

3 Derivation and Representation in Modern Transformational Syntax

HOWARD LASNIK

0 Introduction

The general issue of derivational vs. representational approaches to syntax has received considerable attention throughout the history of generative grammar. Internal to the major derivational approach, transformational grammar, a related issue arises: are well-formedness conditions imposed specifically at the particular levels of representation¹ made available in the theory, or are they imposed “internal” to the derivation leading to those levels? This second question will be the topic of this chapter. Like the first question, it is a subtle one, perhaps even more subtle than the first, but since Chomsky (1973), there has been increasing investigation of it, and important argument and evidence have been brought to bear. I will be examining a range of arguments, some old and some new, concerning (i) locality constraints on movement, and (ii) the property forcing (overt) movement, especially as this issue arises within the “minimalist” program of Chomsky (1995b).

1 Locality of Movement: Subjacency and the ECP

The apparent contradiction between unbounded movement as in (1) and strict locality as in (2) has long played a central role in syntactic investigations in general and studies of the role of derivation in particular:

- (1) Who do you think Mary said John likes
- (2) ?*Who did you ask whether Mary knows why John likes

Based on the unacceptability of such examples as (2), Chomsky (1973), rejecting earlier views (including the highly influential Ross 1967a) proposed that

long distance movement is *never* possible. (1) must then be the result of a series of short movements, short movements that are, for some reason, not available in the derivation of (2). In roughly the terms of Chomsky (1973), movement across more than one bounding node is prohibited by Subjacency and S is a bounding node. Consider (3), a representation of (1), with the successive positions of *Who* indicated by *t* (trace):

(3) [_{S'} Who [_S you think [_{S'} *t* [_S Mary said [_{S'} *t* [_S John likes *t*]]]]]]]

Movement is via "Comp," given the phrase structure:

(4) S' → Comp S

In (2), on the other hand, the lower Comps are occupied, as shown in (5):

(5) [_{S'} Who [_S you ask [_{S'} whether [_S Mary knows [_{S'} why [_S John likes *t*]]]]]]]

Therefore Subjacency is necessarily violated: the movement of *Who* must have been in one long step, crossing three Ss on the way.

Immediately, a question arises about long movement out of certain non-"islands":

(6) Who do you think that Mary said that John likes

(6) has the grammaticality status of (1), but does not seem relevantly different from (2) in its structure:

(7) [_{S'} Who [_S you think [_{S'} that [_S Mary said [_{S'} that [_S John likes *t*]]]]]]]

Chomsky's solution to this problem was a strictly derivational account of locality effects (though he did not make a big fuss about it at the time). The proposal was that in the course of the syntactic derivation, the structure of (6) is identical to that of (1), namely (3). Movement proceeds via successive Comps in both derivations. The difference is that in (6), at a very late level in the derivation the non-matrix Comps are spelled out as *that*. Thus, at that late level, Subjacency appears to be violated, even though in the course of the derivation every step is legitimate. This is the paradigmatic type of situation Chomsky frequently alludes to in his recent writings when he argues for a derivational approach to syntax.

In the early 1980s, particularly under the impetus of Huang (1981/2, 1982), attention began to focus on certain locality "asymmetries." For instance, while movement of an argument across two bounding nodes results in substantial degradation, movement of an adjunct yields total unacceptability:

(8) ?Which problem do you wonder whether Mary solved

(9) *How do you wonder whether Mary solved the problem

Huang, extending ideas from Chomsky (1981), appealed to the Empty Category Principle (ECP), a constraint independent of Subjacency, demanding locality between a (non-argument) trace and its antecedent. (9) is then worse than (8) because (8) violates only Subjacency while (9) violates both Subjacency and the ECP. Now notice that by the above reasoning, the ECP too must be satisfiable “online,” or (10) would incorrectly be assigned the status of an ECP violation:

(10) How do you think that Mary solved the problem

Lasnik and Saito (1984), developing Huang’s ideas further, explored the ECP requirements of intermediate movement positions. Consider the unacceptable long movement in (11):

(11) *How do you wonder whether John said (that) Mary solved the problem

In the derivation of (11), movement can proceed via the lower Comp, as in (10). This step is clearly licit, and the adjunct trace in the lower clause must be in satisfaction of the ECP, given the grammaticality of (10). The next step of movement is in violation of Subjacency, but the descriptive problem is that (11) is worse than a mere Subjacency violation, as in the structurally parallel (12):

(12) ??Which problem do you wonder whether John said (that) Mary solved

Following Chomsky’s proposal (adopted by Huang) that the ECP is specifically a requirement on traces, Lasnik and Saito concluded that even intermediate traces must satisfy the requirement (must be “properly governed”). Consider first the structure of the version of (11) without *that*:

(13) How [_S do you wonder [_{S'} whether [_S John said [_{S'} *t'* [_S Mary solved the problem *t*]]]]]

The initial trace *t* is properly governed by the intermediate trace *t'*. But the intermediate trace is too distant from its antecedent *How*, causing a violation of the ECP.

The version of (11) with *that* seems trivial. The *that* in (14) presumably keeps the initial trace from being properly governed:

(14) How [_S do you wonder [_{S'} whether [_S John said [_{S'} that [_S Mary solved the problem *t*]]]]]

However, this simple account must be rejected. If it were correct, then (10), repeated as (15), would be incorrectly ruled out.

(15) How did you think that Mary solved the problem

Recall that earlier, I indicated, based on (15), that the ECP must be satisfiable “online.” That is, the initial trace must be established as properly governed at a point in the derivation where the *that* is not present. That point might be before *that* is inserted, as suggested above. Alternatively, as proposed by Lasnik and Saito, the point might be after *that* is inserted and subsequently deleted in the LF component. Under the assumption that operations in the “covert” component mapping S-structure to LF are just those that are potentially available overtly (in the D-structure to S-structure portion of a derivation), Lasnik and Saito argued that the availability of an overt Complementizer deletion process entails the availability of a covert analogue. Under either alternative, the question is how we can resolve the apparent contradiction between the unacceptable (11) and the acceptable (15).

2 Constraints on Adjunct Movement: Representational and/or Derivational

The analysis provided by Lasnik and Saito has clear implications for the derivation vs. representation issue. They argued that the proper government requirement must be satisfied specifically at *levels* (S-structure or LF), and not at arbitrary points of the derivation. The reasoning was as follows. We have already seen that allowing online satisfaction of the ECP (gamma-marking of trace, in the Lasnik–Saito technology and notation) has the incorrect consequence of allowing derivations such as (13), where an adjunct makes a short licit move followed by a long illicit move. Yet we still must allow (15). Earlier, I noted that Lasnik and Saito argued that *that* can be deleted in the LF component. Simplifying slightly, immediately following the deletion, the representation of (15) is (16):

(16) How do you think [_S [_S Mary solved the problem *t*]]

Then, nothing prevents Move alpha, the basic and general movement operation of the grammar, from “lowering” *How* to the lower Comp, then re-raising it to the matrix Comp, leaving a trace in the intermediate Comp. Both traces are now in the appropriate locality configuration with their nearest antecedents for gamma-marking to operate, and the sentence is correctly allowed.

Chomsky (1986b) offers an alternative “derivational” approach to the locality constraints on adjunct movement. Chomsky rejects the “representational” stipulation that gamma-marking is only at levels. The crucial example to consider is (11), which I argued above is ruled out on the representational approach but incorrectly allowed on the derivational approach. At this point, I slightly alter the structures of the relevant sentences in accord with the extended X-theory introduced by Chomsky (1986b). Chomsky proposes that *S'* is actually CP, the maximal projection of Comp (and *S* is actually IP, the maximal projection of the tense-agreement inflectional element of the sentence). *Wh*-movement is

then not literally to *C*, but rather to Spec of *C*, in accord with a generalized structure preservation constraint. Note in passing that this renders the acceptable (15) straightforwardly unproblematic. On these assumptions, presence or absence of *that* is evidently irrelevant (though Chomsky's Minimality condition on proper government, which I will put aside here, makes it relevant again in certain circumstances). (11) then has (17) as one step in its derivation, and at that point in the derivation, the initial trace can be gamma-marked:

- (17) [_{CP} [_C do [_{IP} you wonder [_{CP} whether [_{IP} John said [_{CP} How [_C (that) [_{IP} Mary solved the problem *t*]]]]]]]]]]]

How then moves to matrix Spec of CP, violating Subjacency. Further, if this movement leaves a trace, that trace will also violate the ECP. But the worry of Lasnik and Saito was that the movement would not obligatorily leave a trace, since no principle demands a trace in this intermediate position. Or, if a trace is left, no principle prevents its deletion. Chomsky (in lectures in the late 1980s) addressed this worry with the following principle:

- (18) Adjuncts must be fully represented.

The intent of (18) is that every step of adjunct movement must leave a trace, and none of these traces can be deleted. Thus, the continuation of (17) will necessarily be (19), and *t'* will necessarily be in violation of the ECP, an "offending trace" in Lasnik and Saito's term:

- (19) [_{CP} How [_C do [_{IP} you wonder [_{CP} whether [_{IP} John said [_{CP} *t'* [_C (that) [_{IP} Mary solved the problem *t*]]]]]]]]]]]

Note that we also have here the beginnings of an explanation of the adjunct-argument asymmetry in long movement. For argument movement, the variable trace must of course be present, but nothing (including (18)) demands that intermediate traces be present. Parallel to (19), we thus can have (20) with the structure (21), where the offending trace *t'* is deleted:

- (20) ??Which problem do you wonder whether John said (that) Mary solved

- (21) [_{CP} Which problem [_C do [_{IP} you wonder [_{CP} whether [_{IP} John said [_{CP} ~~*t'*~~ [_C (that) [_{IP} Mary solved *t*]]]]]]]]]]]

As noted above, the long argument movement in (20) is not perfect, but is not nearly as degraded as long adjunct movement. The fact that it is not perfect follows on Chomsky's completely derivational view of Subjacency. Since one step of movement was too long, the example is marked as degraded. But since the offending trace is gone following deletion, (21) is not an ECP violation. As Chomsky and Lasnik (1993: 547) describe the situation, "An expression . . . is a Subjacency violation if its derivation forms a starred trace. It is an ECP

violation if, furthermore, this starred trace remains at LF." This is the residue of the Lasnik and Saito distinction between gamma-marking and the gamma filter (i.e., the ECP). Significantly, even on this "derivational" analysis, the ECP is actually not entirely derivational. It is derivational to the extent that potential violations are marked online anywhere in the course of the derivation. But it is representational in that whether the potential violations are actual violations is determined strictly at the *level* of LF.

Thus, as often turns out to be the case, there is empirically little to choose between the representational approach and the derivational one, and even technically, the difference is not very sharp. In this particular instance, Lasnik and Saito's representational account of adjunct movement locality and Chomsky's derivational (and partially representational) one both capture the facts, and both have a cost. For the former, the cost is the stipulation that gamma-marking is *strictly* representational, operating only at levels. For the latter, it is (18), which is evidently relevant solely for the phenomenon at issue.

3 Intermediate Trace Deletion and Economy

Chomsky (1991) (see also Chomsky and Lasnik 1993) proposes a more general way of capturing the effects of (18). Following Chomsky, for concreteness suppose that a trace in violation of the ECP is marked *. Further, continue to assume that deletion is one of the operations affecting phrase markers, but abandon the assumption that movement only optionally leaves a trace. Rather, if a trace is missing in a position from or through which movement has taken place, it is by virtue of deletion. What is then at issue is a deletability asymmetry between intermediate traces in adjunct chains and those in argument chains. Chomsky's account of this asymmetry is in terms of the theory of "economy" that he began to develop in detail in Chomsky (1991) (and which formed the basis for the "minimalist framework"). I briefly summarize the relevant concepts here.

First, the level of LF must satisfy the principle of Full Interpretation (FI). This principle is parallel to economy of derivation. Just as economy of derivation demands that there be no superfluous steps in derivations, FI requires that there be no superfluous symbols in representations. In particular, every element in an LF representation must be "legitimate." Chomsky (1995a: 153–4) suggests that only:²

the following elements are permitted at LF, each a chain . . . :

1. Arguments: each element is in an A-position . . .
2. Adjuncts: each element is in an \bar{A} -position.
3. Lexical elements: each element is in an X^0 position.
4. Predicates, possibly predicate chains if there is predicate raising . . .
5. Operator-variable constructions, each a chain (α_1, α_2) , where the operator α_1 is in an \bar{A} -position and the variable α_2 is in an A-position.

Then deletion, one instance of “Affect α ,” “may apply (and must apply) only to yield such an element, given an illegitimate object.”

Successive cyclic \bar{A} -movement from an A-position will produce a chain that is not one of the legitimate LF objects. The “tail” of the chain is in an A-position while all of the other links are in \bar{A} -positions, so the chain is neither an argument chain nor an adjunct chain. Nor is it an operator-variable construction, since those are limited to two-membered chains. The only way to make such a chain into a legitimate LF object is to delete the intermediate traces, whereupon it becomes an operator-variable construction. Notice that if one or more of those intermediate traces had been marked * in the course of the derivation, those *s would be eliminated when the traces are deleted.

Now consider the case of successive cyclic adjunct movement. This time, all of the members of the chain are in \bar{A} -positions. Hence, the chain is one of the legitimate objects – an adjunct chain. By economy of derivation, no deletion of intermediate traces can take place. Thus, if any of those intermediate traces were marked * (by virtue of a too-long step of movement having taken place), by economy of derivation those *s could not be eliminated. They would remain at the level of LF. The result is a legitimate object, but one that happens not to be well formed. Thus does Chomsky derive the argument-adjunct asymmetry with respect to long movement. It is principled just to the extent that the categorization of legitimate LF objects is, particularly the illegitimacy of a complete chain of *wh*-movement of an argument, as contrasted with the legitimacy of a corresponding adjunct chain. As before, the approach is derivational, in that the marking of a trace as being in violation of the ECP crucially takes place in the course of the derivation. In fact, it can be immediately upon the creation of the trace. Yet it is representational in the way that a derivational violation can be remedied. If no offending trace remains at the level of LF, the resulting structure is not an ECP violation, even if online it was.

Subadjacency remains strictly derivational:³ a long movement, even of an argument, causes some degradation of the sentence, as we have seen. Thus, it is evidently not the LF representation that is responsible for determining violation of Subadjacency. Rather, violation is determined online. This is all rather standard in Chomskyan work of the last several years. Interestingly, if we look back two decades before that, we can find a paradigm intriguingly parallel to the one arguing that the proper treatment of the ECP is partly representational. Ross (1969), in his seminal discussion of Sluicing, argued extensively that ellipsis involves deletion, and, given that, showed that deletion ameliorates island violations. Sluicing is the ellipsis phenomenon often found in embedded questions, given appropriate discourse context:

- (22) Mary hired someone.
Tell me who ~~Mary hired~~.

In (22), the *wh*-movement has been internal to one clause. Sluicing is also possible when the *wh*-movement has been long distance:

- (23) I heard that Mary hired someone.
Tell me who ~~you heard that Mary hired~~.

Now notice that if the long movement is out of an “island” (in this case, an adjunct island), the usual degradation is significantly lessened in the Sluicing construction:

- (24) I resigned because Mary hired someone.
?*Tell me who you resigned because Mary hired.
?Tell me who ~~you resigned because Mary hired~~.

If Sluicing is, indeed, a deletion phenomenon, as Ross argued (and as Chomsky has consistently maintained about ellipsis in general), then (24) provides evidence that Subjacency violation is not determined strictly online. This is because the Sluiced and non-Sluiced versions of (24) are identical throughout the syntactic portion of the derivation, and, in particular, at the point in the derivation where the excessively long step of movement takes place. The improvement created by Sluicing suggests that a Subjacency violation, rather like an ECP violation, places a * at some specific place in the structure, perhaps on the constituent constituting the island (rather than on the trace, in order to maintain the distinction between Subjacency violations and ECP violations). Deletion of (a constituent containing) the island then eliminates the *. If deletion is a PF process, then, rather curiously, we are led to the tentative speculation that it is the PF level that ultimately determines Subjacency violations, while, as we saw, it is the LF level that ultimately determines ECP violations.

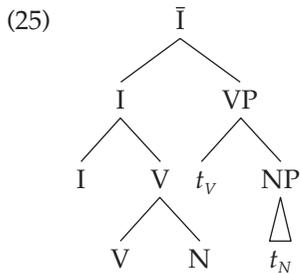
4 When Derivational Locality is Obscured by Later Operations

For both ECP and Subjacency, we have seen evidence that an online violation can be improved (if not remedied entirely) by a later operation that results in a change in the ultimate representation, LF in the first case, PF in the second. Interestingly, Chomsky argues for the reality of syntactic derivations by appealing to virtually the opposite state of affairs: situations where the ECP is *satisfied* online, but where the ultimate LF representation appears to be *in violation*, yet the resulting sentence is good. I quote one of Chomsky’s discussions:

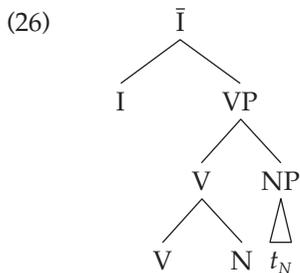
Viewed derivationally, computation typically involves simple steps expressible in terms of natural relations and properties, with the context that makes them natural “wiped out” by later operations, hence not visible in the representations to which the derivation converges. Thus, in syntax, crucial relations are typically local, but a sequence of operations may yield a representation in which the locality is obscured. Head movement, for example, is narrowly “local,” but several such operations may leave a head separated from its trace by an intervening

head. This happens, for example, when N incorporates to V, leaving the trace t_N and the $[_V V-N]$ complex then raises to I, leaving the trace t_V : the chain (N, t_N) at the output level violates the locality property, and further operations (say, XP-fronting) may obscure it even more radically, but locality is observed in each individual step. (Chomsky 1995a: 223–4)

Consider the structure at issue, where, for concreteness, I assume that V right-adjoins to I, and complements are to the right of heads:

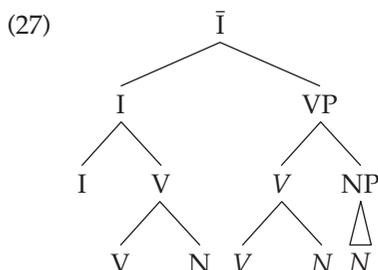


Chomsky's assumption, which I share, is that in this configuration, N is too distant from its trace to satisfy the Head Movement Constraint (a head cannot be separated from its trace by an intervening head), or whatever it derives from (the ECP or some version of relativized minimality,⁴ on a fairly standard view). At the immediately prior point of the derivation, however, HMC is evidently satisfied:



At this stage of the derivation, N is attached to the next head up from its trace, so there is no intervening head. The locality demanded by HMC obtains. But, as Chomsky notes, the next local step of movement has the effect of obscuring the locality of the first step. If we examined the LF output to determine locality satisfaction, we would incorrectly exclude (25), or so it seems.

One can, however, imagine a representational alternative. True, in (25), the V trace intervenes between the moved N and its trace. But what exactly is the V trace? Chomsky, among others, has argued that a trace in a position is a copy of the item that moved from that position. Under that hypothesis, consider the more articulated version of (25), where the 'traces' are in italics:



Now if N “antecedent governs” its trace in (26), presumably the intermediate *N* likewise antecedent governs the initial trace *N* in (27). Thus, under the copy theory of traces, even at the level of LF the locality requirement on head movement (i.e., on the traces of head movement) is arguably satisfied.

Interestingly, Chomsky himself alludes to this alternative to the strictly derivational analysis, but he rejects it:

It is generally possible to formulate the desired result in terms of outputs. In the head movement case, for example, one can appeal to the (plausible) assumption that the trace is a copy, so the intermediate *V*-trace includes within it a record of the local $N \rightarrow V$ raising. But surely this is the wrong move. The relevant chains at LF are (N, t_N) and (V, t_V) , and in these the locality relation satisfied by successive raising has been lost. (1995a: 224)

Chomsky’s objection to the representational alternative raises an important question. If, as I assume Chomsky implies here, the relevant chain at LF consists of just the moved item and its *initial* trace, his argument is fairly compelling. The representational alternative makes crucial reference to intermediate traces, but if they have no relevance at LF, such reference is plausibly illicit. (Note that no such conceptual objection applies to the derivational approach: in the whole range of theories I am considering here, intermediate *steps* exist regardless of whether they still leave evidence of their existence in the form of traces (copies) at LF.) However, there is reason to believe that (at least some) chains at LF are not just two-membered, but rather include intermediate traces along with the initial one. Recall the Chomsky (1991) analysis summarized above of the *wh* adjunct-argument asymmetry with respect to degree of unacceptability of island violations. That analysis demanded that every intermediate step of movement leave a trace, and that when a particular step is too long, the trace created is marked *. At LF, a * marked trace indicates ungrammaticality. This much could, of course, be easily restated in strictly derivational terms. But the account of the *asymmetry* had a fundamentally representational aspect. Both adjuncts and arguments make successive short moves leaving traces; and in both instances, when the step of movement is too long, the trace is marked *, whether the step is the first step of movement or a later step. But intermediate traces of adjuncts cannot be eliminated (given that the chain already is a well-formed object, conforming to the second criterion

– Adjuncts: each element is in an \bar{A} -position). Intermediate traces of arguments undergoing \bar{A} -movement can be eliminated (in fact must be, on Chomsky’s account). Then at LF, if the former kind of derivation involved a long intermediate step, the chain will contain a * marked intermediate trace. A corresponding derivation involving movement of an argument, on the other hand, will not contain such a trace, all intermediate traces having been eliminated so that the chain will conform to the fifth criterion: Operator-variable constructions, each a chain (α_1, α_2) , where the operator α_1 is in an \bar{A} -position and the variable α_2 is in an A-position. As far as I know, there is still no better way to account for the adjunct-argument asymmetry, or even one equally good. To the extent that this is so, we have evidence that (at least some) intermediate traces exist at LF. The strongest possible form of the conceptual argument against intermediate traces as licensors of lower traces can thus be answered.⁵

5 Locality and Reconstruction

There is also evidence that, as Chomsky himself suggests, a trace is indeed a copy of the moved item. This assumption is clearly necessary for the representational treatment of ECP satisfaction in the head movement case under consideration. The evidence comes from so-called reconstruction effects, where for certain purposes, especially those having to do with anaphoric connection, a moved item behaves as if it were in the position of one of its traces. Barss (1986) provides an extensive discussion of such phenomena, and Chomsky (1993) concludes from them that traces are copies and that they exist at LF, the level of representation that Chomsky takes to be relevant to “Binding Theory.” (28) is based on one of Chomsky’s examples, of a standard type:

(28) Mary wondered which picture of himself Bill saw *t*

The point is that *Bill* can serve as the antecedent of *himself* in (28) even though *Bill* does not seem to c-command *himself*. Chomsky’s analysis of this fact relies on the trace being a copy of the moved item. Abstracting away from details that are irrelevant here, the binding requirement of the anaphor *himself* is satisfied by virtue of the copy of *himself* in the “trace.” On this approach, at least initial traces exist at LF, and they are copies. I turn now to the issue of intermediate traces, since intermediate traces were directly relevant to the adjunct-argument asymmetry discussion.

With respect to intermediate traces, there is evidence like that in (28) for the existence of such traces, and for their status as copies. But, as we will see, that evidence will actually run counter to the conclusion based on the adjunct-argument asymmetry. First, the evidence for intermediate traces: Barss (1986) observes that the kind of reconstruction effect with respect to initial position seen in (28) also arises with respect to intermediate position. Consider (29), whose derivation involves successive movement:

(29) Which picture of himself does Mary think that John said that Susan likes

Obviously, with the *wh*-phrase in its surface position, *himself* is not in an appropriate *c*-command relation with its antecedent *John*. Further, the initial position of the *wh*-phrase, object of the most deeply embedded verb, has *himself* too distant from *John*, as evidenced by (30):

(30) *Mary thinks that John said that Susan likes pictures of himself

The appropriate position seems to be Spec of the lowest CP, as seen in the approximately parallel (31):

(31) Mary thinks that John said that pictures of himself, Susan likes

A trace (copy) in the intermediate position would thus provide the necessary structural relation to license the anaphoric connection.

I have thus far summarized two arguments for the existence of intermediate traces. Unfortunately, they rather neatly contradict each other. The argument based on the adjunct-argument asymmetry presupposes an analysis under which intermediate traces of arguments are eliminated prior to the LF level (while intermediate traces of adjuncts remain). But the reconstruction effects just examined, evidently implicating intermediate traces, involved *wh*-argument movement. The two lines of reasoning cannot both be correct. I do not know of a good counter to the first line of reasoning, so I will assume that it is correct. That leaves the intermediate reconstruction facts to be addressed in another way.

As indicated above, Chomsky's specific approach to reconstruction in general centers on traces (copies) visible at the LF level, a strongly representational approach, interestingly enough. According to Chomsky (1993), developing a suggestion of Chomsky and Lasnik (1993), if the basic aspects of Binding Theory hold only at the LF interface, then we can move toward a simple interpretive version of Binding Theory which directly maps structural properties into semantic ones. There is no syntactic filtering of the sort found in the three binding conditions of Chomsky (1981). In fundamental respects, this represents a return to much earlier generative treatments of anaphora, such as RI (Rule of Interpretation), the Chomsky (1973) ancestor of Condition B, and the Disjoint Reference Rule of Lasnik (1976), the ancestor of Condition C. But, as we saw, this approach seems to demand the presence at LF of intermediate traces (even of arguments), a consequence incompatible with the treatment of the adjunct-argument island asymmetry.

An appealing alternative would rely on a more derivational approach to anaphoric connection. Belletti and Rizzi (1988) present just such a theory, at least for one class of binding phenomena. They argue that Condition A can be satisfied online, at any point in the derivation. Their concern is, of course, reconstruction effects abstractly similar to those discussed above. The crucial property is that at the output of the derivation, Condition A is apparently not

satisfied while at some prior point in the derivation, it would have been. In empirical effect, the online proposal is very close to the LF trace account. The technical difference is that it precisely does not rely on LF traces, so it would allow us to preserve the account of the adjunct-argument asymmetry that demands the elimination of intermediate argument traces. Note that if we take the Binding Theory to consist not of conditions on form, but rather of interpretive principles, as suggested in the preceding paragraph, then this amounts to saying that there is no specific *level* of LF. That is, there is no one single representation that uniquely interfaces with semantics.

Within generative grammar, there are numerous antecedents for this possibility. For example, in the classic formulation of generative grammar in the mid-1950s, the closest analogue to LF is the “T-marker,” the record of the transformational derivation. Later, in the orthodox form of the “Extended Standard Theory” (EST) of the early 1970s, the input to semantic interpretation consisted of two representations – Deep Structure and Surface Structure. Variant EST formulations existed at the same time, some of them clearly anticipating the Belletti–Rizzi approach. Jackendoff (1972), for example, proposed a theory of anaphora that included interpretive rules operating at the end of each syntactic cycle. Similarly, Lasnik (1972) suggested “cyclic” interpretation of the scope of negation, and Lasnik (1976) extended this to the scope of other operators. There are analogues on the PF side of the grammar as well, as with the Bresnan (1971) arguments that the rule responsible for the assignment of sentence stress in English applies not at Surface Structure, as had been assumed, but at the end of each syntactic cycle. Needless to say, all of these analyses are fundamentally derivational. There is a modern version of these approaches which is even more derivational. Epstein (in press) (see also Uriagereka in press) suggests that *all* interpretive information is provided online, in the course of the syntactic derivation. In a way reminiscent of Chomsky (1955), there is no *level* of LF per se. Under Epstein’s approach, after each syntactic operation, any structural information relevant to semantics (and presumably phonetics as well, though Epstein’s specific arguments are not concerned with that interface) is available. This approach can easily accommodate the reconstruction phenomena sketched above. However, it seems that it is not consistent with the account of the adjunct-argument asymmetry discussed here, or with any close variants, since such accounts crucially rely on a late representation masking properties present earlier in the derivation (illicitly long steps of movement in this case).⁶ For present purposes, then, I will continue to assume one of the hybrid approaches incorporating derivational and representational aspects.

6 The Motivation for Movement

Thus far, I have been examining the “derivation” vs. “representation” question with respect to general locality constraints on movement. At this point, I turn

to an arguably more fundamental aspect of the question: what determines whether movement occurs in the first place? Chomsky (1995a) argues on conceptual and, to some extent, empirical grounds that movement is always morphologically driven. That is, there is some formal feature that needs to be checked, and movement provides the configuration in which the checking can take place. Chomsky also provides strong reason to believe that, all else being equal, covert movement (that in the LF component) is preferred to overt movement, a preference that Chomsky calls “Procrastinate.” When movement is overt, rather than covert, then, it must have been forced to operate “early” by some special requirement. The major phenomenon that Chomsky considers in these terms is verb raising (which Chomsky takes to be overt in French and covert in English). He also hints at a contrast in object shift, overt in some languages and covert in others. Chomsky (1993, 1994, 1995a)⁷ codes the driving force for overt movement into “strong features,” and presents three successive distinct theories of precisely how strong features drive overt movement. These three theories, which I will summarize immediately, are of interest to my central question, since the first two of them are explicitly representational in the relevant sense, while the third is derivational:

- (32) a. A strong feature that is not checked in overt syntax causes a derivation to crash at PF. (Chomsky 1993)
- b. A strong feature that is not checked (and eliminated) in overt syntax causes a derivation to crash at LF. (Chomsky 1994)
- c. A strong feature must be eliminated (almost) immediately upon its introduction into the phrase marker. (Chomsky 1995a)

All three of these proposals are designed to force overt movement in the relevant instances (e.g. verb raising in French; a strong V feature of Infl will cause a violation in one of the three ways listed in (32)) and all are framed within a “Minimalist” conception of grammar. In addition to what was indicated in the preceding paragraph, the most pertinent aspect of this conception is the reduction of levels from the four that were standard in Government Binding analyses (D-structure, S-structure, LF, and PF) to just the last two, the “interface” levels LF and PF. There is no D-structure level of representation since the work of building a structure is done by generalized transformations, as it was before recursion in the base was introduced in Chomsky (1965). Interestingly, this return to an earlier approach replaces a representational view with a strongly derivational one. As for S-structure, it remains, but merely as the point where the derivation branches off toward LF on one path and toward PF on another. Chomsky calls this point “Spell-Out.” The claim is that it has no further properties. In particular, it is not the locus of satisfaction of any conditions or constraints (as it obviously had been in Government Binding work). With this much background, we can proceed to a more careful consideration of the three proposals in (32).

Chomsky (1993) argues that the treatment in (32a) follows from the fact that parametric differences in movement, like other parametric differences, must

be based on morphological properties reflected at PF. (32a) makes this explicit. Chomsky suggests two possible implementations of the approach:

“strong” features are visible at PF and “weak” features invisible at PF. These features are not legitimate objects at PF; they are not proper components of phonetic matrices. Therefore, if a strong feature remains after Spell-Out, the derivation crashes . . . Alternatively, weak features are deleted in the PF component so that PF rules can apply to the phonological matrix that remains; strong features are not deleted so that PF rules do not apply, causing the derivation to crash at PF. (Chomsky 1993: 198)

There is presumably only one other possible type of representational approach, given minimalist assumptions: one that involves LF, rather than PF. Chomsky (1994) proposes such an analysis, (32b), based on an empirical shortcoming of (32a). What is at issue is the unacceptability of sentences like (33):

(33) *John read what?

Assuming that the strong feature forcing overt *wh*-movement in English resides in interrogative C,⁸ the potential concern is that that C might be introduced in the LF component, where, checked or not, it could not possibly cause a PF crash, since, as far as PF knows, the item does not exist at all. Yet (33) is bad, so such a derivation must be blocked. This problem arises in the general context of fitting lexical insertion into the grammar. In most circumstances, there is no need for a specific prohibition against accessing the lexicon in the PF or LF component. (33) represents a rare problem for the assumption that lexical insertion is free to apply anywhere. I quote Chomsky’s discussion of this point.

Spell-Out can apply anywhere, the derivation crashing if a “wrong choice” is made . . . If the phonological component adds a lexical item at the root, it will introduce semantic features, and the derivation will crash at PF. If the covert component does the same, it will introduce phonological features, and the derivation will therefore crash at LF . . . Suppose that root C (complementizer) has a strong feature that requires overt WH-movement. We now want to say that unless this feature is checked before Spell-Out it will cause the derivation to crash at LF to avoid the possibility of accessing C after Spell-Out in the covert component. (Chomsky 1994: 60)

Chomsky proposes to implement this basic idea in the following way: “Slightly adjusting the account in Chomsky (1993), we now say that a checked strong feature will be stripped away by Spell-Out, but is otherwise ineliminable” (Chomsky 1994: 60).

Chomsky (1995a) rejects the representational approach in (32a), and the conceptual argument he gives evidently applies equally to the alternative representational approach in (32b). He discounts such an account as an “evasion,” and proposes what he claims is a more straightforward statement of the phenomenon:

formulation of strength in terms of PF convergence is a restatement of the basic property, not a true explanation. In fact, there seems to be no way to improve upon the bare statement of the properties of strength. Suppose, then, that we put an end to evasion and simply define a strong feature as one that a derivation “cannot tolerate”: a derivation $D \rightarrow \Sigma$ is canceled if Σ contains a strong feature. (Chomsky 1995a: 233)

Chomsky’s summary of that approach is given in (34):

- (34) “A strong feature . . . triggers a rule that eliminates it: [strength] is associated with a pair of operations, one that introduces it into the derivation . . . a second that (quickly) eliminates it.” (Chomsky 1995a: 233)

This approach is strongly derivational.

In addition to the claim that (32c) avoids the evasion present in (32a) (and presumably in (32b) as well), Chomsky implies that (32c) is superior to the representational theories on another ground as well: that of computational complexity. While under theory (32c), the determination that overt movement is needed is made online, immediately as the head containing the strong feature is introduced into the structure, the two representational theories require considerable “look-ahead.”⁹ At a given point in the overt portion of a derivation, it is necessary to inspect the PF or LF representation to see whether Procrastinate is to be evaded.

At this point, I want to give a fuller elaboration of (32c), since that fuller elaboration is of interest in its own right, but also because a potential problem emerges. Chomsky suggests the following: “The intuitive idea is that the strong feature merged at the root must be eliminated before it becomes part of a larger structure by further operations” (Chomsky 1995a: 234). After considering how derivations work in general, he goes on to indicate that: “the descriptive property of strength is [(35)]. Suppose that the derivation D has formed Σ containing α with a strong feature F . Then [(35)] D is canceled if α is in a category not headed by α ” (1995a: 234).

According to Chomsky, there are two significant consequences of this approach: (i) that cyclicity follows,¹⁰ (ii) that a strong feature is checked by an overt operation: “We . . . virtually derive the conclusion that a strong feature triggers an *overt* operation to eliminate it by checking. This conclusion follows with a single exception: covert merger (at the root) of a lexical item that has a strong feature but no phonological features” (1995a: 233). This exception involves a kind of example we have seen before:

- (36) *John read what

Recall that it was exactly this type of example that led Chomsky to reject the PF crash theory (32a) in favor of the LF crash theory (32b). But, as Chomsky in effect acknowledges, the problem now re-arises in the derivational theory. To be specific, how can derivation (37) be blocked?¹¹

- (37) Spell-Out: John read what
 LF: C [strong Q] John read what

The structure in (37) never needs to become part of a larger structure, so the fact that the strong feature of C was not checked overtly should not hinder the derivation. Chomsky proposes to bar such covert insertion of an item bearing a strong feature with the economy principle (38):

- (38) α enters the Numeration only if it has an effect on output.

The Numeration is basically the set of items chosen from the lexicon that will form the basis for the derivation. (38) is intended to mean that an item can be chosen from the lexicon only if it will have a phonetic consequence or a semantic consequence. This immediately raises a question concerning the central argument for the derivational approach – that it eliminates the look-ahead inherent in the PF and LF approaches. There is extreme look-ahead here, all the way from the very beginning of the derivation, the Numeration, to the very ends, the phonetic and semantic interfaces.

Apart from this conceptual question, there is an empirical question about whether the correct result is obtained. There is reason to think that it is not. Consider the situation at issue, insertion in the LF component of interrogative C in English, a language in which C has a strong *wh* feature. (38) purports to prevent this. To see whether it does, we first have to ask whether this C has an effect on output. By definition, covert insertion of a C will have no phonetic effect. Will it have an effect at the LF output? Either it will or it will not. If it will, then covert insertion is allowed, and we generate (36) with structure (39):

- (39) C [_{IP} John read what]

Since this is not the correct result, suppose instead that C will not have a semantic effect. Then we cannot generate (36) with structure (39), so the problem is apparently solved under the assumption that insertion of interrogative C has no effect on semantic output. As Chomsky states the situation:¹² “the interface representations (π , λ) are virtually identical whether the operation takes place or not. The PF representations are in fact identical, and the LF ones differ only trivially in form, and not at all in interpretation” (1995a: 294). But the goal is actually more general than just ruling out (36) with structure (39). Rather, it is ruling out (36) altogether. Under the assumptions just spelled out, (36) is successfully excluded with C covertly inserted. But what if C is not inserted at all? That is, what if the LF is just the same as the “S-structure”?

- (40) [_{IP} John read what]

(40) violates no morphological requirements, and, if C has no effect on output, the assumption that was necessary in order to exclude (36) *with* C inserted, then it should mean exactly *What did John read?* To summarize, if C has a semantic effect, inserting it in LF should be permitted. And if it does not have

a semantic effect, *not* inserting it should be of no consequence. Thus, even given the new economy condition (38), (36) is allowed as a standard interrogative, an incorrect result. In this regard, then, the derivational theory of strong features ultimately fares no better than the PF theory. Either theory demands an additional stipulation, perhaps just that lexical insertion is prohibited in the covert component¹³ (a result Chomsky was trying to deduce, but, as we have just seen, not completely successfully).

7 The Nature of Strong Features

There are further phenomena that might bear on the question of representational vs. derivational approach to strong features. The derivational approach outlined above demands that a strong feature triggering overt movement is always a feature of the position that an item is moving to (a feature of an “attracting head”). This is so since a strong feature in an item that will move would virtually never be able to be eliminated before that item is embedded, by a generalized transformation, in a larger structure. The derivation would therefore fatally terminate before the feature ever had a chance to be checked and deleted. Evidence for a strong feature in an item forcing that item to move would therefore be evidence against the derivational approach, at least as formulated in (32c). Certain ellipsis paradigms provide such evidence, at least on the face of it. I turn to one such paradigm now, involving Pseudogapping.¹⁴

(41) presents a few examples of Pseudogapping from the classic study by Levin (1978):

- (41) a. If you don't believe me, you will \emptyset the weatherman
 b. I rolled up a newspaper, and Lynn did \emptyset a magazine
 c. Kathy likes astronomy, but she doesn't \emptyset meteorology

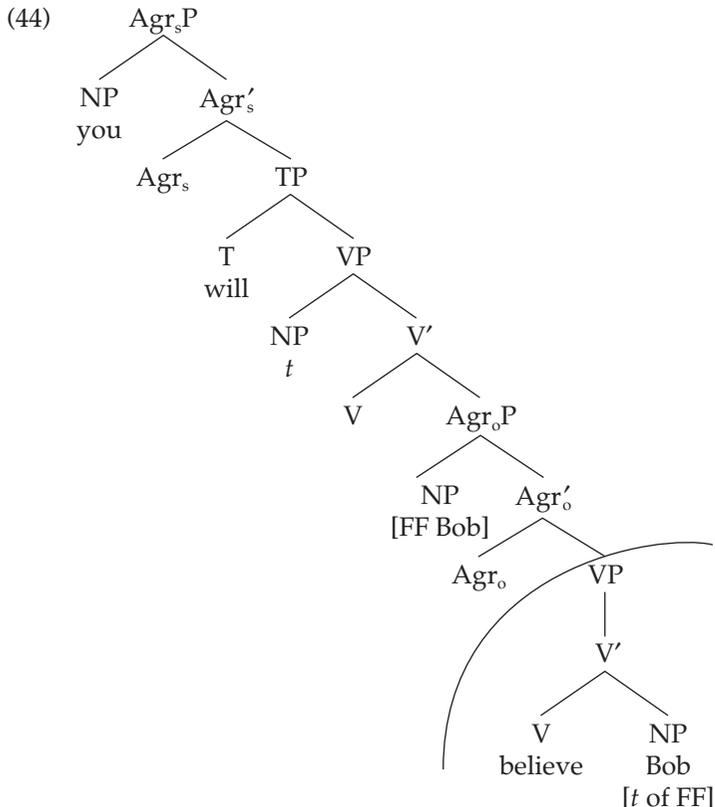
While in many instances, it might appear that the process is simply elision of the main verb, there is evidence that more is involved. In the examples in (42), the ellipsis site includes the main verb plus (a) the small clause predicate or (b) the second object in a double object construction:

- (42) a. The DA proved Jones guilty and the Assistant DA will ~~prove~~ Smith
 ~~guilty~~
 b. ?John gave Bill a lot of money, and Mary will ~~give~~ Susan ~~a lot of~~
 ~~money~~

Rejecting the possibility of an ellipsis rule affecting a discontinuous portion of the structure, Jayaseelan (1990) proposes that Pseudogapping constructions result from VP ellipsis, with the remnant NP having moved out of the VP by Heavy NP Shift. In Lasnik (1995c) I argue that this proposal is correct in its essentials, though wrong in certain details. I modify Jayaseelan's analysis by positing raising to Spec of Agr_o, instead of HNPS, as the process removing the remnant from the ellipsis site.

Under “standard” Minimalist assumptions, going back to Chomsky (1991), raising to Spec of Agr_o is covert, taking place in the LF component. Given Jayaseelan’s goal, adopted also in Lasnik (1995c), of analyzing Pseudogapping as affecting a constituent, the ellipsis process must then be analyzed as copying in the LF component, rather than deletion in the PF component. However, on the theory of LF movement advocated by Chomsky (1995a), and further defended by Lasnik (1995c, 1995d), the necessary structure would not be created even in *covert* syntax. On that theory, since movement is invariably triggered by the need for formal features to be checked, all else being equal only formal features move. When movement is overt (triggered by a strong feature), PF requirements demand that an entire constituent move, via a sort of pied-piping. However, when movement is covert, PF requirements are irrelevant so economy dictates that movement *not* be of the entire constituent. But it is very difficult to see how covert raising of just the formal features (FF) of NP to Spec of Agr_o could possibly create a suitable ellipsis licensing configuration. The structure of the second conjunct of (43) would be as in (44), with *believe* incorrectly within the ellipsis site rather than outside of it:

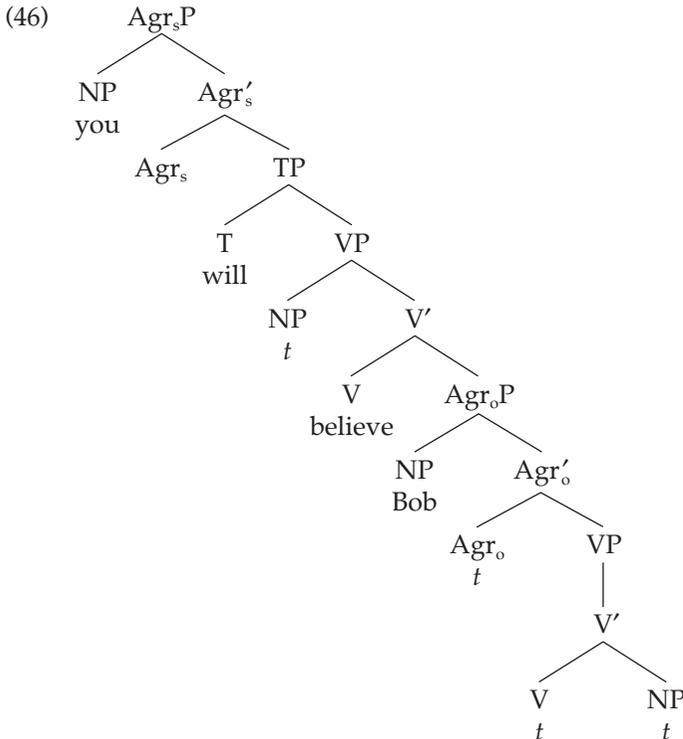
(43) Mary will believe Susan, and you will Bob



It seems then that if movement newly creates a configuration licensing ellipsis, the movement must be overt rather than covert. Before I indicate how that is possible in the present instance, I note that if the movement is overt, then the conclusion above, that ellipsis must involve LF copying, no longer follows. If the licensing configuration must be created prior to the LF/PF split regardless, then ellipsis could just as easily be a PF deletion phenomenon.

Now early Minimalist literature, such as Chomsky (1991, 1993) and Chomsky and Lasnik (1993), did have accusative NP raising to Spec of Agr_o, but covertly rather than overtly. However, Koizumi (1993, 1995), developing ideas from Johnson (1991), argues, instead, that that raising is always overt, driven, as usual, by a strong feature. In Lasnik (1995f) I suggest that the strong feature in this instance is an “EPP feature” residing in Agr, hence the same feature that drives overt subject raising, the modern technological implementation of the Extended Projection Principle of Chomsky (1981). I will have little more to say here about this particular strong feature. I will, however, address another strong feature that must be involved in simple transitive sentences without ellipsis. Given that word order in English is V-O rather than O-V, if object raises out of VP, verb must normally raise still higher. Koizumi’s proposal, which he calls the split VP hypothesis, is that V raises, via Agr_o, to a higher “shell” V position, as shown in (46) for the sentence in (45):

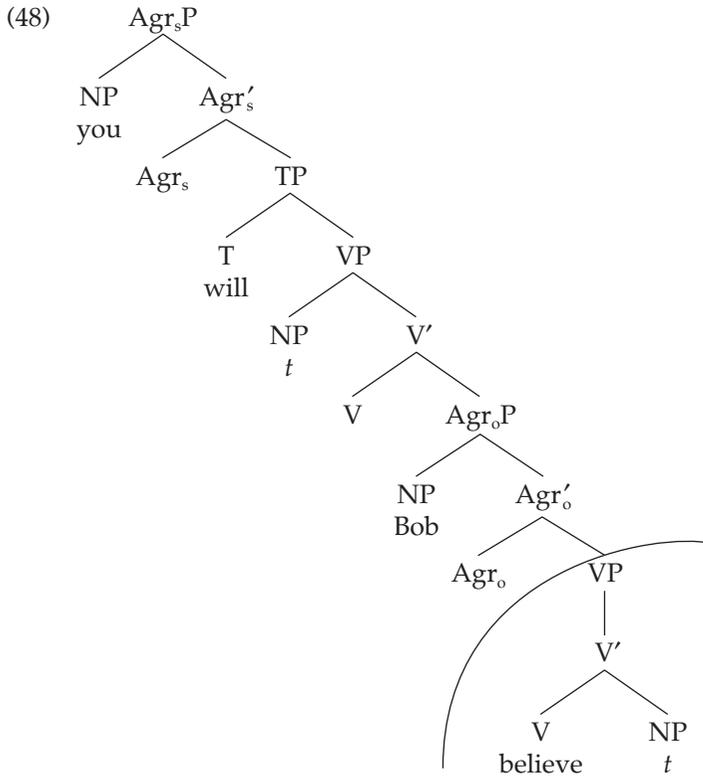
(45) You will believe Bob



Under this general hypothesis, consider a simple Pseudogapping example such as (47):

(47) You might not believe me but you will Bob

If *Bob* overtly raises to Spec of Agr_o while *believe* remains in situ, then deletion of the residual VP produces (47). The relevant structure is shown in (48):



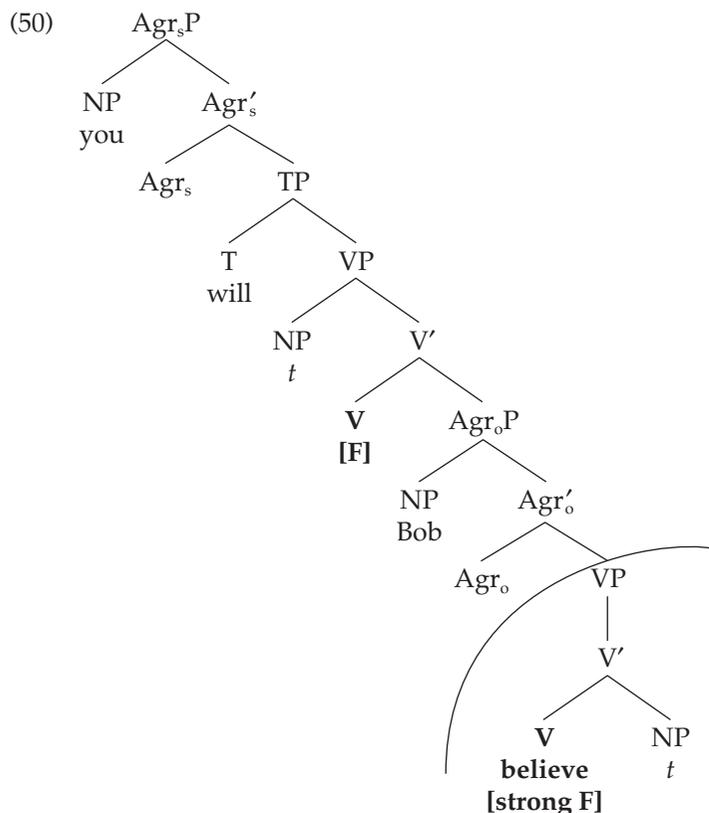
Deletion of the lower VP yields the Pseudogapping example in (47).

The question that now arises is why the V need not raise in Pseudogapping constructions, given that in non-elliptical sentences it must:

(49) *You will Bob believe

By hypothesis, a strong feature is involved. Yet there seem to be two possibilities for a convergent derivation. The verb can raise as in (46), presumably checking the relevant strong feature. Alternatively, the verb can be deleted along with its containing VP as in (48). This state of affairs receives a

straightforward account under theory (32a) of strong features, the PF approach, under the new hypothesis that the strong feature forcing the verb to raise overtly is a feature of the lexical verb itself, rather than of the target position it raises to. The overt raising derivation is essentially unaffected by this change in perspective. The ellipsis structure is more interesting. Consider (48) from this point of view:



If *believe* fails to raise, and no other relevant process takes place, the strong feature that is not overtly checked causes (50) to crash at PF. But if the lower VP containing *believe* is deleted in the PF component, then, patently, the strong feature cannot cause a PF crash, since the (category containing the) feature will be gone at that level. It is not obvious how to capture this result under theory (32b), the LF crash account of strong features (since even if the strong feature is deleted in PF, it will still be present at LF), or theory (32c), the derivational theory. Thus, we apparently have an argument for theory (32a).

Surprisingly, it turns out that the ellipsis facts can be reconciled with the derivational theory. There is a possible alternative analysis of those ellipsis facts, based on the Chomsky (1995a) theory of “pied-piping,” particularly as explicated by Ochi (1997, 1999).¹⁵

Ochi, following Chomsky, considers the nature of pied-piping, the usual reflex of movement triggered by a strong feature. Chomsky (1995a) gives the following characterization:

For the most part – perhaps completely – it is properties of the phonological component that require pied-piping. Isolated features and other scattered parts of words may not be subject to its rules, in which case the derivation is canceled; or the derivation might proceed to PF with elements that are “unpronounceable,” violating F[ull] I[n]terpretation]. (Chomsky 1995a: 262)

Overt movement consists of a complex of operations under this approach:

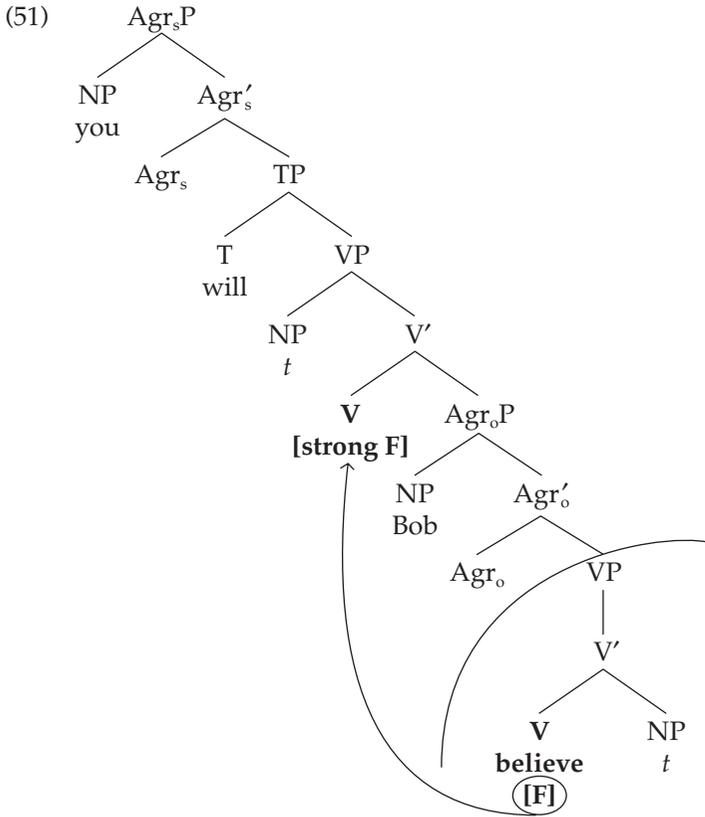
Applied to the feature F , the operation Move thus creates at least one and perhaps two “derivative chains” alongside the chain $CH_F = (F, t_F)$ constructed by the operation itself. One is $CH_{FF} = (FF[F], t_{FF[F]})$, consisting of the set of formal features $FF[F]$ and its trace; the other is $CH_{CAT} = (\alpha, t_\alpha)$, α a category carried along by generalized pied-piping and including at least the lexical item containing F . CH_{FF} is always constructed, CH_{CAT} only when required for convergence . . . As noted, CH_{CAT} should be completely dispensable, were it not for the need to accommodate to the sensorimotor apparatus. (1995a: 265)

Note that this seems to assume the second of the two possibilities Chomsky mentioned in the prior passage, that is, that failure of pied-piping causes a violation specifically at PF. Chomsky goes on to observe that even overt movement might be possible without pied-piping under certain circumstances, if no phonological requirement is violated: “Just how broadly considerations of PF convergence might extend is unclear, pending better understanding of morphology and the internal structure of phrases. Note that such considerations could permit raising without pied-piping even overtly, depending on morphological structure” (1995a: 264).

Consider now how the ellipsis phenomena examined above might be reanalyzed in terms of this theory. Recall my analysis of Pseudogapping in terms of the PF crash theory of strong features. Assuming the split VP hypothesis, in a non-elliptical transitive sentence, for example, the object raises to Spec of Agr_o and the lexical V raises to the higher shell V position in order that a strong feature of the lexical V will be checked. If the V does not raise, a PF crash will ensue, but only if the offending item exists at that level. Hence, under a deletion account of ellipsis, ellipsis provides another way to salvage the derivation. When the lower VP is deleted without the V having raised, a PF crash is avoided and the result is acceptable Pseudogapping.

The alternative account preserves the idea of deletion averting a PF crash, but the potential crash now has another cause. The feature driving overt V -raising could be a strong feature of the *higher* V . Once the matching feature of the lower lexical V is “attracted” out of the lower V , the lower V becomes defective. A PF crash will be avoided if either pied-piping or deletion of a

category containing the lower V (VP Deletion = Pseudogapping in the relevant instances) takes place. This is illustrated in (51):¹⁶



Even under the derivational theory of strong features, then, there is a way to capture the saving effect of ellipsis in the Pseudogapping construction.

Thus, the essence of the PF account of the ellipsis facts based on the PF theory of strong features can be captured under the derivational theory as well, a rather surprising result, and, perhaps, a welcome one if Chomsky's conceptual arguments for the derivational theory are accepted. It must be noted, though, that even the account of the ellipsis facts relying on the derivational approach to strong features does ultimately rely on a property of the PF interface level. Before concluding, I will briefly mention one well-known, and rather powerful, argument that strong features reside in some moving categories (not just in attracting heads), and that the basic premise of the derivational theory (32c) is therefore incorrect. There is a great deal of literature, going back to Toman (1982) and Rudin (1982, 1988), discussing the phenomenon of multiple *wh*-movement in the Slavic languages. Bošković (1998)

presents a treatment of Serbo-Croatian multiple *wh*-movement in terms directly relevant to the present discussion. Bošković argues that in Serbo-Croatian, *wh*-phrases have a strong focus feature, and that that is why they all have to move overtly:

- (52) Ko šta gdje kupuje?
who what where buys "Who buys what where?"
*Ko kupuje šta gdje?
*Ko šta kupuje gdje?
*Ko gdje kupuje šta?

Whether he is right about the precise identity of the feature will not be of concern here. Regardless, the fact that all of the *wh*-phrases must move overtly strongly suggests Bošković's basic conclusion, that the strong feature driving the movement resides in the *wh*-phrases themselves. If, instead, it resided in the head to which they move (or just in that head), why would not the movement of just one of them suffice (as in English)? But if the *wh*-phrases contain strong features to be checked against higher heads, there is no way that strong features can invariably be eliminated immediately upon their insertion into the structure. Thus, the technology behind the Chomsky (1995b) strictly derivational approach to strong features cannot be accepted in its entirety.

8 Summary

In summary, the issue of derivation vs. representation in transformational grammar turned out to be every bit as subtle as it seemed that it would. But there is some reason for optimism that the topic is not completely intractable. Consideration of certain movement locality asymmetries (adjunct vs. argument; full form vs. elliptical) at least helped bring the issue into sharper focus. Those locality effects seem to demand a hybrid account that is crucially derivational, measuring length of each successive step of movement, but is partly representational as well, inspecting the LF and PF representations for violation markers. The attempt to reconcile the treatment of long adjunct/argument movement asymmetry with intermediate position reconstruction effects led to the tentative conclusion that Binding Conditions (or at least condition A) are satisfied "derivationally." Also derivational is the third of Chomsky's three analyses of the strong features which provide the driving force for overt movement. Interactions between ellipsis and (lack of) overt movement at first appeared to argue against that derivational view and in favor of a PF account, but on closer inspection turned out to be consistent with it, given plausible assumptions about feature movement and pied-piping. How to reconcile this with the facts of multiple *wh*-fronting in Slavic, and, indeed, how to fit all of these pieces together into a coherent whole, remains a task for further investigation.

NOTES

- * I am indebted to Željko Bošković for very helpful suggestions at every stage of this research, and to Chris Collins for valuable recommendations for improvement of the presentation.
- 1 In the technical sense of Chomsky (1955).
 - 2 See Browning (1987) and Déprez (1989) for antecedents of this typology.
 - 3 See Freidin (1978) for an early representational view of Subjacency as well as for important discussion of the whole derivation–representation question.
 - 4 The history of this particular condition is interestingly relevant to the theme of this chapter. The earliest version (though not under this particular name) appears in Rizzi (1986b), in the context of an argument against derivation and for a strictly representational approach. Chomsky (1991) and Chomsky and Lasnik (1993), however, present a derivational formulation, based on Rizzi (1990).
 - 5 Note that this line of reasoning also seems to run directly counter to the Chomsky (1995a: 224) argument cited above. This is so since the logic of the intermediate trace deletion account of argument vs. adjunct movement also should extend to head movement, precluding intermediate trace deletion in that situation as well. In fact, Chomsky (1991), after spelling out the adjunct movement analysis, indicates that head movement should behave in parallel, though he gives no examples. A relevant example would involve licit short head movement followed by illicit long head movement, with the resulting sentence totally unacceptable. A potential instance is shown in (i), where *is* has raised to Infl in the lower clause, then to Comp in the matrix:
 - (i) *Is Mary has said John clever
 - 6 Chris Collins (personal communication) suggests an interesting reinterpretation of the mechanism of trace deletion outlined above that might render it consistent with the Epstein approach: suppose that immediately upon their creation, intermediate traces of argument *wh*-movement are deleted, while, as before, adjunct traces are not eliminated. Then the presence of a starred trace at any step in the derivation would yield an ECP violation. I hope to explore this possibility in future research.
 - 7 Page references to Chomsky (1994) will be to pages in Campos and Kempchinsky (1995), one of two books where the article was published (the other being Webelhuth 1995a). I use the 1994 citation for ease of exposition, and to keep the historical development of the ideas I am exploring clear. Page references to Chomsky (1993) will be to the reprint in Chomsky (1995b).
 - 8 Notice that in English, the relevant strong feature could not reside in the *wh*-phrase, since in multiple interrogation, all but one of the *wh*s remain in situ, hence unchecked in overt syntax:
 - (i) Who gave what to who
 - 9 See Collins (1997) for extensive discussion of the issue of

- computational complexity and look-ahead.
- 10 At least for overt movement, though Chomsky does not add this qualification.
- 11 As Máire Noonan pointed out in the discussion following my presentation of some of this material at the 1997 Open Linguistics Forum in Ottawa, even *overt* insertion of C in the matrix without overt *wh*-movement seems to be incorrectly allowed by Chomsky's formulation: "[A] strong feature merged at the root must be eliminated before it becomes part of a larger structure by further operations" (1995a: 234). Chomsky elaborates this as follows: "Suppose that the derivation D has formed Σ containing α with a strong feature F. Then . . . D is canceled if α is in a category not headed by α " (1995a: 234). When, as in the example now under consideration, the interrogative will not be embedded, hence will never be part of a larger structure, nothing demands that the strong feature be checked overtly.
- 12 Here I am somewhat reinterpreting what Chomsky actually said, since prior context indicates that Chomsky was referring to the operation of "insertion of strong features." But I do not see how to fit such an operation (insertion of strong features independent of the item of which they are features) into the theory. Possibly I am missing something crucial.
- 13 Or at least lexical insertion of an item with a strong feature.
- 14 In Lasnik (1999) I provide a parallel argument based on Sluicing.
- 15 Ochi's concern is the locality of movement, in particular the fact that only relativized minimality effects follow in any natural way from Attract F. Other island effects seem to make sense only from the point of view of the moving item, rather than the target. Ochi proposes that the feature chain, created by Attract F, is responsible for the relativized minimality effects while the pied-piping chain, created by Move α in order to remedy the defect in α created by the movement of the formal features out of α , is responsible for other island effects.
- 16 The entire tree is shown in (51) just for expository purposes. In the actual derivation, the strong feature of the higher V would attract the corresponding feature of *believe* immediately upon the introduction of the former into the phrase marker, in accord with the derivational theory.