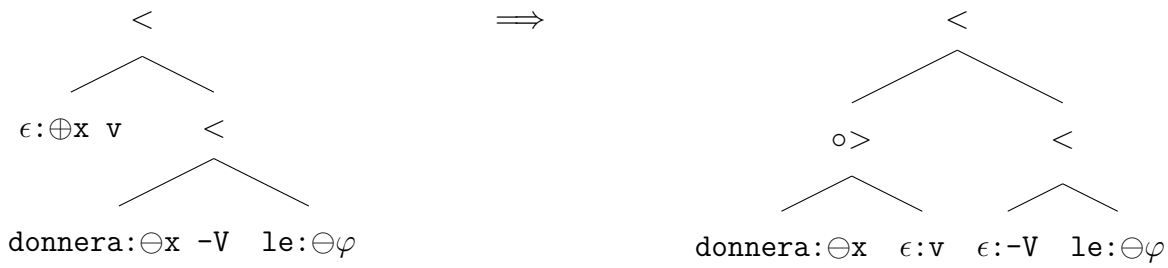
additionally trigggers movement of a $\ominus\varphi$ head (i.e. a clitic), whereas (75b) does not. Thus if (75a) selects for a VP [*donnera le*], both *donnera* and *le* will move as heads to adjoin to it, as in (76):

(76)



If instead we use the lexical item in (75b), then the clitic *le* will be left behind in the VP to be attracted by a higher head later in the derivation, as in (77):

(77)



Thus this system can provide explicit triggers for all movements involved in the derivations that Roberts proposes. However, although it can generate the PCC facts seen above, it still falls short of explaining them, since there is nothing that actually determines which version of the v head should be merged. (75a) must be used when the IO that will be merged later is first or second person, and (75b) must be used when the IO is third person, but there is nothing in the analysis so far that enforces this. Therefore it is doubtful whether either version of this analysis is the best possible characterization of the PCC facts in French. Incidentally, Matushansky (2011, 544) reviews some of Roberts' proposed analyses for other phenomena – including Slavic second position clitics and clitic doubling in Montesover Trentino – and finds that they are similarly problematic.

However, it is unclear to what extent we are forced to an analysis like the above by the formal mechanisms of the minimalist grammars presented in §3.2. The exact details of any derivation will be determined by the feature specifications of the lexical items involved, and the order in which they are merged. Thus the actual head movement operation presented in this paper is in principle compatible with a variety of different analyses for any particular syntactic phenomenon.

I leave it to others to determine how best to account for the PCC. A brief survey of the current literature on the subject reveals little consensus. One open question regards the universality of the PCC: it was originally claimed (Bonet 1991) that either a strong or weak version of the PCC holds universally of phonologically reduced dative and accusative elements, meaning that the constraint applies to phenomena much broader than just Romance clitic clusters. However, there appear to be exceptions in Romanian (Ciucivara 2009) and many other languages (Anagnostopoulou 2003, Haspelmath 2004). There are also questions about whether the person-case constraint actually makes distinctions on the basis of person, or if instead the relevant factor is animacy (Ormazábal and Romero 2007), or an even more general factor: topicality (Haspelmath 2004). Among those who argue that the PCC is in fact based on person, there are still differences of opinion. Early proposals tended towards morphological solutions involving templates for clitic order (Perlmutter 1971, Bonet 1991). More recent accounts are more likely to be syntactic: Anagnostopoulou (2003) and Béjar and Rezac (2003) propose variants on an analysis that is somewhat reminiscent of Roberts' proposal, involving split checking of person and number features, while Ciucivara (2009) has a cartographic approach where within the functional sequence of the clause, there is a hierarchy consisting of a Person-field followed by dative and accusative Case heads, and clitics must check their case and person features by moving to the appropriate heads, in sequence.

In short, there is no one standard analysis of the PCC, and so although Roberts (2010) may not have the right answer here, it is not yet clear that anyone else does either. Thus it

would be premature to draw too many firm conclusions about the import of the Romance clitic cluster facts on the larger question about the status of head movement in minimalist syntax. There is no *a priori* reason that head movement of the style proposed in this paper could not be compatible with any of the stories about the PCC just discussed. Therefore I do not claim that the Romance clitic cluster data offers any conclusive evidence either that head movement is a necessary part of syntactic theory, or that it is incompatible with mainstream assumptions about the narrow syntax. Nevertheless, I believe that this example illustrates several points that are relevant to an overall assessment of my proposed system of head movement. But before turning to that assessment, a summary of the proposal is in order:

3.4 Interim summary

Head movement is triggered by specific head movement features, which are disjoint from the phrasal movement features. When Move encounters a head movement licensor feature $\oplus$x, it searches down through the tree for a matching head movement licensee feature $\ominus$x, and moves the head features of the minimal projection of the $\ominus$x head, adjoining them to the left of the $\oplus$x head. Any projecting features the moved head may have had (i.e. those that trigger phrasal movement) are left behind at its original position. In the process of head movement, the $\oplus$x feature is deleted, but the $\ominus$x feature persists in order to allow for further movement of that head, via excorporation or pied-piping. Because we have seen no convincing evidence to the contrary, I assume for present purposes that head movement is in fact subject to the SMC, thus ensuring that derivations involving head movement remain computationally tractable.

4 Assessment

The version of head movement summarized in §3.4 is successful in several respects. It is a relatively simple and internally coherent treatment of syntactic head movement that is largely in line with standard minimalist assumptions about syntax. It provides an answer to the longstanding questions of why a head moves rather than an XP, and why that head moves to adjoin to another head, rather than to a specifier position: head movement and phrasal movement are triggered by disjoint sets of features, and the operation Move simply treats each set of features slightly differently. Also, if we assume that head movement is indeed subject to the SMC, then these MGs with head movement have some nice computational properties: they should be provably equivalent to the MGs without head movement discussed in Stabler (2010), in the sense that both kinds of MGs define the same set of languages (though I leave the actual proof of this to future work). This also suggests that they can be shown to be efficiently parsable using standard methods, such as Earley and CKY parsing, which is a desirable result (although the validity of these methods for modeling human parsing remains an open question (Stabler 2010, 8)).

Still, some aspects of the system may seem unsatisfyingly stipulative. Recall that in Roberts (2010), in order for a goal to incorporate with its probe, its formal features must be a proper subset of the probe's. Matushanksy's (2011) review of Roberts (2010) argues that this requirement often forces Roberts to arbitrarily enrich the features of the probe or impoverish the features of the goal in order to produce a subset relation, in ways that are not always motivated by any independent empirical considerations. Although the proposal summarized in §3.4 does not rely on the defectivity of the goal, and thus avoids some of the difficulties of Roberts' approach, it nevertheless does not constrain in any principled way the kinds of feature specifications that a lexical item may have. This leaves plenty of room for arbitrariness in the lexicon: the analyst is still free to propose that a lexical item has whatever features are necessary to make the derivation work out the way they want,

whether those features are empirically motivated or not. The fact that it is possible to make the derivation work out properly in §3.3 by adding or removing head movement features from the initial feature specification of the lexical items involved is troubling, and may only be a symptom of a larger problem.

The core of my proposal is that lexical items enter the derivation already specified for whether they will undergo or trigger head movement. This is a necessary step in order to differentiate between head and phrasal movement, but it can introduce some unfortunate arbitrariness into the derivation. For instance, let us revisit example (2), repeated below:

(78) a. Harvey **should** buy milk.

b. **Should** Harvey __ buy milk?

As the theory currently stands, we must assume that the auxiliary *should* has a different feature specification in (78a) than it does in (78b). In both cases it selects a `V` as complement and is itself selected as a `T`, but because it has moved in (78b), it must additionally have a head movement licensee feature ⊖`x` there:[11]

(79) a. `should::=V T` In (78a)

b. `should::=V T` ⊖`x` In (78b)

Because the feature sets are different, these actually count as two different lexical items. The version of *should* in (79a) must be used when there is no higher head with a ⊕`x` feature,

[11] One possible alternative explanation might be to say that actually, in both cases the lexical item is `should::=V T` ⊖`x`, and the only difference between (78a) and (78b) is the presence or absence of a higher ⊕`x` head. This would take advantage of the existing idea that head movement licensee features are persistent, which implies that there is already some process to delete head movement licensee features once their role in the derivation is complete, so that the derivation will not crash at Spellout. Thus any lexical item that can ever be head-moved would enter the derivation marked with a head movement licensee feature that could later be deleted by the same process, whether the item actually moved as a head in the course of the derivation or not. However, this solution would not work for a case like Roberts' (2010) analysis of French clitic clusters, since there the difficulty is in finding a principled way to choose between the version of a lexical item with or without a licensor feature (not with or without a licensee feature, as in (78)). If possible, we would like to avoid proposing that licensor features are also persistent and thus able to be deleted by any process other than movement.

and the version in (79b) must be used when there is such a higher head. The trouble is that at the point in the derivation where *should* is merged, there is no available information in the tree about what other lexical items will be merged higher up, and thus no basis for the decision between (79a) and (79b).

A standard response to this sort of problem is to say that the derivation may proceed up to Spellout regardless of whether all features are checked (Chomsky 1995). It is only at the LF interface that any lingering uninterpretable syntactic features will cause the derivation to crash, so Spellout acts as a kind of post-hoc filter on derivations. From the perspective of processing or acquisition, this is not entirely satisfying, but note that in this respect, head movement is no different from phrasal movement. Any syntactic theory that relies so heavily on feature checking will essentially force much of the ugliness and arbitrariness of the grammar into the lexicon. It is possible that the lexicon is the proper place for such things (though it would require psycholinguistic experimentation to say for sure), but until the field develops a serious theory of the syntactic feature system, such lexical solutions will continue to feel stipulative. However, this is not a problem specific to head movement; rather, it applies to all minimalist syntax. Therefore, it does not constitute an argument against including narrow syntactic head movement in minimalist theory.

So far, I do not think we have seen any convincing argument that head movement cannot work at all, though certainly it may not be the right answer in every situation where it has been proposed. For instance, Koopman & Szabolcsi (2000) argue that a head movement analysis of complex verb formation in Hungarian is untenable on a number of grounds: head movement can explain only a some of a variety of phenomena that seem clearly related, while the rest must involve XP-movement; it would be redundant to propose both kinds of movement; and allowing head movement at all makes incorrect predictions in some cases. They propose instead that the cases that appear to involve head movement can be better treated as remnant XP-movement. However, this does not count as an argument against

the concept of head movement as a whole – only its application to the Hungarian and West Germanic phenomena that Koopman & Szabolcsi discuss. It remains unclear to what extent remnant movement could replace head movement analyses more broadly.

Furthermore, remnant movement poses theoretical challenges of its own, since in principle it is possible to derive any order of elements using remnant movement, which is constrained primarily by the specific functional projections that host the movement. Because these projections are often not phonetically realized, it requires very careful analysis to motivate their existence, just as all the real work in proposing a head movement analysis under the system in §3.4 comes in motivating the feature sets for each lexical item in the derivation. So again, head movement has not been shown to be clearly inferior to its existing alternatives.

5 Conclusion

I make only one firm claim in this paper: narrow syntactic head movement is potentially compatible with standard minimalist theory. This does not mean that head movement is a necessary component of minimalist syntax, or that it is necessarily involved in the derivation of every phenomenon that has been claimed in the literature to require head movement. But the fact that so many such analyses continue to be proposed attests to the continued usefulness of the concept of head movement, and the arguments against its inclusion in narrow syntax so far seem inadequate to justify its removal. Thus I have proposed a new version of head movement that is internally coherent and (most likely) computationally equivalent to other varieties of minimalist grammars that have been better studied. Its mechanisms are simple, yet powerful – possibly too powerful, since in principle they could be used to build structures that might not be attested in the languages of the world. However, the formal features of the lexical items manipulated by the mechanisms of the grammar can always rule out such structures. That is, the structure fails to occur because the lexical items that would

be required to build it are simply not part of the lexicon. This claim may sound stipulative, but it is a kind of stipulation that is not specific to head movement. Rather, it is shared with current minimalist syntactic theory in general, and that is why I argue that head movement is compatible with that theory.

6 References

Anagnostopoulou, E. 2003. *The Syntax of Ditransitives: Evidence from Clitics.* Berlin, New York: Mouton de Gruyter.

Baker, M. 1988. *Incorporation.* Chicago: University of Chicago Press.

Bejar, S. and M. Rezac. 2003. Person Licensing and the Derivation of PCC Effects. In A.T. Pérez Leroux and Y. Roberse, eds., *Romance Linguistics: Theory and Acquisition: Selected papers from the 32nd Linguistic Symposium on Romance Languages, Toronto, April 2002.* John Benjamins.

Bonet, E. 1991. *Morphology after Syntax: Pronominal Clitics in Romance Languages.* Doctoral dissertation: MIT.

Bonet, E. 2007. The Person-Case constraint and repair strategies. In R. d'Allesandro, S. Fischer, G.H. Hrafnbjargson, eds., *Person restrictions.* Mouton de Gruyter.

Borsley, R., M. Rivero, and J. Stephens. 1996. Long Head Movement in Breton. In I. Roberts and R. Borsley, eds., *The Syntax of Celtic Languages*, 53-74. Cambridge: Cambridge University Press.

Chomsky, N. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press

Chomsky, N. 2001. Derivation by phase. In M. Kenstowicz, ed., *Ken Hale: A Life in Language*, 1-52. Cambridge, MA: MIT Press

Ciucivara, O. 2009. *A syntactic Analysis of Pronominal Clitic Clusters in Romance: The view from Romanian.* Doctoral dissertation: NYU.

Halle, M., and A. Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from Building 20: Essays on linguistics in honor of Sylvain Bromberger*, Kenneth Hale and Samuel Jay Keyser, eds. 111-176. Cambridge, Mass.: MIT Press.

Haspelmath, M. 2004. Explaining the ditransitive person-role constraint: A usage-based approach. *Constructions 2/2004.*

Heim, I., and A. Kratzer. 1998. *Semantics in generative grammar.* Oxford: Blackwell.

Higginbotham, J. and R. May. 1981. Questions, quantifiers, and crossing. *Linguistic Review*, 1: 41-80.

Koopman, H., and D. Sportiche. 1986. A note on long extraction in Vata and the ECP. *Natural Language and Linguistic Theory* 4: 357-374.

Koopman, H., and A. Szabolcsi. 2000. *Verbal Complexes.* Cambridge, MA: MIT Press

Marantz, A. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, and A. Williams, eds,. *Proceedings of the 21st Annual Penn Linguistics Colloquium*, University of Pennsylvania Working Papers in Linguistics.

Marantz, A. 2001. Words. Talk given at WCCFL, University of Southern California.

Marantz, A. 2006. *Phases and words.* MS., New York University.

Matushansky, O. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37: 69-110.

Matushansky, O. 2011. Review article: Ian Roberts' Agreement and head movement: Clitics, incorporation, and defective goals. *Journal of Linguistics* 47/2: 538-545.

McCloskey, J. 1996. On the scope of verb-movement in Irish. *Natural Language and Linguistic Theory* 14: 47-104.

Nunes, J. 2004. *Linearization of Chains and Sideward Movement.* Cambridge, MA: MIT Press

Ormazábal, J. and J. Romero. 2007. The Object Agreement Constraint. *Natural Language and Linguistic Theory* 25.

Perlmutter, D.M. 1971. *Deep and Surface Structure Constraints in Syntax.* New York: Rinehart and Winston, Inc.

Rivero, M.L. 1991. Long head movement and negation: Serbo-Croation vs. Slovak and Czech. *The Linguistic Review* 8: 319-351.

Roberts, I. 2010. *Agreement and Head Movement: Clitics, Incorporation, and Defective Goals*. Cambridge, MA: MIT Press

Salvati, S. 2011. *Minimalist grammars in the light of logic.* INRIA Research Report No. 00563807.

Stabler, E.P. 2001. Recognizing head movement. In P.De Groote, G. Morrill, C. Retoré, eds, *Linguistic Form and its Computation*, 245-260. Springer.

Stabler, E.P. 2010. Computational perspectives on minimalism. In C. Boeckx, ed. *Oxford Handbook of Linguistic Minimalism*, 616-641.

Travis, L. 1984. *Parameters and Effects of Word Order Variation.* Doctoral dissertation: MIT.

Wu, H.I. 2008. *Generalized Inversion and the Theory of Agree.* Doctoral dissertation: MIT.