Investigations of covert phrase movement

by

Jonathan W. Nissenbaum
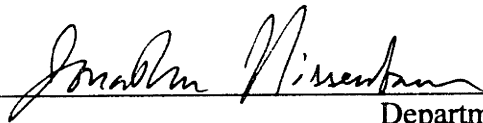
B.A., Oberlin College, 1992

SUBMITTED TO THE DEPARTMENT OF LINGUISTICS AND PHILOSOPHY IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
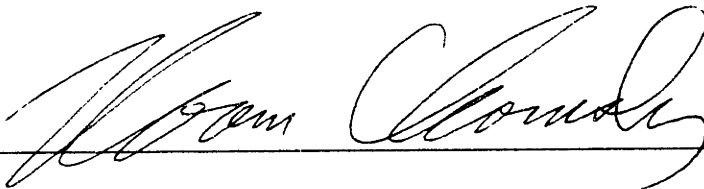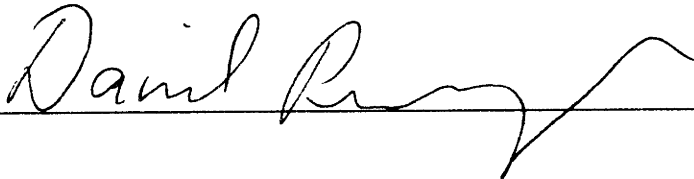
SEPTEMBER 2000

Signature of Author:_____
Department of Linguistics and Philosophy
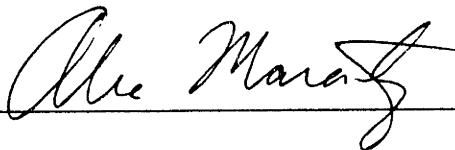September 11, 2000

Certified by:_____
Noam Chomsky
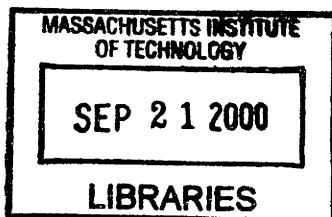Institute Professor
Thesis Supervisor

Certified by:_____
David Pesetsky
Ferrari P. Ward Professor
Thesis Supervisor

Approved by:_____
Alec Marantz
Chair, Department of Linguistics and Philosophy

Investigations of covert phrase movement

by

Jonathan W. Nissenbaum

Submitted to the Department of Linguistics and Philosophy
on September 11, 2000 in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in
Linguistics

ABSTRACT

The status of covert movement in Universal Grammar has been a perennial source of trouble in the study of language. What kinds of structures does it derive? To what extent is it similar to overt movement? What is its place in the overall architecture of the grammar?

In this thesis I present several case studies bearing on these questions, providing new evidence for the existence of covert phrase movement. These studies contribute to the growing body of evidence that grammatical conditions hold only at interface levels [Chomsky 1993]. Further, I attempt to show that, taken together, the investigations reported here lead to a model of grammar in which the interface representations are computed cyclically, by successive applications of the basic grammatical operations *merge, move* and *spellout,* on each phase of derivation.

The first studies demonstrate that covert movement licenses parasitic gaps and feeds Condition A, reversing longstanding assumptions. The apparent counterevidence that has obscured these properties of covert movement, I argue, results from a general constraint on movement (the *Tucking-in condition* [Richards 1997]) that prevents the formation of the required configurations in the classic experimental paradigms. In addition, the study on parasitic gaps provides evidence for the Y-model's sequencing of overt before covert operations. However, an investigation of adjunct extraposition from NP (a report of joint work with D. Fox) yields evidence for the opposite conclusion: that a covert operation (QR) can be followed by an overt one (late adjunction to the raised NP).

Finally, I show that these conflicting results are resolved by a theory of successive-cyclic computation of structure in which *spellout* applies repeatedly throughout a derivation. I argue that the correct characterization of the cyclic model captures Y-model effects such as the failure of covert movement (typically) to license PGs, while allowing 'anti-Y-model effects' typified by extraposition. I propose a condition that limits countercyclic adjunction to the linear edge of already computed structures. This condition in turn predicts an intricate pattern of further generalizations about extraposition. The resulting theory thus unifies the overt and covert cycles in a manner consistent with the evidence for covert phrase movement.

Thesis Supervisors: Noam Chomsky   David Pesetsky
Title: Institute Professor   Ferrari P. Ward Professor

Table of Contents

CASE STUDIES ON THE COMPUTATION OF LF REPRESENTATIONS

Chapter 1. Introduction

The status of covert movement in Universal Grammar has been a perennial source of trouble in the study of language. Claims that *wh*-in-situ and other kinds of phrases are subject to an unpronounced movement operation have faced a well-known obstacle: these ostensible movements fail to show characteristic properties of their overt counterparts (sensitivity to islands, feeding of Condition A, parasitic gap licensing). On the other hand speakers' knowledge of the meanings of sentences with quantificational elements, Antecedent-Contained Ellipsis and the like seems to yield very strong evidence that UG provides something like a covert movement mechanism.

This conflict has led to some not very satisfying hypotheses about architecture of the grammar and the place of covert movement within it. For example, the 'Y-model' incorporated a level of representation (s-structure) that feeds both a phonology and a covert syntactic component (whose operations are interpreted but do not have a phonetic reflex). The s-structure hypothesis attempted to sweep the problematic aspects of covert movement under the rug, by stipulating that the conditions from which covert movements are exempt are simply s-structure properties. But this move left the theory in a strange position. Why don't conditions which appear to be semantic in nature (like those governing co-reference and the interpretation of gaps) hold at the level of semantic interpretation (Logical Form)?

Such considerations, together with recent empirical discoveries concerning binding theory (Fox 2000a, Romero 1997), suggest the s-structure hypothesis should be abandoned in favor of a theory in which conditions hold at the LF and PF interfaces (cf. Chomsky 1993). This leaves open once again the question of covert movement. For an 'interface theory' of this sort to

be more explanatory than the s-structure theory, it will have to account in a principled way for the apparent failure of covert movement to show properties of overt movement, and it will have to say something about what distinguishes covert from overt movement. What is the nature of the structures that it derives? What is its place in the overall architecture of the grammar?

In this thesis I present several case studies bearing on these questions, providing new evidence for the existence of covert phrase movement and arguing that covert movement has more in common with overt movement than has been generally assumed. These studies contribute to the growing body of evidence that grammatical conditions hold only at interface levels. Further, I will attempt to show that, taken together, the investigations reported here lead to a model of grammar in which the interface representations are computed cyclically, by successive applications of the basic grammatical operations *merge, move* and *spellout,* on each phase of derivation.

The first studies reverse longstanding assumptions by demonstrating that covert movement licenses parasitic gaps and feeds Condition A. The apparent counterevidence that has obscured these properties of covert movement, I will argue, results from a general constraint on movement, namely the *Tucking-in condition* (Richards 1997) that prevents the formation of the required configurations in the classic experimental paradigms. In addition, the study on parasitic gaps provides evidence for the Y-model's sequencing of overt before covert operations. However, an investigation of adjunct extraposition from NP (a report of joint work with Danny Fox) yields evidence for the opposite conclusion: namely that a covert operation (QR) can be followed by an overt one (late adjunction to the raised NP).

I will argue that these conflicting results—the existence of both 'Y-model effects' and 'anti-Y-model effects'—are resolved by a theory of successive-cyclic computation of structure in

which *spellout* applies repeatedly throughout a derivation. I will argue that the correct characterization of the cyclic model captures Y-model effects such as the failure of covert movement (typically) to license PGs, while allowing 'anti-Y-model effects' typified by extraposition. I propose a condition that limits countercyclic adjunction to the linear edge of already computed structures. This condition in turn predicts an intricate pattern of further generalizations about extraposition. The resulting theory thus unifies the overt and covert cycles in a manner consistent with the evidence for covert phrase movement.

1. An argument for a vP-peripheral landing site in successive-cyclic movement (chapter 2)

As a preliminary to the investigations of covert movement and the architecture of grammar, it is important to draw some conclusions about the kinds of structures that are interpreted. In chapter two I will argue that we can learn some non-obvious lessons about covert structure from an investigation of parasitic gap (PG) constructions. Specifically, I will argue that (1) and (2) hold.

(1)   Successive-cyclic A-bar movement targets a specifier position of every vP along the way to the final landing site (in addition to every CP, a fact that is already fairly well established).

(2)   It is possible to pinpoint quite precisely the relative structural positions of these intermediate traces with respect to vP-adjoined modifiers. If the vP-trace is above a vP-adjunct, a PG appears; the absence of a PG indicates that the trace is below the adjunct.

The argument starts with evidence for an important background assumption, namely Chomsky's (1986) hypothesis that the PG is not bound by the antecedent of the 'licensing gap' upon which it is parasitic, but rather by a separate, phonetically empty operator (3).

(3)   **The PG is bound by a separate antecedent**   (Chomsky 1986)

[Which article]$_1$ did John file t_1 [O_2 without reading t_2]

The claim that PGs are null operator constructions (i.e. semantic predicates) raises a question: how does an adjunct in which null operator movement has applied compose semantically with the rest of the sentence to yield an interpretation? I will suggest that the question would have a straightforward answer if PGs arose only in the environment of *heavy-NP-shift* (HNPS), illustrated in (4)-(5). Specifically the predicate-adjunct containing the PG modifies a vP which is itself interpreted as a derived predicate of the raised DP. The two predicates yield, in effect, a conjoined interpretation akin to that of an NP modified by a relative clause.

(4)   John put on the table without reading a recent article about global warming.

(5)
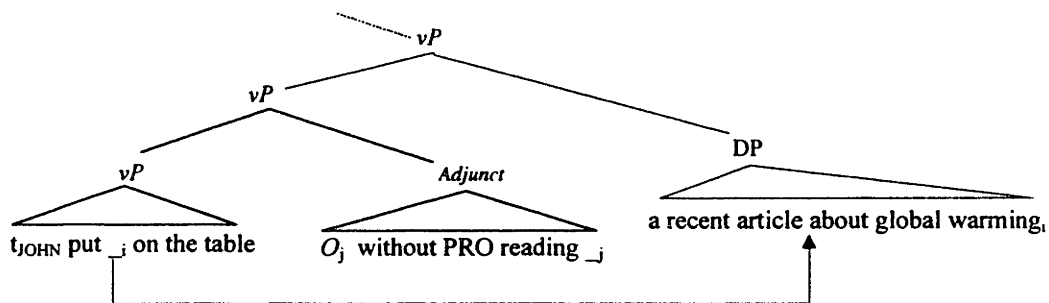
I will then argue that PGs do in fact arise only in this configuration. This entails that long-distance movement always leaves an intermediate trace in the position of HNPS (1), yielding configurations like (6). The last section of Chapter 2 will present an array of evidence that this claim is correct.

(6)

```
                    CP
                   /\
      which article  /\
                  John /\
                      /  \
                     /   vP
                    /   / \
                  vP   /   \
                 / \  /     \
               vP  Adjunct   t_WHICH ARTICLE
              / \    / \
        ...file t_WHICH ARTICLE   O_j without reading t_j
```
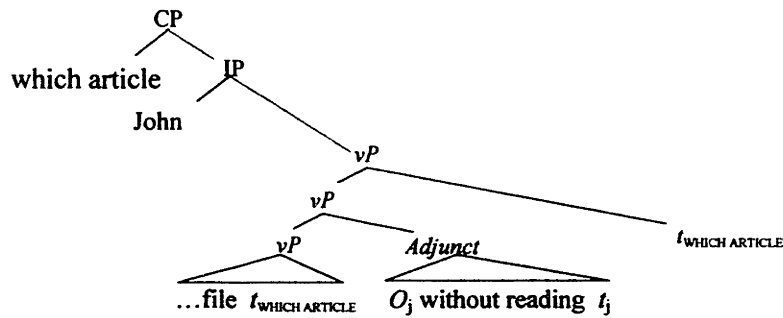
In addition to providing evidence for the presence of intermediate traces at the vP level, an important result of Chapter 2 will be the demonstration that PGs and vP-adjuncts provide a useful diagnostic for invisible structure.

2. 'Y-model effects' and 'anti-Y-model effects' (chapters 3-4)

The investigations of PGs and extraposition from NP in chapters 3 and 4 will yield an apparent contradiction that suggests neither the Y-model or its single-output alternatives is correct. The investigation of covert movement and PGs in chapter 3 will provide support for the Y-model's sequencing of overt before covert operations; the investigation of extraposition in chapter 4 will then show that the Y-model's assumption of overt and covert 'components' in the grammar is not the correct way of characterizing the sequencing restriction.

2.1 Covert movement and PGs: a 'Y-model effect' (chapter 3)

Chapter 3 will take as its point of departure a well-known generalization concerning the distribution of PGs. In one of the first studies on PGs, Elisabet Engdahl (1983) observed that *wh*-in-situ does not license PGs (7)b—unexpected if the licensing conditions on PGs hold at LF.

(7) **Engdahl's generalization**: PGs are not licensed by *wh*-in-situ

    a.   Which paper$_i$ did you [file _$_i$][without reading _]

    b.   *Who$_i$ [filed *which paper$_j$*][without reading _]

I will show that a theory in which PGs are licensed at LF by fairly standard semantic mechanisms is capable of explaining Engdahl's generalization, once we factor in a general constraint on movement that is motivated on grounds quite independent of PG constructions. The explanation will then be shown to have a very surprising empirical consequence. In a well-defined class of cases involving multiple interrogatives, Engdahl's generalization breaks down: Each of the relatively acceptable examples in (8)-(9) contains a PG that is licensed by *wh*-in-situ.

Engdahl's generalization is only partially correct:

(8) a. ?Which senator$_1$ did you persuade _$_1$ to borrow *which car$_2$* [after getting an opponent of
       _$_{pg1}$ to put a bomb in _$_{pg2}$]?

    b. * *Which senator$_1$ did you persuade _$_1$ to borrow which car$_2$ [after putting a bomb in _$_{pg2}$]*

(9) a. ?Which kid$_1$ did you give *which candy bar$_2$* to _$_1$ [without first telling a parent of _$_{pg1}$
       about the ingredients in _$_{pg2}$]?

    b. * *Which kid$_1$ did you give which candy bar$_2$ to _$_1$ [without looking at the ingredients
      in _$_{pg2}$]*

In addition, I will argue that the different behavior of overt and covert movement with respect to PG licensing (i.e. the fact that Engdahl's generalization is largely correct) supports a traditional claim about the organization of grammar: namely that the difference between overt and covert movement itself reduces to a derivational sequencing imposed by the point at which *spellout* applies (10). This specific claim was part of the model proposed in Chomsky 1993

Error! Reference source not found. and has its roots in the Y-model that was motivated by the s-structure hypothesis.[1]

(10)   **The difference between overt and covert movement is due to sequencing:**
Operations that precede *spellout* are overt, while those that follow it are covert

(11) a.   **The 'Y-model' (as revised in Chomsky 1993)**

The argument will have the following form. If it could be shown that the configuration in which PGs are licensed can only be derived by *overt* movement, we would have an explanation for Engdahl's generalization. The first step of the argument (section 1) consists of motivating a general condition on movement: namely, that second (and subsequent) movements to a single projection do not extend the tree, but rather 'tuck in' as argued by Richards (1997). A consideration of multiple *wh*-constructions in Bulgarian will illustrate the 'tucking in' property of movement. 1 will argue that this condition blocks movement from forming the structural configuration needed for licensing a PG (cf. Chapter 2), on the grounds that the vP-step of movement would be required to tuck in below the adjunct. Consequently, that configuration can

---

[1] In Chapter 4 I will give an argument (based on joint research undertaken with Danny Fox) that this picture cannot be correct; a different model of grammar that resolves the conflict will be motivated in Chapter 5. The central proposal will be to embed the claim (10) in a cyclic theory of spellout, with the result that 'Y-model effects' largely carry over. For the remainder of the discussion in this chapter I will keep to more familiar picture and continue to speak in terms of the 'Y-model.'

only be derived in the 'reverse' order, with movement targeting the vP *before* insertion of the adjunct (to the position between the mover and its target.

Crucially, the second of these two steps is necessarily overt: insertion of the adjunct feeds pronunciation. If this is the only derivation that licenses PGs, it would follow from the Y-model that covert movement is incapable of licensing PGs; a covert operation cannot be followed by an overt one.

This account makes an interesting prediction for cases of multiple extraction. I will show that an unexplored consequence of the *Tucking-in condition* is that additional movement possibilities are created when the structure provides more than one position below which a phrase could tuck in. The PG-licensing configuration provides such an environment. Once that configuration has been formed, a *subsequent* movement should be able to raise past the adjunct as long as it can tuck in below the phrase that moved first, forming the structure

$[_{vP} WH-1 [_{vP} WH-2 [_{vP} \text{adjunct} [_{vP}...t_1...t_2...]]]]$

A predicted consequence is that an adjunct that modifies the resulting derived two-place predicate will contain two PGs (one licensed by each raised XP). Since the Y-model prohibits only the first of these two movements from being covert, the account leads us to expect that a second PG can be licensed by a covert movement. These are exactly the environments exemplified by (8)-(9).

2.2 Covert movement and extraposition from NP: an 'anti-Y-model effect' (chapter 4)

The central conclusion of chapter 3—that overt and covert operations are distinguished by means of a sequencing restriction imposed by the timing of *spellout*—receives an apparent challenge in chapter 4. This chapter, based on joint research carried out with Danny Fox, argues that 'covert'

operations like Quantifier Raising (QR) can precede 'overt' operations. Specifically Fox and I argue that there are overt operations that must take the output of QR as their input.

If this argument is successful the conclusion reached in Chapter 3 must be re-evaluated; there cannot be a 'covert' (i.e. post-spellout) component of the grammar. That is, what distinguishes operations that affect phonology from those that do not cannot be a single point in the derivation (spellout) before which the former apply and after which the latter do. One alternative model, which we can call the 'single output' model of grammar, was suggested by Brody (1995), Bobaljik (1995), Pesetsky (1998), Groat and O'Neil (1994). These authors proposed that the distinguishing property has to do with principles of the syntax-phonology interface. Assume that movement is a copying operation with phonology targeting one copy in a chain for pronunciation. The distinction between 'overt' and 'covert' movement, these authors suggest, is this: 'overt' movements are the result of phonology targeting the head of a chain for pronunciation, while 'covert' movements result from phonology targeting the tail of a chain.

(12)   **The Single-Output model**
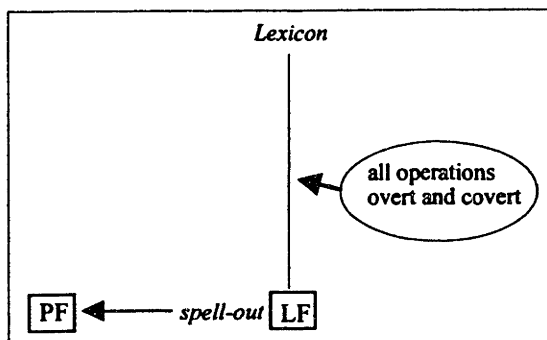
The argument that covert operations sometimes precede overt operations is based on extraposition from NP. Specifically, we argue that certain instances of extraposition result not from movement of the extraposed material but rather from QR of an NP and subsequent merger of an adjunct phrase. Phonology will determine that the NP is pronounced in its pre-QR position.

But the late-inserted NP-adjunct is not present in the pre-QR position—it can only be pronounced in the position in which it was merged into the structure. QR followed by merger of an adjunct which is "overt" is impossible if covert operations apply after spell-out.

However, while this investigation provides evidence against the Y-model, the single-output model cannot be the correct one either, given the result from chapter 3. An alternative theory that is consistent with both the evidence for 'Y-model-effects' and the evidence for 'anti-Y-model effects' is developed in chapter 5.

3. Cyclic spellout: Covert movement in a single-cycle grammar (chapter 5)

Given the untenability of both the Y-model and the single-output alternative, what distinguishes overt from covert movement? One answer which is possible in principle is that there are no covert syntactic operations that involve movement of phrasal categories. This is the answer that is assumed in Chomsky (1998,1999); a covert operation is simply an agreement relation between probe and goal, without movement of the goal. This answer would explain why such phrases are pronounced *in situ*—the *in situ* position is the head of the chain. But it would fail to account for the evidence for covert movement of phrasal categories.

An alternative possibility which I will develop in chapter 5 is that the correct theory of the unified cycle predicts the existence of Y-model effects. Specifically, the theory that we are after will have the property that it predicts just those Y-model effects that block *post-covert-movement* insertion of a vP-adjunct (which would overgenerate PGs)—but at the same time it will not rule out the cases of *post-covert-movement* insertion of NP-adjuncts that yield extraposition.

The central claims about a cyclic theory of spellout that are needed to derive Y-model effects are those stated in (13) and (14). (13) is inherited from the Y-model. But the Y-model assumed that spellout applied just once, to the entire structure. I propose that spellout applies on each cycle, to a designated sub-part of the structure, namely the internal domain (14).

(13)   **Pronunciation of chains:**
A syntactic chain is spelled out in exactly one position—the highest position possible.

(14)   *Spellout* **applies to the internal domain on each cycle**
The *spellout* property of a head H is satisfied by applying rules of phonology to the sister of H.

The difference between overt and covert movement will then reduce to the timing of the spellout operation on each cycle. If raising precedes spellout, as in (15), then the chain pronunciation algorithm (13) will determine that the chain will not be assigned a pronunciation in the position inside the domain that gets spelled out on that cycle.

(15) **Overt movement:**

    a.   Raising: $[_{HP}\, \alpha... \, H[...\alpha...]]$

    b.   Spell out the internal domain: $[_{HP}\, \alpha...H[\text{``}...\langle\alpha\rangle...\text{''}]]$ $(\langle\alpha\rangle$ unpronounced)
                                                       domain of spellout

On the other hand, if spellout precedes raising, as in (16), then (13) will determine that the chain is assigned a pronunciation in the pre-raising position, yielding 'covert' movement.

(16) **Covert movement:**

    a.   Spellout the internal domain: $[_{HP}... \, H[\text{``}...\alpha...\text{''}]]$
                                                domain of spellout

    b.   Raising: $[_{HP}\, \langle\alpha\rangle...H[\text{``}...\alpha...\text{''}]]$ $(\langle\alpha\rangle$ unpronounced)

Anti-Y-model effects arise in the cyclic model because of a 'loophole' in the definition of the cycle. Assume that the cycle is feature-driven, as stated in (17).

(17)   **Operations are driven by the head of the root projection**
Properties of the selector/attractor must be satisfied before a new selector/attractor is accessed to drive further operations.

An important consequence follows from (17). Merge of adjuncts is not governed by the cycle, since this operation is not 'triggered' (i.e. not driven by selection or attraction). Therefore, merge after the spellout operation has applied on a given cycle is in principle free; the late-merged adjunct will be incorporated into the phonological representation on the subsequent spellout.

In sum, the cyclic theory of spellout predicts that movement and argument merge will be subject to Y-model effects. These feature driven operations are forced to apply cyclically by the end of a phase; operations that apply before spellout applies to the internal domain of the phase will have phonological consequences while those that apply after spellout will not. After the phase is complete its head is inert, meaning its head cannot trigger further operations.

Non-feature driven operations—i.e. merge of adjuncts—are predicted to be immune from Y-model effects. These operations are not triggered by any property of the phase or its head. Therefore the cycle does not itself prevent adjuncts from merging indefinitely low inside previously computed syntactic objects. This consequence is desirable given the evidence from extraposition that countercyclic merge is possible (post-QR adjunction of the relative clause). However, the result is that anti-Y-model effects are overgenerated under this theory: Covert movement is apparently predicted to license PGs, by late merge of the adjunct following the movement.

I will suggest that the important distinction between these two cases lies in the fact that late merge would be forced to break up a previously computed linear sequence in order to license

a PG. The argument had two parts. First I will suggest that merge after spellout is prohibited *during* a phase. A requirement that spellout cannot apply on a phase until the array is exhausted dictates that nothing will be left to merge after spellout:

(18)   Apply spellout only if the array is exhausted.

As a consequence, anti-Y-model effects can only arise by adjunct merge on a subsequent phase, into an already constructed syntactic object that is accessed in the new array. I will argue that merge into a syntactic object is constrained to apply at its linear edge:

(19)   **Linear Edge Generalization:**
         Merge of new material is possible inside a phase P after P is complete, but only at the linear edge.

This condition will be shown to hold more generally, constraining word-formation as well. The linear edge generalization bars countercyclic derivation of PGs, but allows extraposition. Finally, I will show that even extraposition is tightly constrained. When extraposition would violate the linear edge generalization, it is impossible. The evidence will be drawn from cases of multiple extraposition to the same vP, extraposition from *wh*-in-situ and Antecedent-contained deletion (ACD) constructions.

Chapter 2.   What *Parasitic Gaps* can tell us about the hidden structure of chains

The goal of this chapter is to show how it is possible to use the *parasitic gap* (PG) construction to discover non-obvious properties of sentence structure. Isolation of specific kinds of structure that are signaled by the presence or absence of PGs will allow us to make inferences in the next chapter about principles involved in deriving those structures, based on peculiarities in the distribution of PGs.

Here I will argue, specifically, that a close study of PGs reveals two important facts about the structures created by phrase movement:

(1)   Successive-cyclic A-bar movement targets a specifier position of every vP along the way to the final landing site (in addition to every CP, a fact that is already fairly well established).

(2)   It is possible to pinpoint quite precisely the relative structural positions of these intermediate traces with respect to vP-adjoined modifiers. If the vP-trace is above a vP-adjunct, a PG appears; the absence of a PG indicates that the trace is below the adjunct.

These conclusions will play an important role in subsequent chapters, in particular serving as a foundation for the investigation of covert movement in chapter three.

The line of argumentation will be as follows. First, I will summarize several puzzling properties of PGs that have largely guided research on the topic. I will then review evidence for two important background assumptions that I believe can be established with reasonable certainty (although they have been challenged in the literature): Chomsky's (1986) null operator hypothesis, and the assumption that the 'sentential' adjuncts that typically host PGs are right-adjoined modifiers to vP. In the second section, I will try to show that there is nothing special

about the PG construction—i.e. no special mechanisms or licensing conditions are involved. I will argue that the construction is licensed instead by ordinary semantic composition principles that are used in the interpretation of other 'null operator structures' (relative clauses being the canonical case). The argument will start with a demonstration that PGs are a predictable consequence of the structure that is formed by heavy-NP-shift (HNPS). I will then show that nothing further need be said if we assume that this is how PGs *always* arise. Even in long-distance movement the appearance of PGs is expected if an intermediate trace is always in the structural position of HNPS. This will constitute the first argument for the vP-level successive-cyclic position (claim (1)): it allows a simple interpretation of PGs with no stipulated mechanisms. Next, I will briefly review independent evidence for the vP landing site, discovered by Fox (1999,2000a), from correlations between binding theory and scope reconstruction. Finally, I will present an array of evidence for the claim that PG constructions provide a diagnostic for the precise position of this intermediate copy (claim (2)). The evidence will come from surprising asymmetries that arise with stacked vP-adjuncts, extraposition and ACD.

1   Background: the syntactic properties of Parasitic Gaps

The PG construction displays two very puzzling properties, illustrated by the examples in (3)-(4). The first property is the presence of a gap inside a domain that is ordinarily an island for extraction (CED), made relatively acceptable by the addition of a gap *outside* the island. The parasitic gap (inside the island) and the 'licensing' gap (outside the island) are both apparently bound by the same antecedent. This property is illustrated by the contrasts in (3).[1] The second

---

[1] In these and subsequent examples, the PGs are distinguished by means of bracketing around the islands that contain them, as in ...[..._...].... Where necessary, the brackets will be in bold type to distinguish the island from other bracketed material (as in (4)).

puzzling property is the fact that the 'licensing gap' cannot c-command the parasitic gap, as

shown in (4). These anomalous properties have been the chief concern of research on the topic

since the earliest studies in the late 1970s.

(3) a. **Parasitic gaps inside adjunct islands**

Which article did John file _ [before reading _]?
??Which article did John file his papers [before reading _]

What movies did Mary claim she liked _ [in order to get you to see _]?
??What movies did Mary claim she liked The Godfather [in order to get you to see _]

John's the guy that they said they'll hire _ [if I criticize _ publicly]
* John's the guy that they said they'll hire me [if I criticize _ publicly]

b. **Parasitic gaps inside Subject islands**

Who do [friends of _] often end up hating _ ?
* Who do [friends of _] often end up hating you?

Which judge did [my talking about _] offend _ ?
* Which judge did [my talking about _] offend you

Mary is one person that [everyone who talks to _] becomes friends with _
* Mary is one person that [everyone who talks to _] becomes famous

(4) **The 'licensing gap' cannot c-command the PG**

a. the person who I [claimed _ was lonely] [in order to get you to visit _]

b. *the person who _ [claimed I was lonely] [in order to get you to visit _]

The examples in (3)-(4) all involve *wh*-extraction/relativization, but PGs are not limited to these

environments—a variety of movement constructions give rise to them. A partial sample is given

in (5)-(7), including topicalization (5)a, tough-movement (5)b, and comparative deletion (5)c. In

addition to the unbounded A-bar movement environments of (5), PGs are also licensed by *heavy-*

NP-shift (6). In languages that allow it, scrambling of an NP to the left of an adverbial licenses a

PG, illustrated for German in (7).

(5)   **PGs in other A-bar movement environments:**

a.   Fred, I talked to _ [in order to impress _]

b.   Mary is easy to talk to _ [without offending _]

c.   Cool Hand Luke ate more eggs than they were able to boil _ [without breaking _]

(6)   **PGs licensed by *Heavy-NP-Shift* (HNPS):**

John filed _ [without reading _] a recent article about global warming.

(7)   **PGs licensed by scrambling:**   (German example due to Martin Hackl, p.c.)

...weil der Hans das Formular [ohne _ vorher auszufüllen] _ abgeschrieben hat
because the H.   the form   [without _ first to-fill-out] _ copied   has
'...because Hans copied the form without filling (it) out first'

cf.   ...weil der Hans [ohne *(es) vorher auszufüllen] das Formular abgeschrieben hat
                              ...*(it)...

Besides the properties just illustrated, PGs differ in an important respect from other structures involving what is apparently a single binder of multiple gaps. The *across-the-board* (ATB) movement dependencies in (8), for example, require simultaneous extraction from two independent clauses; raising out of either one of them alone is impossible as shown in (8)b and (8)c. This is not true of the environments in which PGs appear. As (9) shows, a single gap may be independently licensed in the main clause but not in the adjunct.

(8) a.   an article that John read _ and Mary glance at _

   b.   * an article that John read _ and Mary glance at it

   c.   * an article that John glance at it and Mary read _

(9) a.   an article that John read _ [before Mary glanced at _]

   b.   an article that John read _ [before Mary glanced at it]

   c.   * an article that John read it [before Mary glanced at _]

On the other hand there can be little doubt that PGs arise from movement of some kind, given the discovery by Kayne (1983) and Chomsky (1986) that island violations re-appear *internal* to the CED islands containing the PG:

(10)   **Island effects re-appear inside the parasitic domain:**

    a. Who did John visit _ [without claiming that he knew _]

    b.   *wh-island*   ??[without asking [whether I knew _]]

    c.   *relative clause*   * [without consulting [the person who'd talked to _]]

    d.   *CNP*   * [without denying [the claim that he disliked _]]

    e.   *CNP (infinitive)*   ??[without announcing [the plan to hire _]]

    f.   *CED: subject*   * [before [friends of _] could talk him out of it]

    g.   *CED: adjunct*   ??[after offending me [by not introducing me to _]]

The right generalization thus seems to be the following: exactly one island boundary can appear between the PG and the apparent antecedent. Otherwise standard conditions on movement apply. The latter fact would be difficult to explain if PGs were not the result of a movement operation.

1.1   Two kinds of theories
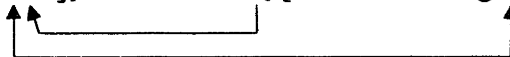
While it's clear from (8)-(9) that PG environments are not subject to an ATB movement condition, more sophisticated proposals have been offered which do not postulate such a condition. The proposals fall into two basic categories, schematically illustrated in (11)-(12). The first category has a number of quite different variants. All of them accept (in one form or another) the premise that the relationship between the PG and the overt antecedent is what needs to be explained. More specifically, the proposals in this category have in common the view that the dependency involving the PG crosses the island without inducing a violation; i.e. the dependency is blind to the island separating the PG from the antecedent. Implicit in these

proposals is an assumption that both gaps (parasitic and main) are bound by the same antecedent. We can call this class of proposals the 'shared antecedent' viewpoint. The other category rejects the premise that there is a relationship between the PG and the overt antecedent that needs to be explained. This view posits that, contrary to appearance, no island intervenes between the PG and its antecedent—rather the antecedent for the PG is distinct from the antecedent of the licensing gap, and internal to the island. We can call the proposals in this category the 'separate antecedent' viewpoint.

(11)   **Category 1: 'Shared antecedent' proposals**

(Taraldsen 1981, Chomsky 1982, Sag 1983, Kayne 1983, Longobardi 1984, Pesetsky 1982, Brody 1995, Nunes 1995, Richards 1997, Steedman 1997)
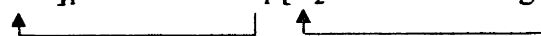
[Which article]$_1$ did John file t_1 [without reading PG_1]

(12)   **Category 2: 'Separate antecedent' proposals**

(Chomsky 1986, Postal 1998)

[Which article]$_1$ did John file t_1 [O_2 without reading t_2]

The 'Shared antecedent' proposals of category 1 divide into several variants. An early version, defended by Taraldsen 1981 and Chomsky 1982, holds that the PG is a base-generated (unmoved) empty pronominal. While this view is consistent with the facts in (3)-(9), it is incompatible with Chomsky's later discovery of facts like (10).

Three other important early contributions in the 'shared antecedent' class were those of Kayne (1983), Longobardi (1984) and Pesetsky (1982), which showed that the *position* of the island plays a crucial role in licensing the PG. Longobardi, extending observations of Kayne, demonstrated that an island containing a PG must be positioned 'along the path' between a

licensed movement and its antecedent (i.e. it must c-command a licensed gap).[2] This generalization yields a stronger version of the anti-c-command requirement that was illustrated in (4). Namely, not only must the primary gap not c-command the PG; it must be c-commanded by the (smallest) island containing the PG:

(13)   **The Kayne-Longobardi generalization:**
        An island containing a PG must c-command a licensed gap

    a.   a person who I hang out with _ [because I admire friends of _]

    b. * a person who I hang out with _ [because [friends of _] are famous]

    c.   a person who I hang out with _ [because [friends of _] admire _]

In (13)b, the gap in the main clause is c-commanded by the adjunct. However in this example the PG is also contained in a subject island that does *not* c-command (or dominate) the primary gap. Therefore the Kayne-Longobardi generalization is violated. Note that a 'licensed' PG (such as the gap after *admire* in (13)c) can in turn serve to license an *additional* PG inside another island (e.g. the gap inside the embedded subject in (13)c), so long as that island obeys the Kayne-Longobardi generalization. Other proposals similar to Kayne's (but using quite different theoretical frameworks) include those of Sag (1983) and Richards (1997) and Steedman (1997). Common to these studies are (i) the assumption that movement (or a movement-like relation) holds between the PG and the position of the overt antecedent, and (ii) the claim that the island is obviated because it falls along the path of a licensed movement.[3]

---

[2] Kayne originally stated the generalization in such a way as to limit its scope to subject islands. As Longobardi noted in his (1984) extension of Kayne's theory, however, the more general statement seems to be correct.

[3] Among these 'path-based' proposals, only those of Steedman and Sag (as far as I can tell) derives the crucial claim from general principles. Sag's proposal (which was further developed in Gazdar, Klein, Pullum and Sag 1985) makes use of the framework of assumptions of *Generalized Phrase Structure Grammar* (and has been adopted in subsequent work in *Head-Driven Phrase Structure Grammar*), which characterizes the movement relation by means of Slash-categories. Steedman's work in Categorial Grammar similarly characterizes the movement relation as information-projection. In effect, the existence of the non-violating gap in the main VP 'percolates up' through the

Yet another type of proposal in the 'shared antecedent' camp, due to Nunes 1995 and adopted with modifications by Hornstein 1998, holds that the PG and the trace of the licensing movement both belong to the same chain. The obviation of the island arises due to the way the sentence is claimed to be derived: the NP *which article* first merges in its position in the adjunct, then re-merges to the direct object position (the position of t_1 in (11)) in the main VP *prior* to the merger of the adjunct island to the main clause.

The other class of proposals, which I have termed the 'Separate antecedent' class, also assumes that PGs arise by movement. However, under this alternative view, the appearance of a shared antecedent is misleading, and so is the apparent island obviation. There is no dependency between the PG and the antecedent of the licensing movement because the PG is the trace of a separate movement that doesn't cross the island boundary. This proposal was first made by Chomsky (1986), who argued that the PG construction is an instance of *null operator* movement of the type that is found in relative clauses and has been argued to characterize (among others) *tough*-movement (Chomsky 1977) and purpose clauses (Browning 1987). Specifically, Chomsky 1986 argued that the PG arises from movement of a separate, phonetically empty, operator to a position that takes scope just over the island (as in (12)).[4] Under somewhat different assumptions, Postal 1998 proposed a similar idea: he argued that PGs belong to a class of extractions that involve movement of a null resumptive pronoun to a position with scope over an island (1998: 2.4.3).

---

tree to the antecedent. Thus, on GPSG/HPSG and Categorial Grammar assumptions, an adjunct with a gap of the same category is able to compose semantically with nodes along the path from the main extraction site to the node dominating the antecedent. The facts I will discuss in section 3 and those in the first appendix to this chapter constitute empirical arguments against these proposals (and against 'path-based' theories in general).

[4] Chomsky's proposal was actually that the null operator scopes over a CP-complement of the head of the island—i.e. below the preposition *without* in (12). This was based on theory-internal considerations about available landing sites; I will argue below that the operator targets the island as I have represented it in (12).

Of course, while Chomsky's *null operator* hypothesis (and Postal's variant) accounts for the absence of island effects in PG constructions, it opens up a new question: how the sentences receive the interpretation that they do. A theory that explicitly denies the relation between the PG and the apparent antecedent must nonetheless account for the fact that the PG (largely) behaves like a variable bound to that antecedent. Moreover, other ostensible instances of null operator constructions have a more-or-less straightforward compositional semantics in terms of predication: null operator movement turns a constituent with a clausal meaning into a semantic predicate, relative clauses being the canonical case. Chomsky's proposal that PGs involve null operator movement—and much of the subsequent research that has adopted it—does not in itself answer the question of how the construction receives the right (or any) interpretation.[5]

One interesting study that is exceptional in this regard is Larson 1988a. Larson showed how Chomsky's proposal in fact does allow PG constructions to fall together with other null operator structures, arguing that the structures formed by null operator movement are VP-internal predicates which compose semantically with the verb to form a complex predicate of the direct object. I will argue in section 2 that something like Larson's answer to the interpretability question is correct. However I will try to demonstrate that Larson's specific proposals about the syntax of HNPS and vP-adjuncts cannot be correct. Instead, I will provide some evidence for a more conventional set of assumptions and show that Larson's insight can be maintained under those assumptions. The simplicity of that explanation, in my view, constitutes sufficient reason to accept Chomsky's null operator hypothesis.

---

[5] Chomsky suggested that a special interpretive mechanism is involved, allowing two chains in the appropriate configuration to be interpreted by a rule of *Chain Composition*. But this new rule begs the question; it is construction-specific and moreover it is not clear that the empirical evidence for the null operator (reviewed in section 1.2) is actually predicted if the two chains receive a 'composed' interpretation.

However, since there are alternative proposals of the 'shared antecedent' category, I will first briefly review some independent evidence that argues against such proposals and in favor of the null operator thesis.

1.2 Arguments for the *Null Operator* hypothesis

As already noted, it is by now little in doubt that PGs are gaps left by movement (or its equivalent in theoretical frameworks that characterize the movement relation differently), rather than the unmoved empty pronouns postulated by Chomsky 1982. But we would still like to find empirical predictions that distinguish the null operator thesis from the various 'shared antecedent' theories that also assume movement. If we could show that there are environments in which the two gaps (the PG and the main-clause gap) behave as though they are bound by separate antecedents, we would have a very strong argument for the null operator hypothesis. I will argue below that various kinds of reconstruction phenomena create just such environments.[6]

The evidence below will be presented in the following manner. In order to show how the null operator hypothesis makes empirically different predictions from the various proposals of the 'shared antecedent' class, we need to consider *two* kinds of alternative predictions. The path-based theories of Kayne 1983, Pesetsky 1982 and Sag 1983/Gazdar et al. 1985, assume that the relation holding between the PG, the main gap and the ostensibly shared antecedent is essentially (a complicated sort of) ATB movement.[7] The 'sideward movement' theory of Nunes

---

[6] I will follow recent convention and use the term 'reconstruction' as a descriptive term for the phenomenon; I will assume here that reconstruction effects are a consequence of the copy theory of movement, as has been argued by Chomsky 1993, Fox 2000a and Romero 1997.

[7] It is possible in principle to restate some path-based theories in terms of movement of a separate (null) operator *to the position of the overt antecedent.* (Such a restatement is not possible for the GPSG/HPSG and Categorial Grammar theories, which do not regard movement as relating discrete terms of a chain.) These theories would then predict the asymmetries discussed in this section, without giving up the central idea that movement from

1995/Hornstein 1998 assumes that the relation holding between the PG, the main gap and the antecedent is essentially successive-cyclic movement. So for each of the arguments discussed below, I will compare the predictions of the null operator hypothesis with predictions made both by the assumption of ATB movement and by the assumption of successive-cyclic movement.

Reconstruction asymmetry 1:   Condition A

If both the parasitic and the licensing gap are bound by a single antecedent, then both gaps ought to provide potential reconstruction sites—and reconstruction effects ought to hold symmetrically with respect to both gaps. But as Kearny (1983)[8] showed, Condition A reconstruction effects are asymmetric in PG constructions: reconstruction is impossible to the position of the PG:

(14)   **Asymmetric pattern of reconstruction:**

    a.   Which pictures of himself did John sell before Mary had a chance to look at?

    b.   * Which pictures of himself did Mary sell before John had a chance to look at?

This pattern contrasts with that found in standard ATB movement (15) as well as with the pattern found in successive-cyclic movement through intermediate positions (16).[9]

---

the parasitic domain is allowed to escape the island. However, one would want to know why movement of a separate null operator would be necessary, given the assumed insensitivity to islands. Furthermore I believe that the challenge of showing how such structures would receive the appropriate interpretations would be difficult to meet. In any case I will continue to conflate the property of "shared antecedenthood" with that of "ability to escape the parasitic island," noting here that they should be regarded as distinct.

   [8] Cited in Chomsky 1986.

   [9] David Pesetsky (p.c.) informs me that many speakers find a fairly strong contrast between examples like (15)a and (15)b, parallel to the contrast found with PGs in cases like (14) (with reconstruction permitted to the gap on the left). While I do not find (much of) a contrast in (15), such contrasts are reported in work by Munn (1992, forthcoming). To the extent that reconstruction is asymmetric in (15)—and particularly for speakers who find the asymmetry to be as strong as the asymmetry for PGs—this pattern of judgments supports Munn's proposal that null operator movement is involved in (15) as well as in (14). If so, (15) does not actually represent a case of ATB movement; the argument here therefore remains sound. It is not obvious how the other facts in this section could be explained if standard ATB constructions reduce to null operator movement, but see Munn (forthcoming).

(15)   **ATB movement: Reconstruction must be** *across-the-board*

    a. * Which pictures of himself did John sell and Mary buy?

    b. * Which pictures of himself did Mary sell and John buy?

(16)   *Successive-cyclic* **movement: Reconstruction has multiple options**

    a. Which picture of himself was John upset that Mary filed (before she had a chance to look at)?

    b. Which picture of himself was Mary upset that John filed (before she had a chance to look at)?

The patterns of reconstruction in (15) and (16) are both symmetric. A reflexive contained in an ATB-moved *wh*-phrase needs to be licensed in both extraction sites, the expected pattern when one antecedent binds both gaps. Similarly, when intermediate positions are targeted by successive-cyclic raising (16), a reflexive can be bound under reconstruction to any position in the chain, as has been known since Barss (1986).

The reconstruction asymmetry in the PG construction can also be seen in examples like (17), where the PGs are inside subject NPs (and hence to the left of the licensing gap):

(17) a. ? Which portrait of himself do people who see for the first time usually think Picasso spent the most time on?

    b. * Which portrait of herself do people who know Gertrude Stein hated usually think Picasso spent a lot of time on?

    (cf. Which portrait of herself do people who know Gertrude Stein hated usually think she should have liked?)

While (17)a seems to me to have a more marginal status than (16)b and (14)a, it is clear that this example differs from the quite sharp violation of Condition A in (17)b.

The conclusion from (14)-(17) seems to be that the asymmetry results from some property of the construction itself: reconstruction of the overt *wh*-phrase is in principle possible

only into the position of the main gap, and not to the position of the parasitic gap. This is the predicted result given Chomsky's null operator hypothesis. Since the PG is not part of the same chain as the overt *wh*-phrase, it is not a possible site for reconstruction:

(18)   [Which picture of himself]$_1$ did John sell t_1 [O_2 before Mary had a chance to look at t_2]

The asymmetric pattern of these examples is, on the other hand, unexpected under the alternative theories that assume the overt *wh*-phrase is the antecedent for the PG. If we assume that PGs result from a kind of ATB movement we would wrongly predict both examples in (14) to be unacceptable. The alternatives of the Nunes 1995/Hornstein 1998 variety apparently predict that (14) should have the successive-cyclic pattern of (16).[10]

Reconstruction asymmetry #2: Condition C

The same result can be seen again in cases of reconstruction-induced violations of Condition C. The picture-NP contained in the *wh*-phrases in (19) contains an r-expression (*John*), and hence should induce a Condition C violation if the *wh*-trace is c-commanded by a pronoun co-referent with *John*. In (19)a, where the main gap is c-commanded by a pronoun, a Condition C effect holds, blocking co-reference. By contrast, in (19)b only the parasitic gap is c-commanded by the pronoun, and co-reference is allowed.

---

[10] We might try to supplement the Nunes/Hornstein theory with an assumption (not implausible) that the two separate chains formed by the peculiar successive-cyclic movement each have to satisfy a well-formedness condition requiring the reflexive to be bound in some position. But even with an assumption of this sort, the Nunes/Hornstein theory does not predict the attested pattern—it then strongly predicts ATB reconstruction. The same observation will apply below to variable-binding reconstruction effects.

(19)   **No anti-reconstruction to the position of the PG**

    a.   *Which picture of John$_i$ did he$_i$ buy without letting Mary look at?

    b.   Which picture of John$_i$ did Mary buy without letting him$_i$ look at?

Again the unexpected pattern is unique to PGs. In both ATB-movement (20) and successive-cyclic raising constructions (21), anti-reconstruction phenomena like Condition C are created in any position in which a lower copy is c-commanded by a co-referent pronoun.

(20)   **ATB movement: Anti-reconstruction in both conjuncts**

    a.   *Which picture of John$_i$ did he$_i$ buy and not let Mary look at?

    b.   *Which picture of John$_i$ did Mary buy and not let him$_i$ look at?

(21)   **Successive-cyclic movement: Anti-reconstruction all the way down**

    a.   *Which picture of John$_i$ was he$_i$ upset that Mary sold?

    b.   *Which picture of John$_i$ was Mary upset that he$_i$ sold?

Like the pattern of Condition A reconstruction, the asymmetric behavior of PG constructions with respect to Condition C is a consequence of the null operator hypothesis, but surprising under alternative theories.

Reconstruction asymmetry #3:   variable binding

A similar pattern of results can be found with variable binding reconstruction effects, as argued in Nissenbaum 1998b.[11] The contrast in (22) shows that a variable contained in a *wh*-phrase can be bound by a quantifier under reconstruction to the position of a gap in the main

---

[11] Thanks to Danny Fox for suggesting this test and for very helpful discussion about the examples.

clause, but not under reconstruction to the PG. (23) shows that even ATB reconstruction is impossible: the variable in the overt antecedent simply cannot be bound in the island.[12]

(22)   **No variable-binding reconstruction into the position of the PG**

   a.   Which picture of his mother did you give to every Italian after buying from Mary?
        (Possible answer: "I gave every Italian the picture of his mother that he liked best after buying it from Mary.")

   b. # Which picture of his mother did you give to Mary after buying from every Frenchman?

(23)   **No ATB binding of PGs (a sub-case of (22)):**

   # Which picture of his mother did you give to every Italian after buying from every Frenchman?
   (IMPOSSIBLE answer: "I gave every Italian the picture of his mother that he liked best after buying from every Frenchman the picture of his mother that HE liked best.")

As shown in (24)a, ATB binding is possible in a standard ATB movement structure. That is, (24)a has an interpretation in which different pictures are being received by every Italian and promised to every Frenchman. And (24)b,c show that not only is ATB reconstruction possible in this environment, it is obligatory.

(24)   **ATB movement requires ATB binding:**

   a.   Which picture of his mother did you give to every Italian and sell to every Frenchman?
        (POSSIBLE answer: "I gave to every Italian and sold to every Frenchman the picture of his mother that he liked best.")

   b. #?Which picture of his mother did you give to every Italian and sell to Mary?

   c. #?Which picture of his mother did you give to Mary and sell to every Italian?

---

[12] Example (22)a raises interesting questions concerning interpretation. The indicated interpretation suggests that QR may be involved. If so, the contrast in (22) is irrelevant for the argument. However, the contrast between (23) and (24) remains relevant.

'Shared antecedent' theories do not give us any reason to expect a difference between (22)-(23) and (24), since they assume that both gaps are the result of movement to the position of the overt antecedent. The null operator hypothesis, on the other hand, predicts the pattern of (22)-(23) since there is no copy of the *wh*-phrase in the adjunct island.

The Nunes/Hornstein theory likewise seems to make an incorrect prediction: since both gaps are claimed to be copies left by successive movements of the overt *wh*-phrase, we would expect[13] to find the pattern of reconstruction possibilities in (25), which allows variable binding by any quantifier that c-commands one of the copies.

(25)   **Successive-cyclic raising allows variable-binding reconstruction in all positions**

    a.   Which picture of his mother did no boy think that every girl would like?
        (Possible answer: "No boy thought every girl would like the picture of his mother that I took")

    b.   Which picture of his mother did no girl think that every boy would like?
        (Possible answer: "No girl thought every boy would like the picture of his mother that I took")

1.3   Right-branching vs. adjunction to vP

Assuming that we are on relatively solid ground in accepting Chomsky's null operator hypothesis, it is still necessary to review the evidence bearing on *where* in the clause these null operator islands are located. With regard to subjects that contain PGs, I will take their position to be an established matter: their surface position is the specifier of an inflectional category high in the clause (I will assume TP for concreteness, although nothing hinges on that decision). I will assume (crucially—see the appendix to chapter 2) that the surface position of subjects is derived

---

[13] With the qualification noted in footnote 10.

from a thematic position in spec-vP (cf. Kitagawa 1986, Koopman and Sportiche 1991, Kuroda 1988, Speas 1986 and much subsequent work).

As for the kinds of adjuncts that are able to contain PGs, their position has been more contentious. I will assume that they are adjoined to vP, as illustrated in (26). This subsection will review evidence for and against the assumption, and attempt to make the case against a lower attachment site (i.e. internal to VP).

(26)

```
                CP
               /\
              /  \
    which book     IP
               \  /\
                John \
                      \
                       vP
                      / \
                     /   \
                   vP     Adjunct
                  / \    / \
                 /   \  /   \
            ...bought t_i  O_j without reading t_j
```
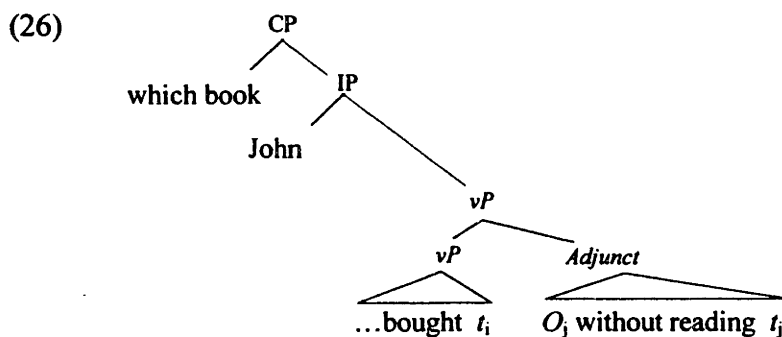
The assumption that adverbial clauses may be adjoined to vP appears to be at odds with Kayne's (1995) Linear Correspondence Axiom[14] and subsequent related work which assumes that there are no right-hand adjuncts. But this appearance may be superficial: it could in principle be the case that non-right-branching structures like (26) are derived from underlying right-branching structures by a series of leftward movements (the kind of derivation that Kayne 1998 argued to be prevalent). If good reasons are discovered for believing that (apparent) non-right-branching structures have such a derivation, then the conclusions reached in this chapter would not necessarily present counterevidence. What this section will argue for is that that PG adjuncts are

---

[14] At least under Kayne's formulation of the LCA. But one can imagine alternative formulations of Kayne's theory which allow a distinction between adjuncts and specifiers, and which require an unambiguous linearization only for specifiers, heads and complements. (c.f. K. Johnson 1997, esp. pp. 36-38.)

not vP-internal (i.e., not complements) *at the level of representation at which binding conditions, scope relations and the interpretation of PGs are relevant.*

The kinds of adjuncts that typically host PGs are clausal: those headed by temporal prepositions *(before, after, while),* rationale clauses *(in order to..., because...),* *if*-clauses, and participial clauses headed by *without* and *by.* One argument that these are not complements in a Larsonian right-branching structure is provided by constituency tests of the kind used by Pesetsky (1995) to argue for "layered" representations. The tests in (27) appear to argue strongly for a structure like (26). The *without* clause must be adjoined higher than the vP-internal material to its left, on the assumption that only syntactic constituents can be coordinated (27)a, deleted (27)b and fronted (27)c. The remnant must be external to the constituent:

(27) a. John [filed the papers and shelved the books] without reading them

b. John filed the papers without reading them, and Peter did _ without even looking at them.

c. ...and file the paper, John did, without even bothering to read it.

Unfortunately, the tests in (27) are not reliable on their own, since (as Pesetsky and others have shown) they sometimes conflict with c-command tests of the type used by Larson (1988b) (following Barss and Lasnik 1986) to argue that double object VPs and some adverbials are right-branching structures. However, in the cases that are of concern to us (namely, clausal adjuncts of the kind that can contain PGs), even the Larsonian c-command tests do not systematically support a right-branching structure.

Condition C of the binding theory provides one such test for c-command relations. The sentences in (28) are typical of those argued by Larson (1988b) to have right-branching VPs. The contrast between (28)a and (28)b suggests that (28)a is a violation of Condition C, with the r-

expression *John* c-commanded by a co-referring pronoun. (28)c is a control showing that in a sentence similar to (28)a, with the pronoun embedded so that it does not c-command the r-expression, the violation disappears.

(28) a.   * We gave him a book for John's birthday

   b.   We gave John a book for his birthday

   c.   We gave his favorite charity a donation for John's birthday

(29) a.   We gave him a book in order to impress John's mother

   b.   We gave him a book because John's mother couldn't afford one

   c.   We gave him a book before/after/without talking to John's mother

The sentences in (29) differ markedly from (28)a.[15] They seem to be on a par with (28)c, in which the pronoun also appears to the left of the r-expression but does not c-command it. Since there is no Condition C violation, we are entitled to conclude that the structure is not right-branching.

Another test that is frequently used to show c-command relations is variable-binding. It is an elementary fact about the semantics of variables that if a variable is bound in a structure, the quantifier must c-command it. However, this test does not unambiguously diagnose c-command

---

[15] Examples similar to those in (29) have sometimes been reported in the literature (cf. Thompson 1996) as yielding Condition C effects, contrary to the judgment claimed in the text. (Cf. also Reinhart 1983, who reports judgments that accord with mine on such examples.) Several points seem in order. First, in cases like (29) where an r-expression follows a co-referent pronoun, the prosodic characteristics of the sentence seem to make a difference — the r-expression needs to be de-accented (or at least it becomes much worse with any prosodic prominence). That is why the examples in (29) (as well as Reinhart's) embed the r-expression in a possessive NP, to prevent nuclear stress from falling on the r-expression. (Note that real Condition C violations appear even when the r-expression is de-accented: *He talked to John's MOTHER. Second, the important fact here is the contrast between (29) and (28)a. The r-expression in (28)a is similarly embedded, but the Condition C effect remains even if great care is taken to de-accent *John*. Finally, it should be noted that Pesetsky's (1995) arguments for a right-branching "cascade" representation that is computed in parallel with the layered structure noted above, were made on the basis of adverbials other than those under consideration (*for*-adverbials as in (28), as well as *in, on, about, to,* and the like. In fact, Pesetsky (p. 181) leaves open the possibility that some adverbials are in fact outside the VP and therefore not part of a cascade-structured VP (based on the Reinhart examples similar to my (29)).

at *s-structure*. As long as syntactic scope-changing operations like QR are available for certain kinds of phrases, the ability to bind a variable is an unreliable test for s-structure c-command relations. And in fact, the results of the variable binding test appear to conflict with those from the Condition C test, as can be seen from the examples given in (30)a, all perfectly acceptable.

(30) a.   John read every/no book thoroughly in order to memorize it
                                without managing to understand it
                                before criticizing it
                                after learning its author had died

   b.   John [every book]$_i$ [$_{VP}$ [read t$_i$ thoroughly][before criticizing it$_i$]]

Clearly what must be said to reconcile these apparently conflicting results is that the Condition C test accurately diagnoses the s-structure relations, and the variable-binding of (30)a merely indicates that QR has raised the quantifier phrases to a position from which it can bind the variable in the adjunct, as shown in (30)b. If so, the question of Weak Crossover arises: why doesn't the structure in (30)b create a WCO violation? It should be noted, however, that the constraint against WCO still lacks a satisfactory explanation. The facts in (30)a are consistent with a statement of the WCO generalization as a prohibition against QR that raises a quantifier to the left of a variable (cf. Chomsky's (1973) "Leftness Condition").

In support of this conjecture about the irrelevance of WCO, consider (30)c,d below. Assuming that the in-situ wh-phrase and QP undergo covert raising, (30)c-d should also be WCO violations. The rationale clause in (30)c is unambiguously associated with the higher clause, psych predicates being incompatible with rationales (cf. *#I enjoyed that movie in order to impress my friends*). And under the most natural reading of (30)d the rationale clause similarly modifies the embedding vP.

(30) c.   Who [said that he'd enjoyed what (movie)][in order to get you to see it]?

     d.   Which boy did you [refuse to introduce to every girl][in order to make her angry]?

The fact that these examples are fully acceptable, yielding no WCO violation, thus bolsters the argument that QR has similarly applied in (30)a and that WCO is irrelevant.

The final test that can be marshaled to support a non-low attachment site for the kinds of adjuncts that host PGs comes from NPI licensing. NPI tests often involve QPs like *nobody* as the licenser, and are thus susceptible to the same problem as the variable binding test.[16] However, certain verbs are known to license NPIs as well, and these are not susceptible to the problem of QR creating a difference between s-structure and LF c-command relations. The contrast between (31)a and (31)c provides further support for a VP-external attachment site for the adjuncts. (31)a illustrates the NPI-licensing capability of *deny;* (31)b shows that the NPI is not licensed internal to the NP. If the rationale clause or temporal adjunct were inside the VP, then *deny* should license the NPIs in (31)c. However, the NPIs are not licensed. (31)d is an additional control to show that if the adjunct containing the NPI modifies a clause embedded under *deny,* the NPI becomes acceptable.

(31) a.   John denied rumors that he'd accepted any illegal contributions (/even a single illegal contribution)

     b.   * Rumors that he'd accepted any (/even a single) illegal contributions destroyed John's candidacy.

     c.   * John denied rumors of wrongdoing in order to impress anyone / even a single person
                                  ...after talking to anyone / even a single person

     d.   John denied [that he'd spread rumors of wrongdoing in order to harm even a single person]

---

[16] Urribe-Etxebarria (1995) argued that NPI-licensing applies at LF and is subject to the output of scope-changing operations. (Although she herself considered only reconstruc-tion, QR of a licenser should in principle be able to license an NPI.

Admittedly these are not minimal pairs. Although it is hard to see a way for a theory of NPI licensing to accommodate a right-branching structure for the VP in (31)c, it has been argued by Laka (1992) that predicates like *deny* do not themselves license NPIs, but rather select a clausal complement that in turn provides a licensing polarity head. But a somewhat more minimal contrast is provided by the double-object structure that *deny* allows:

(32) a.   Smith denied his workers even a small raise
      *(cf: *Smith gave his workers even a small raise)*

    b.   * Smith denied his workers a raise after talking to even a single lawyer /any lawyer

    c.   * Smith denied his workers a raise in order to starve even a single family

    d.   Smith denied that he'd [closed the factory in order to starve even a single family]

In (32)a, the NPI *even* is licensed in the second NP argument after the verb. On the other hand, *even* is not licensed in (31)b or (31)c, where it is contained in what would be considered the innermost argument of *deny* if the VP had a Larsonian structure. (31)d, by contrast, allows an NPI in a rationale clause if the latter is construed with the embedded clause. This sentence could felicitously be uttered as a follow-up to *Although he was accused of shutting down the factory merely out of a cruel wish to starve hundreds of families,....*

Thus, Barss-Lasnik c-command tests, applied to vPs that are modified by PG-compatible adjuncts, yield fairly strong evidence for *non*-right-branching structure, and apparently no compelling evidence against.

42

1.4   Summary

So far, we have seen that there are good reasons to assume that the sentential adjuncts that contain PGs are not VP-internal, but are higher—adjoined to vP or to some other projection below the surface position of the subject. In addition we have seen evidence of a fairly strong kind that PGs are gaps left by movement of a (phonetically empty) element, and that the resulting chain is distinct from the chain formed by the licensing movement. While we do not yet have an explanation for why the licensing chain must be present alongside of a PG, we have a promising hypothesis (Chomsky's null operator thesis) that provides a characterization of the distinctness of the two antecedents, and relates that characterization to an account for the fact that PGs appear in islands for movement.

However, in addition to the still unanswered question of why there needs to be a licensing movement, the fact that PGs are bound by a null operator raises a new problem, noted at the end of section 1.1: What principles allow the null operator structure to receive an interpretation at LF? In the next section I will argue that these two questions have a straightforward answer that unifies PGs with other null operator constructions. Namely, null operator movement forms a predicate, whose open position is available to apply to a local DP. Therefore any construction in which a DP raises to target the vP should allow an adjunct with a PG to form a modified predicate structure with the vP—yielding an interpretation as a conjoined predicate of the raised DP. I will try to show that the presence of a PG systematically correlates with this structure, whether the DP is an overt category (HNPS) or an unpronounced copy left by successive-cyclic movement.

2. Why PGs appear in natural language

In section 1.2 we saw evidence that the appearance of a PG indicates a structure like (33), formed by movement of an empty operator O to a position with scope over the island containing the PG. And in section 1.3 we saw that there are plausible reasons for believing that these structures are adjoined to vP.

(33)   **Predicate-adjunct formed by null operator movement:**

Adjunct

O_j without PRO reading t_j

Such structures are familiar. Relative clauses, for instance, have a similar configuration, with a gap left by movement of either an overt relative operator (34)a or a null operator (34)b:[17]

(34) a.   the book [which$_i$ [John read t_i]]

  b.   the book [O_i [(that) John read t_i]]

In (34)b operator movement from the position of the gap takes a constituent with a propositional meaning (*that I read* t) and yields a constituent that is interpreted as a property of individuals that the proposition (with a variable in place of the gap) is true of. We can represent the interpretation as the lambda abstract $\lambda x.John\ read\ x$. The sister of the relative clause, the NP *book*, is also a one-place predicate, which we can likewise represent by lambda abstraction ($\lambda x.x\ is\ a\ book$).

---

[17] Other constructions that have been argued to have this structure include purpose clauses (i)a, *tough-*movement (i)b, comparative-adjectival constructions (i)c, and comparative DPs (i)d (Cf. Chomsky 1977, Browning 1987).

(i) a.   We brought John along [O_i to talk to t_i]

  b.   I find Mary easy [O_i to appreciate t_i]

  c.   Gertrude is too stubborn (/stubborn enough) [O_i to disregard t_i]

  d.   Cool Hand Luke ate more eggs [O_i than they were able to boil t_i]

These two constituents compose together by an interpretive rule (predicate modification) that, in effect, conjoins their two meanings. The open argument position of each constituent is passed up to yield a one-place predicate which expresses the property of being both a book and an individual that was read by John, represented by the lambda abstract $\lambda x.x$ *is a book and John read x.*[18]

The claim that PGs involve operator movement suggests that adjuncts like (33) are likewise interpreted as predicates. This claim therefore raises a question: If the adjunct island (33) is a predicate, how does its open argument position become associated with a DP that is typically in a non-local position?

Of course, in a subset of environments in which PGs appear—for example HNPS—a DP does appear in a local enough position to enter into a predicate-argument relation with a null operator structure like (33). In section 2.1, I will show how raising of the heavy NP to target vP allows a straightforward compositional interpretation of adjuncts like (33)—i.e. the existence of HNPS makes the appearance of PGs unsurprising. In section 2.2 I will try to demonstrate that this is the end of the story. The account of why PGs appear in the environment of HNPS extends to all of the environments in which they appear, I will argue. The argument will be based on preliminary evidence that successive-cyclic raising targets the position of HNPS at every vP through which it passes. Further evidence for this claim will be presented in section 3.

2.1 A proposal: HNPS and vP-adjoined predicate modifiers

Ross (1967) argued that sentences like (35) involve rightward movement of the direct object past material inside the verb phrase. A number of subsequent studies have concluded that this so-

---

[18] Cf. Heim and Kratzer (1998).

called *heavy NP shift* (HNPS) operation targets the verb phrase (Bresnan 1976, Stowell 1981), as

shown in (36).[19]

(35)   John put on the table a recent article about global warming

(36)   **HNPS:**
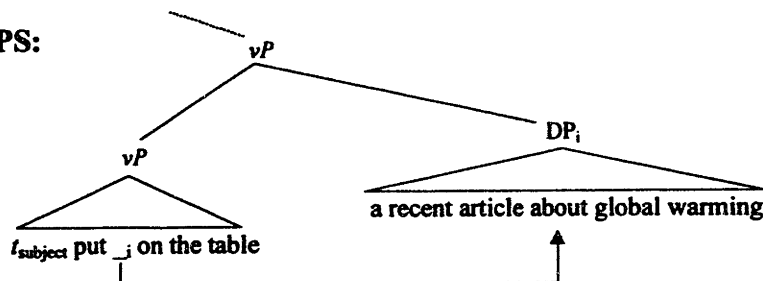
If it is correct that sentences like (35) are derived by a movement operation, then we can assume

that the semantic principles used to interpret these structures are the same as those used for

interpreting chains in general. Movement leaves a variable in the position of the gap, which is

bound to the antecedent. Here again we can represent the relation by lambda abstraction: the

target of movement is interpreted as a predicate abstract that binds the variable, and enters a

predicate-argument relation with the raised XP.[20] If HNPS targets the full, clausal vP (i.e. the

domain in which the external argument is thematically licensed) as in (36), then the

interpretation of that constituent shifts from clausal/propositional (*John put* t *on the table*) to a

one-place function ($\lambda x.John put x on the table$) predicated of the raised DP.

---

[19] In recent years this conclusion, and more generally Ross' original rightward movement analysis, have been challenged (cf. Larson 1988a, Postal 1995, Kayne 1995, 1998). In the second appendix to this chapter I will give arguments against the (very different) proposals of Larson and Postal. As noted in section 1.3, it is possible in principle that the proposals I will argue for below are consistent with those of Kayne 1998 despite superficial appearances.

[20] For present purposes we can think of predicate abstraction as a simple type-shifting mechanism that applies in the interpretive component (i.e. a complication of the semantics), rather than the reflex of a syntactic operation that adds an index to the target (i.e. a complication of the syntax, proposed in Heim and Kratzer 1998). Although an argument was given for the latter view in Nissenbaum 1998a,b, I will show in Chapter 3 that the argument was wrong on empirical grounds.

The structure (36) would therefore provide an optimal environment for a predicate-adjunct like (33) formed by null operator movement. Specifically, if (33) were adjoined to the lower node labeled vP in (36)—as in (38)—the two predicates could, in effect, yield a conjoined interpretation akin to that of an NP modified by a relative clause.

(37)   John put on the table without reading a recent article about global warming.

(38)
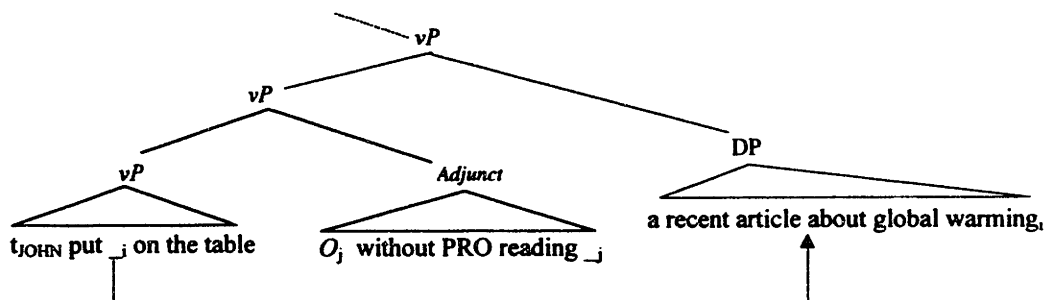
In short, raising of the DP in (38) yields an interpretation of the lowest vP-node as a predicate abstract, exactly as in (36).[21] In the case of (38), this constituent composes semantically with its

---

[21] The assumption being that predicate abstraction created by movement ignores adjuncts. One way of making this assumption explicit is by stating predicate abstraction as a rule that type-shifts the lowest saturated projection of the head that attracts the moving phrase, from type τ to $\langle e,\tau \rangle$. (See footnote20 and also section 2.1.)

sister, a predicate formed by null operator movement, yielding a conjoined predicate whose open argument position is bound by the raised DP.[22]

This is the manner in which a structure like (38) would be interpreted on minimal assumptions—just those that are required for the interpretation of movement in general, given the assumption that movement leaves a variable. Therefore if HNPS derives structures like (36),(38), the appearance of PGs is simply a predicted consequence. Given that UG provides mechanisms to derive the kinds of predicate-adjuncts formed by null operator movement (such as those that modify NPs), we would expect those adjuncts to appear in an environment in which they could be interpreted.[23] Verb phrases that are targeted by a movement operation like HNPS provide one such environment.

Note that this explains—at least for a subset of the constructions in which PGs are licensed—why a licensing movement is required. If no movement targeted the vP, there would be no derived predicate with which the predicate-adjunct containing the PG could compose. Moreover, we have an explanation (still limiting ourselves to PGs that are licensed by HNPS) for

---

[22] Throughout, I am making simplifying assumptions about the interpretation of vPs and vP-modifiers. In particular, I am abstracting away from temporal and event arguments that have been argued to be a part of the meanings of verb phrases. To be a bit more explicit about how the composition would work, assume that what I say in the text can be replaced with (i)-(iii), adopting Davidsonian event variables. Predicate abstraction would then yield a two-place relation between individuals and events, and the type-shifting referred to in the previous footnote would have to be restated in the appropriate way (i.e. "...the lowest saturated projection modulo event arguments...").

(i) lowest vP node of (38) interpreted as a predicate abstract (derived by raising of DP):
$\lambda x. \lambda e. e$ is a putting of x on the table by John

(ii) vP-adjoined null operator structure of (38):
$\lambda x. \lambda e.$ There is no e', e' a subpart of e (or of any relevant super-event E that includes e), such that e' is a reading of x by PRO

(iii) conjunction of (i) and (ii) (by predicate modification):
$\lambda x. \lambda e. e$ is a putting of x on the table by John and there is no e', e' a subpart of e (or of any relevant super-event E that includes e), such that e' is a reading of x by PRO

[23] Something independent would need to be said about the what constrains the distribution: for example, why do relative clauses—although they are predicates—only appear as sisters of NP. Questions of this sort arise independently of the assumptions I am defending here.

the *anti-c-command condition* that was illustrated in section 1 (by the examples in (4)), and the Kayne-Longobardi generalization (illustrated in (13)). The anti-c-command condition is just a consequence of the fact that a movement targeting vP is necessary to license a PG in a vP-modifier. The modifier is by necessity adjoined just below the landing site, since it composes semantically with the predicate derived by movement. The trace of the licensing movement will necessarily be lower in the structure than the landing site, hence cannot c-command into the adjunct. The stronger Kayne-Longobardi generalization (that the parasitic adjunct must c-command the licensing gap) follows for the same reason.

The proposal described above can be summarized as follows:

(39)   **Parasitic Gaps (in HNPS constructions) are licensed by predicate modification:**

PGs in HNPS constructions are gaps left by null-operator raising internal to a phrase that modifies the vP targeted by HNPS. The target vP and the modifier together form a (conjoined) predicate of the shifted NP.

2.2   Generalizing the proposal: successive-cyclic movement through spec-vP

The discussion so far has been limited exclusively to PGs that are licensed by HNPS. If it could be shown that *every* PG (including those licensed by long-distance movement) is similarly licensed by a movement that targets vP, we would then have a general explanation for the appearance of this puzzling phenomenon.[24] I will argue here (and further in section 3) that this

---

[24] For now I will ignore PGs contained in subject NPs. In the first appendix to this chapter I will argue that subject PGs fall together with PGs in adjuncts under (a slight enrichment of) the explanation proposed in this section. Specifically, I will try to show that subjects with PGs are interpreted in the vP-internal position (yielding obligatory reconstruction effects as demonstrated in Nissenbaum 1998a/b). I will argue that if null operator movement within the subject yields a category that is interpreted as a function from individuals to individuals (or to generalized quantifiers), then a recursive definition of the predicate modification rule will allow such a category to compose semantically with the derived predicate created by the licensing movement that targets vP.

explanation is correct. Specifically, I will present evidence that long-distance movement targets vP on the way to the final landing site—in effect, passing successive-cyclically through the position of HNPS. This claim, stated in (40), in turn allows for the more general statement of the conditions under which PGs arise (41):

(40)   **The vP-step:**
        Successive-cyclic A-bar movement targets every vP along the way to the final landing site.

(41)   **Parasitic Gaps are licensed by predicate modification:**
        A PG is the gap left by null-operator raising internal to a phrase that modifies a vP targeted by movement. The target vP and the modifier together form a (conjoined) predicate of the moved XP.

According to (40), a sentence like *Which article did John file* has the structure shown in (42), with an intermediate trace of *which article* in the structural position of HNPS.[25] By (41), this structure would license a PG, with the predicate-modifier containing it adjoined just below the intermediate trace as shown in (43).
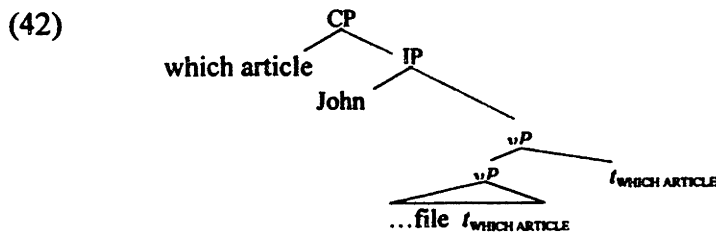
(42)

---

(43)

```
                        CP
                       /\
          which article   IP
                       /\
                   John
                        \vP
                       /\
                      vP
                     /  \
                    vP    Adjunct        t_WHICH ARTICLE
                   /\     /\
              ...file t_WHICH ARTICLE   O_j without reading t_j
```
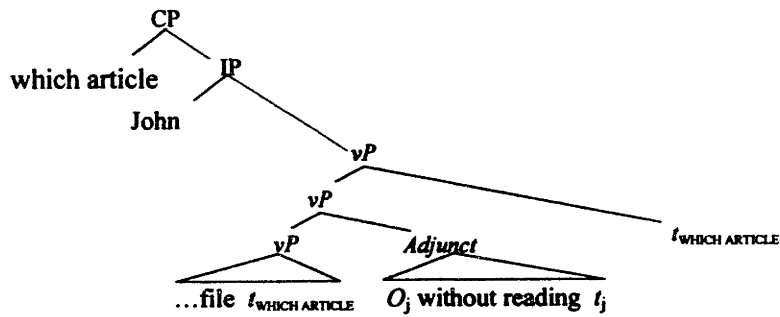
In (43)—exactly as in the HNPS structure (38)—the predicate-adjunct containing the PG composes with the lower segment of vP to form a semantically conjoined predicate of the raised XP, in this case an intermediate trace. Here, as before, predicate modification is made possible by the fact that movement of *which article* has targeted vP, yielding a derived predicate.

Note that (40) incorporates the claim that *every* potential vP is targeted by successive-cyclic movement. This claim is forced on empirical grounds if we accept the proposal (41) that the vP-step is what licenses PGs. Recall from the examples in (3)a (repeated below as (44)a,b) that the adjuncts containing PGs can be adjoined to a vP in a higher clause than the one from which the licensing extraction takes place.

(44) a. What movies did Mary [claim she liked _][in order to get you to see _]?
         cf. #What movies did Mary like _ in order to get you to see _

     b. John's the guy that they said they'll [hire _][if I criticize _ publicly]

Both examples in (44) involve a licensing extraction from the most embedded clause. But while in (44)b the adjunct containing the PG modifies the vP of the embedded clause, the adjunct in (44)a can only be adjoined to the matrix vP (since rationale clauses presuppose agentivity and are incompatible with psych-predicates such as *like*). Therefore, (40) and (41) require the

postulation of the underlying structure (45) (omitting, for the sake of visual clarity, the intermediate trace in the embedded spec-CP).

(45)
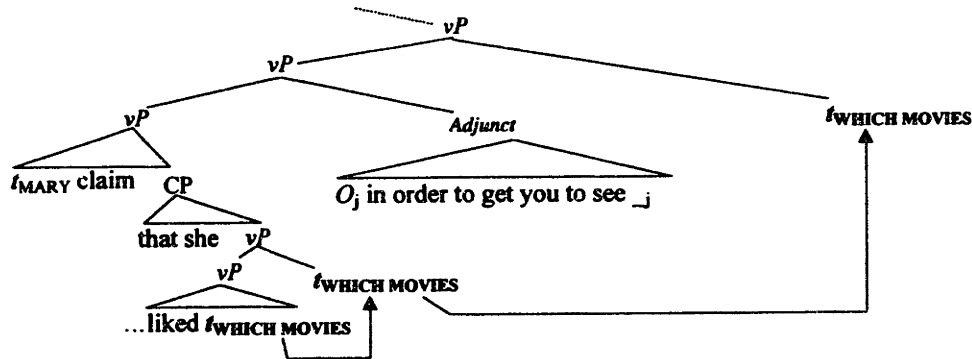
Although we don't have direct evidence for successive-cyclic movement leaving a trace in the *lower* spec-vP in (45), (41) requires the presence of a trace in that position in cases like (44)b, where the modifier containing the PG is adjoined to the lower vP.

A sentence like (46), on the other hand, does provide evidence for intermediate traces in both embedded and matrix spec-vP positions if we assume (41).

(46) John's the guy they [said they'll [hire _ [if I criticize _ publicly]]][in order to get me to praise _]

This sentence has the property that *both* the matrix and embedded vPs are modified by adjuncts that contain PGs. Therefore (41) requires that the underlying structure include the two intermediate traces shown in (47) (again, omitting the trace in the embedded spec-CP).

(47)

vP
vP
vP
t_{THEY} said   CP
...they'll   vP
vP
vP   Adjunct
...hire $t_{\text{RELATIVE OP.}}$ O_i if I criticize $\_i$ publicly
$t_{\text{RELATIVE OP.}}$
Adjunct
O_j in order to get me to praise $\_j$
$t_{\text{RELATIVE OP.}}$
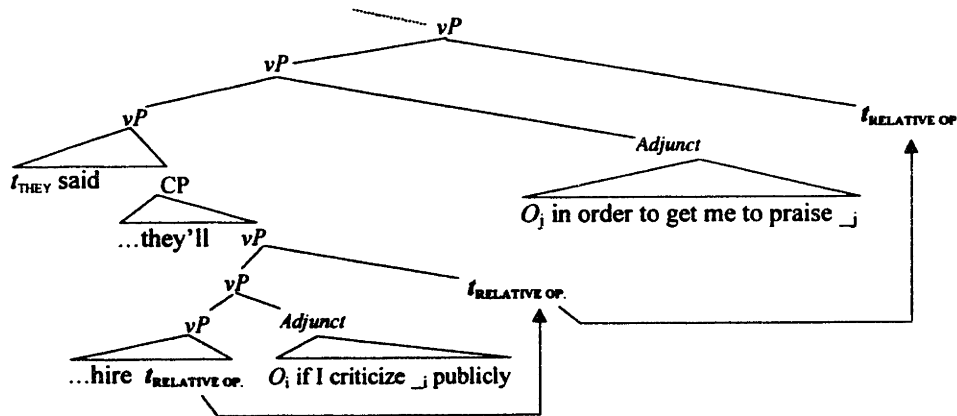
In short, the proposal (41) amounts to the empirical claim that whenever a PG appears in the environment of a long-distance licensing movement, there must be an intermediate trace of that movement local to the parasitic adjunct, in the structural position of HNPS. This claim leads to the expectation that we will find cases of long-distance HNPS in sentences parallel to those in (44)a and (46), with PGs contained in matrix adjuncts. The expectation is borne out:

(48) a.   Mary [claimed that she liked _][in order to get me to see _] that movie with Fred Astaire and Audrey Hepburn

   b.   They [said they'll [hire _ [if I criticize _ publicly]]][in order to get me to praise _] the man who rejected my proposal

The HNPS in (48)a,b raises out of the embedded clauses to target the matrix vP. In both cases a PG is licensed in a matrix adjunct; in (48)b an additional PG is licensed in a modifier adjoined in the embedded clause—suggesting again that the movement was successive-cyclic, targeting the embedded vP. The structures for these two sentences are exactly those of (45) and (47), except for the fact that the higher of the two intermediate traces shown in each of those structures is replaced by the overt shifted DPs.

The proposal that PGs are licensed by familiar and well-motivated principles of compositional semantic interpretation (specifically predicate modification) is attractive if it can

be sustained. As I argued in section 2.1, it would answer several of the longstanding puzzles associated with PGs: their parasitic nature, the anti-c-command condition and the Kayne-Longobardi generalization. In simple cases of HNPS, the proposal is natural and straightforward. What remains to be seen is whether the claimed vP-step of long-distance movement (40) can be empirically justified. If so, we would have good reason for viewing PGs as just one instance of the more general phenomenon of null operator constructions. The latter appear in environments in which predicate modification is possible. Such environments include those in which a vP is targeted by movement, becoming a derived predicate. The null operator constructions that appear in that particular environment have been singled out under the designation PGs, but are no different in principle from others such as relative clauses, purpose clauses and the like.

In the next sub-section I will briefly review independent evidence for a vP-step in successive-cyclic movement, discovered by Fox (1999,2000a). In section 3, I will demonstrate that a number of surprising predictions follow from the picture that has been developed so far, providing further evidence of a quite striking nature for the existence of the vP-step.

2.2.1 Independent support for the vP-step

Fox (1999,2000a) gives an argument that A-bar movement passes through an intermediate position between the surface subject position and the highest internal argument of the VP, as I proposed in (40). The copy left by this intermediate step is always unpronounced in English, but its presence at LF can be detected by means of an ingenious experiment devised by Lebeaux (1990).[26] The argument is based on a correlation between variable binding reconstruction effects

---

[26] Lebeaux 1988 developed this technique to argue for certain effects of intermediate copies of successive cyclic wh-movement in Spec,CP. It was extended by Fox to test for effects of successive cyclic movement through other positions as well.

and Condition C of the binding theory. I will present an abbreviated version of the argument here, consisting of two parts. The first part (due to Lebeaux 1990) concerns the contrast in (49). Fox explained this contrast on the assumption (contra Lebeaux) of a single level of representation (namely LF) at which variables must be bound and Condition C must be satisfied, together with the copy theory of movement.

(49) a. Which paper that he$_j$ wrote for her$_i$ seminar did Mary$_i$ say every student$_j$ should revise _?

  b. * Which paper that he$_j$ wrote for Mary$_i$'s seminar did she$_i$ say every student$_j$ should revise _?

In neither (49)a nor (49)b is the bound variable pronoun *he* apparently in the scope of *every student*. On the assumption that the copy theory is what underlies reconstruction effects, the variable can be bound because a full copy of the *wh*-phrase is present at the tail of the chain in the LF representation. An LF representation of (49)a embodying this assumption is illustrated in (50)a, with a copy of the *wh*-phrase still present in its thematic position.[27]

(50) a. [Which paper that he$_j$ wrote for her$_i$ seminar]$_k$ did Mary$_i$ say <u>every student$_j$</u> should revise ⟨which paper that <u>he$_j$</u> wrote for her$_i$ seminar⟩$_k$

There is a contrast in acceptability between (49)a and (49)b. The only difference between the two sentences is that the positions of the r-expression *Mary* and the co-indexed pronoun are reversed. In (49)b the r-expression is contained in the same relative clause as the bound variable. The variable-induced reconstruction (i.e. the need for the full copy to be present at the tail) thus

---

[27] Something has to be said about how this structure would be interpreted. We could adopt Engdahl's (1980) choice function approach, together with the copy theory. The NP-restrictor of the *wh*-phrase would be interpreted in the reconstructed position:

(i) *For which choice function F, Mary said every student, should revise F(paper that he, wrote for her seminar)*

creates a conflict. If a copy of the *wh*-phrase is present at the tail of the chain, *Mary* will be c-commanded by *she,* resulting in a violation of Condition C. This is shown in (50)b. On the other hand, if the relative clause is absent from the tail of the chain at LF (an *anti*-reconstruction effect) as in (50)c, then the only instance of the variable is the one in the copy at the head of the chain, outside the scope of the quantifier. Hence the variable will remain unbound.

(50) b. [Which paper that he_j wrote for $Mary's_i$ seminar]$_k$ did she_i say [every student]$_j$ should revise ⟨which paper that he_j wrote for $\underline{Mary's_i}$ seminar⟩$_k$

   c. [Which paper that he_j wrote for $Mary's_i$ seminar]$_k$ did she_i say [$\underline{every\ student}$]$_j$ should revise ⟨which paper⟩$_k$

Two important controls should be noted, to show that variable binding and Condition C are indeed the two requirements that are in conflict. (51)a differs from (49)b only by the absence of a quantifier. This minimal difference yields a contrast in acceptability: it shows that as long as the pronoun *him* in the relative clause doesn't need to be bound by a quantifier, the *wh*-phrase doesn't need to reconstruct, hence co-reference can obtain without violating Condition C.[28]

(51) a. Which paper that he wrote for Mary's seminar did she say John should revise _?

   b. Which paper that he wrote for Mary's seminar did [her T.A.] say every student should revise _?

---

[28] The fact that movement ordinarily bleeds Condition C when the r-expression is embedded in a relative clause seems, at first glance, difficult to reconcile with the copy theory. Chomsky (1993) argued that there is no conflict. Adapting a proposal of Lebeaux 1988, Chomsky proposed that relative clauses (and adjuncts generally) can be merged into the structure at a point in the derivation after the NP undergoes movement. Such a derivation would yield (i) as the LF for (51)a (i.e. the relative clause containing the offending r-expression is merged to the NP only in spec-CP), and would explain the absence of Condition C effect:

(i) [Which paper that he wrote for $Mary's_i$ seminar]$_k$ did she_i say John should revise ⟨which paper⟩$_k$

Chomsky's adaptation of Lebeaux's late-merger proposal is taken up again in chapters 4 and 5.

(51)b likewise provides a minimal contrast with (49)b. The condition C violation in (49)b resulted from the reconstructed r-expression *Mary* being c-commanded by a co-referring pronoun, but in (51)b the pronoun is embedded in the subject. As a result, condition C is avoided, and the sentence is acceptable under a co-referent interpretation.

So far, the argument has demonstrated only that there is a correlation between variable binding reconstruction and Condition C effects. The tests that have shown this, however, are consistent with a theory in which reconstruction is always to the thematic position. Lebeaux (1990) made use of the same experimental logic to demonstrate the existence of an intermediate position in an embedded spec-CP. Fox (1999/2000a) extended Lebeaux's test further, to show the effects of another intermediate position *below* the subject.

Consider the examples in (52). If reconstruction were available only to the thematic position, then these sentences should be unacceptable, on a par with (49)b. The LF structures that include full copies at the tail of the chain are shown in (53).

(52) a. Which paper that he$_i$ wrote for Mary's$_j$ seminar did every student$_i$ give her$_j$ a revision of _?

b. Which paper that he$_i$ wrote for Mary's$_j$ seminar did every student$_i$ promise her$_j$ that he$_j$ would revise _?

(53) a. [Which paper that he$_i$ wrote for Mary's$_j$ seminar]$_k$ did every student$_j$ give her$_i$ a revision of ⟨which paper that he$_j$ wrote for Mary's$_i$ seminar⟩$_k$

b. [Which paper that he$_i$ wrote for Mary's$_j$ seminar]$_k$ did every student$_j$ promise her$_i$ that he$_j$ would revise ⟨which paper that he$_j$ wrote for Mary's$_i$ seminar⟩$_k$

The twin requirements of having the variables in the scope of their binders and simultaneously ensuring that *Mary* is not in the c-command domain of the (co-indexed) indirect object should prove to be a fatal conflict, as they are in (49)b. However, the sentences in (52) are relatively

acceptable, far more so than their close counterparts (54)a and (54)b. Both of those sentences contain violations of Condition C, the one in (54)b induced by the forced presence of the full copy at the tail of the chain. Contrary to expectation, the status of (52)a,b is similar to (54)c, a sentence in which there is *no* copy-theory induced violation of Condition C.

(54) a. * Every student$_i$ gave her$_j$ a revision of a paper that he$_i$ wrote for Mary's$_j$ seminar
/ promised her$_j$ that he$_i$ would revise a paper that he$_i$ wrote for Mary's$_j$ seminar

    b. * Which paper that he$_j$ wrote for Mary's$_i$ seminar did she$_i$ give every student$_j$ an F on _?

    c. Which paper that he$_j$ wrote for Mary's$_i$ seminar did every student$_j$ give [her$_i$ TA] a revision of _ / promise [her$_i$ T.A.] that he$_j$ would revise _?

The contrast in acceptability between (52) and (54)b provides us with the crucial argument for the existence of an intermediate position. The fact that (52)a,b allow co-reference shows that there must be an intermediate position for the relative clause—a position that is in the scope of the subject *every student,* but higher than the indirect object pronoun. That is, their LFs cannot be (53)a-b but could be (55)a-b:[29]

(55) a. [Which paper that he$_j$ wrote for Mary's$_i$ seminar]$_k$ did every student$_j$ ⟨which paper that he$_j$ wrote for Mary's$_i$ seminar⟩$_k$ give her$_i$ a revision of ⟨which paper⟩$_k$

    b. [Which paper that he$_j$ wrote for Mary's$_i$ seminar]$_k$ did every student$_j$ ⟨which paper that he$_j$ wrote for Mary's$_i$ seminar⟩$_k$ promise her$_i$ that he$_j$ would revise ⟨which paper⟩$_k$

These facts provide fairly strong independent evidence for an intermediate reconstruction site for *wh*-movement, a site in between the subject and the highest internal argument of the VP. Taken together with copy theory of movement as an explanation for reconstruction effects, this

---

[29] Evidently, QR of *every student* does not rescue the sentence in this case—either because such an operation would result in a WCO violation, or because QR past the subject does not take place in this sentence. Cf. Fox 2000a.

evidence constitutes independent support for an intermediate landing site in long-distance movement. The intermediate position is consistent with the vP-landing site that I proposed in (40). Note further that the crucial intermediate copy in (55)b is in the matrix clause. No copy of the *wh*-phrase inside the embedded CP could include the relative clause without inducing a Condition C effect. We thus have independent evidence that the intermediate steps of movement are successive-cyclic, passing through a position internal to every clause along the way to the final landing site, exactly as claimed in (40).

3. The utility of PGs as a diagnostic for invisible structure

I proposed in section 2.2 that essential properties of PGs (including the fact that they exist) are explained on the basis of general principles of compositional interpretation together with the claim that vP is targeted by successive cyclic movement (40). The explanation starts with the premise that PGs are the gaps left by null operator movement in a vP-modifier (yielding interpretation as a predicate). In HNPS constructions, the vP with which this modifier semantically composes is itself the target of movement (hence a derived predicate); the modified vP is interpreted as a predicate of the raised DP. I suggested in (41) that this account provides a model that can be generalized to all PG licensing if long-distance movement successive-cyclically targets the structural position of HNPS.

In this section I will give additional arguments for this picture, by showing that it is possible to see indirect effects of the claimed intermediate traces in spec-vP. The argument has the following logic. If we could isolate some distinctive pattern of effects that is directly

attributable to HNPS, and then show that long-distance movement produces the same pattern, we would have a powerful demonstration that the long-distance movement involves an intermediate step of HNPS. I will argue that a generalization discussed by Larson (1988a) provides a tool of this sort. Larson observed that HNPS can place a DP to the right of a vP-adjunct only if the adjunct contains a PG. These patterns, I will argue, reveal the presence (and the precise structural position) of an intermediate trace in the position of HNPS.

3.1   Larson's generalization: HNPS past the adjunct makes PGs obligatory

Larson (1988a) observed an interesting fact about PG licensing by HNPS. HNPS past a vP-modifier of the relevant type *forces* a PG to appear.[30, 31, 32]

(56) a.   John filed _ [without reading _] a recent article about global warming
    b.   *?[*without reading it*]
    c.   *?[*without reading your e-mail*]

(57) a.   I hired _ [without even interviewing _] a person who was wearing fancy shoes
    b.   *? [*without even interviewing him*]
    c.   *? [*without even talking to the chair*]

(58) a.   I offended _ [by not recognizing _ immediately] my favorite uncle from Cleveland
    b.   *? [*by not recognizing him immediately*]
    c.   *? [*by not recognizing you immediately*]

---

[30] Engdahl (1983) reported that this pattern holds for Swedish, but pointed out that some English speakers find the examples without PGs acceptable. I find the contrasts very strong, but the deviance of examples like (56)-(59)b,c. can be neutralized with a strong pause between the adjunct and the shifted NP, setting them both off as separate intonational phrases. (Engdahl remarks [p. 12] that "the possibility of having a coreferent pronoun…increases if the adverbial clause is taken to be a prenthetical and set off from the rest of the sentence by heavy intonational breaks.") The crucial fact, then, is that the a. examples of (56)-(59) can be pronounced as single intonational phrases (without any pause), whiie the b. and c. examples, if acceptable at all, need to be broken up intonationally. Larson (1988a, note 33) adds that others have independently reported these contrasts as quite sharp.

[31] Larson argued, partly on the basis of this pattern, for a radically different analysis of HNPS that does not involve movement of the NP. See the second appendix for an argument that Larson's analysis cannot be correct.

[32] Thanks to Sabine Iatridou for very helpful discussion of this paradigm and the examples in this section.

(59) a. We invited _ to the party [in order to impress _] the guy who criticized us in public

b.   *? [in order to impress him]

c.   *? [in order to impress you]

In contrast to these examples, (60) shows (unsurprisingly) that the same adjuncts without PGs may appear *to the right* of an NP that has undergone HNPS (past an argument or a non-clausal adverbial.)[33]

(60) a. John put _ on the table a recent article about global warming [without reading it]

b. I hired _ with no hesitation a person with fancy shoes [without even interviewing him]

c. I offended _ very badly my favorite uncle from Cleveland [by not recognizing him]

d. We invited _ to the party the guy who criticized us in public [in order to impress him]

We can state Larson's generalization as follows:

(61)   **Larson's generalization:**
         HNPS cannot appear to the right of a vP-adjunct unless that adjunct contains a PG

This generalization is interesting because it runs counter to the otherwise free alternation of PGs with pronouns—in fact much of the early literature reported PGs as slightly marginal. But Larson's generalization identifies an environment where the opposite is true—an environment that is degraded *without* a PG.

Let us assume provisionally (contra Larson) that the claims underlying the preceding discussion (section 2) are sound. Specifically, HNPS creates structures like (38), repeated below;

and interpretation of the resulting chain involves predicate abstraction below the vP-modifier (allowing the modifier to compose with vP by predicate modification).

(38)
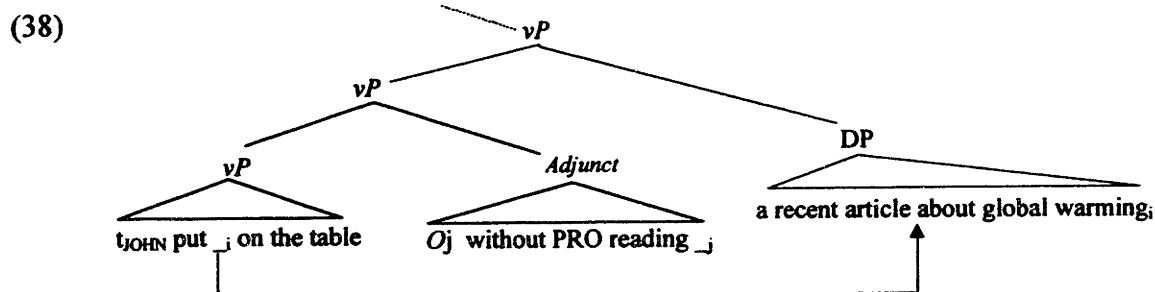
This picture allows us to draw an interesting conclusion from Larson's generalization. Namely, the obligatoriness of the PG in (38) tells us that predicate abstraction not only can but *must* apply to the lowest vP node—i.e. the interpretive rule 'skips' the adjunct.[34] If this were not the case then we would expect the adjunct in (38) to alternate freely with an adjunct *lacking* a PG (with abstraction applying to the sister of the moved constituent, i.e. to [vP[vP...][adjunct...*(no PG)*...]]).

Predicate modification requires both the vP and vP-modifier to have the same semantic type: if one of them is an *n*-place predicate, the other must be as well.[35] That, together with the lesson that we learn form Larson's generalization, entails that in any structure of the general form (62), the lowest vP node (a derived predicate of the raised DP) can only be modified by an adjunct that is a predicate, i.e. an operator-gap structure like (33).

---

[33] Non-clausal adverbials like the one in (60)b are low in the VP, as shown by the kinds of tests used in section 1.3. Larson 1988a argued that clausal modifiers have a higher attachment as assumed here.

[34] Cf. footnote 21.

[35] In the appendix to this chapter I will propose a minor modification to this statement, to generalize the account to include PGs contained within subject NPs. The modification will not affect any of the predictions discussed here.

(62) *Ill-formed unless there is a PG (predicate modification would fail):*

```
                           vP
                  vP
           vP           Adjunct              DP
        ... _ ...    ... OBLIGATORY PG ...
           |_____|
```
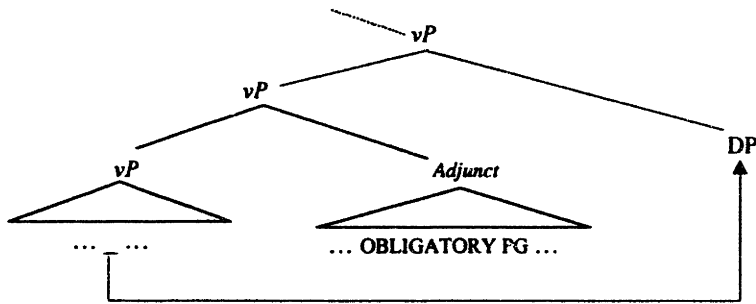
If this is the correct way of thinking about Larson's generalization, we can give a more general statement of the environments in which PGs are predicted to be obligatory:

(63)   **Obligatory PG generalization (OPG):**
A movement that targets vP will force the appearance of a PG inside a vP-modifier adjoined below the landing site.

Larson's generalization (and its extension to (63)) is further supported in cases where multiple vP-modifiers are adjoined to the same vP. Given (63) we expect to find that HNPS to the right of more than one vP-adjunct forces a PG to appear in each of them (illustrated schematically in (64)a). Just as in the previous examples, predicate modification should fail in (64)a unless both adjuncts are operator-gap structures. Since the lowest vP-node is a derived predicate of the raised DP, *Adjunct 1* has to be a predicate. And since these two constituents compose by predicate-modification, passing their open argument position up, *Adjunct 2* must be a predicate as well.
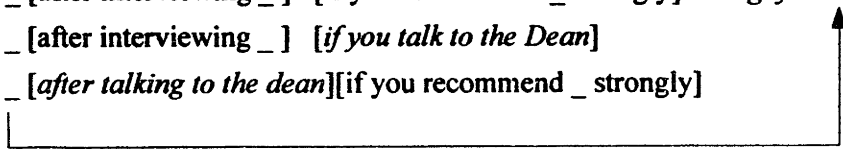
(64)   **Stacked vP-adjuncts**
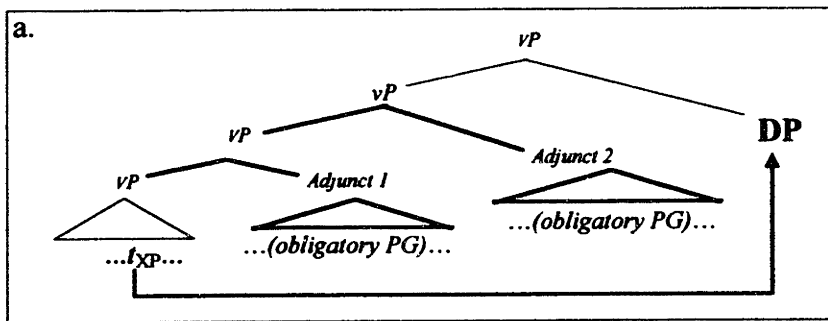
HNPS can also position a DP to the *left* of a vP-modifier, as we saw in (60). This means that when there are two vP-modifiers, there are *three* possible landing sites for HNPS. The DP can move past both adjuncts, as in (64)a. Alternatively, it can move past just the first one, or past neither. These two alternative possibilities are illustrated in (64)b and (64)c, respectively.
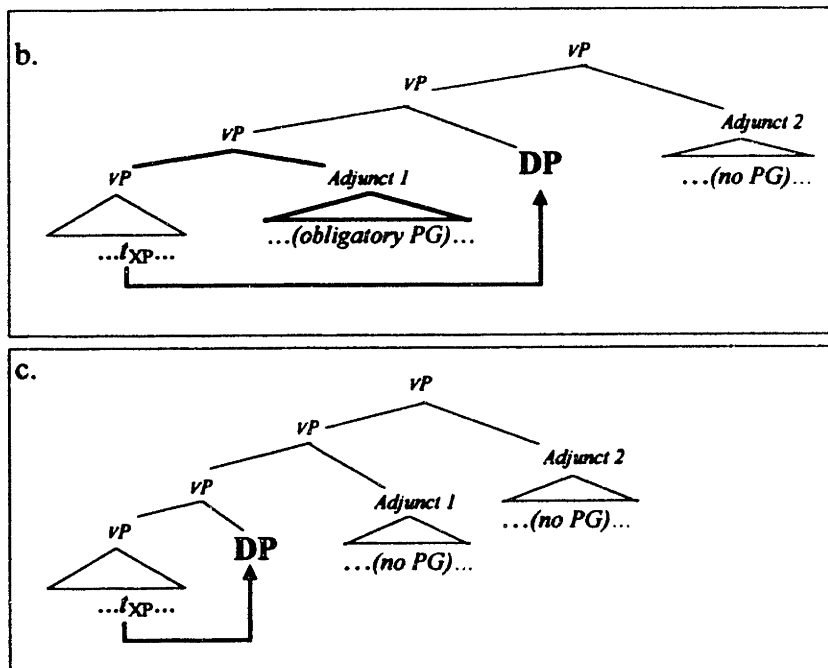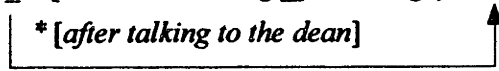
The examples in (65)-(67) correspond to these three configurations. The pattern of judgments verify the generalization, just as in the structurally simpler (56)-(60).

(65)　**HNPS past both adjuncts (as in (64)a)**

    a.　I'll hire _ [after interviewing _]　[if you recommend _ strongly]　*the guy with fancy shoes*

    b.　* I'll hire _ [after interviewing _]　[*if you talk to the Dean*]

    c.　* I'll hire _ [*after talking to the dean*][if you recommend _ strongly]

(66)　**HNPS past adjunct 1 only (as in (64)b)**

    a.　I'll hire _　[after interviewing _]　*the guy with fancy shoes* [if you recommend him]

    b.　　　　* [*after talking to the dean*]

(67)　**HNPS past neither adjunct (as in (64)c)**

    I'll praise _ to the sky *the guy with fancy shoes* [after visiting him][if you recommend him]

Descriptively, HNPS appears to be unique among the environments in which PGs are licensed, in that the landing site of the shifted DP determines whether or not a PG can appear. As noted above, this is simply a consequence of the fact that HNPS creates a derived predicate; any modifying adjunct need to be of the same semantic type, hence an operator-gap structure. Larson's generalization thus provides us with a potentially useful means of probing cases of long-distance licensing of PGs. In the next section I will show that long-distance licensing of PGs, when probed, reveals the distinctive pattern characteristic of the more local HNPS.

3.2　A predicted correlation: PGs and the positions of intermediate traces

I argued in section 2.2 that whenever a PG appears in the environment of a long-distance licensing movement, there must be an intermediate trace of that movement local to the parasitic adjunct, in the structural position of HNPS. The argument was based solely on the fact that this claim allows a reduction of long-distance PG licensing to principles of local semantic composition. We can use Larson's generalization to test the claim further. The vP-trace

hypothesis leads us to expect that the pattern exhibited by HNPS should extend to long-distance movement.

Larson's generalization correlates the presence vs. absence of a PG with the position of HNPS. If HNPS raises a DP past a vP-adjunct, a PG is obligatory; HNPS below the adjunct does not license a PG in that adjunct. Given this correlation, it should be possible in principle to see the effects of HNPS *even if the shifted NP is invisible*. As noted in the last section, Larson's generalization has a natural extension to all movement, and should not be limited to (overt) HNPS. I stated this extended condition as the Obligatory PG generalization (63), repeated below.

(63)   **Obligatory PG generalization (OPG)**
        A movement that targets vP will force the appearance of a PG inside a vP-modifier
        adjoined below the landing site.

By giving us an independent means of detecting the structural position of HNPS with respect to a vP-adjunct—a means that does not depend on the overt word order—the Obligatory PG generalization allows us to make inferences about the exact position of an intermediate trace. Specifically, the presence of a PG indicates the presence of a trace just above the adjunct (this follows from the predicate modification hypothesis (41) that was proposed in section 2.2). The absence of a PG, on the other hand, forces the conclusion that the trace is below the adjunct. (This follows from the Obligatory PG generalization.) We can then corroborate those inferences by means of the intricate patterns of obligatory and illicit PGs that resulted from HNPS.

These two inferences about the position of traces are illustrated in (68)a-b:

(68) a. ***Presence of a PG signals the presence of an intermediate trace above the adjunct:***
(follows from (41))

```
                              vP
                  vP        /    \
              vP      Adjunct     \
           /  \      /      \      obligatory trace
          ···_···   O_i [ ... PG_i ... ]
            |_____|
```
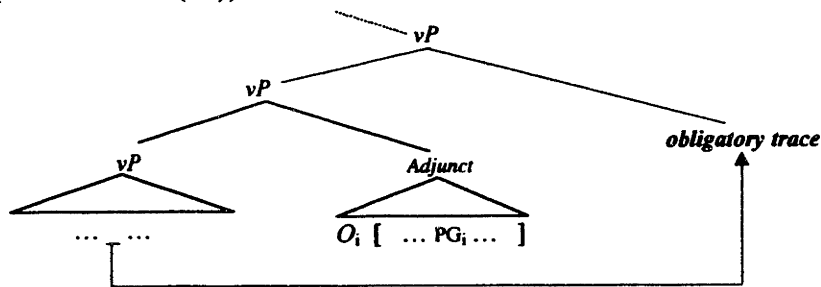
b. ***Absence of a PG signals the absence of an intermediate trace above the adjunct:***
(follows from (63))

```
                              vP
                  vP        /    \
              vP      trace       \
           /  \    |               no trace permitted
        vP            Adjunct
       /  \          /      \
      ···_···      ... (NO PG) ... ]
        |_____|
              (*)
```
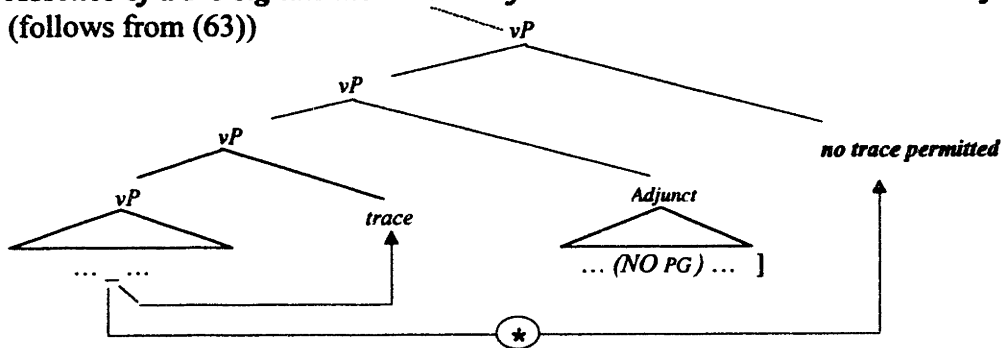
The remainder of section 3 is an attempt to demonstrate the correctness of the Obligatory PG generalization and of the corollary assumption that PGs are licensed by traces in the structural position of HNPS. Three sorts of evidence will be presented. In section 3.2.1 I will show that the effects of HNPS in the environment of stacked vP-adjuncts carries over to cases of long-distance movement, yielding an asymmetric pattern: a PG cannot appear in the higher adjunct unless one also appears in the lower one. Section 3.2.2 will present an argument based on an additional assumption, argued for in Fox and Nissenbaum 1999 (chapter 4 of this dissertation). The additional assumption is that relative clause extraposition marks a position in a chain. Given that assumption the effects of HNPS are predicted to emerge in extraposition from *wh*-phrases, a prediction that is borne out. Finally, I will argue in section 3.2.3 that a surprising pattern of Antecedent-contained deletion follows from the interaction of the Obligatory PG generalization with an independently motivated generalization about ACD.

3.2.1 Stacked vP-adjuncts in *wh*-movement environments

We saw in the last section that Larson's generalization yields a distinctive pattern when HNPS interacts with a pair of vP-adjuncts: the choice of landing site determines whether a PG appears in both, just one, or neither of the adjuncts. If the more general OPG holds, then the three possible landing sites for the proposed intermediate step of movement should yield a similar pattern. Specifically, the pattern of obligatory and illicit PGs that would result from the OPG is illustrated schematically in (69)a-c.

(69)

Crucially, the intermediate trace should reveal itself by producing an asymmetric pattern if the OPG holds. There should be nothing wrong with long-distance movement licensing a PG in each

adjunct (as in (69)a), or *no* PG in either adjunct (as in (69)c). However, if a PG appears in *only one* of the two adjuncts, the OPG predicts that it has to be the innermost ((69)b). There is no way for a movement to license a PG in Adjunct 2 without also forcing one to appear in Adjunct 1. The movement would have to raise past Adjunct 2, resulting in the structure (69)a.

This pattern emerges as predicted. Examples (70) and (71) show, respectively, that both adjuncts may contain PGs, and both may lack them.
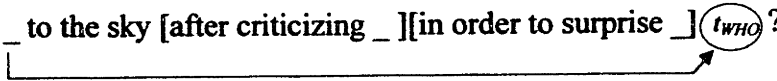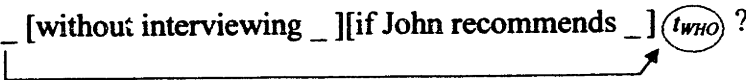
(70) a.   Who did you praise _ to the sky [after criticizing _][in order to surprise _]?

    b.   Who will you hire _ [without interviewing _][if John recommends _]?

(71) a.   Who did you praise _ to the sky [after criticizing him][in order to surprise the poor man]?

    b.   Who will you hire _ [without interviewing him][if John recommends him]?

However, if just one of the two adjuncts contains a PG it must be the innermost (72). Sentences in which only the outermost adjunct contains a PG are unacceptable (73):
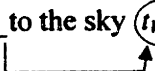
(72) a.   Who did you praise _ to the sky [after criticizing _][in order to surprise **him**]?

    b.   Who will you hire _ [without interviewing _][if John recommends **him**]?

(73) a.   *Who did you praise _ to the sky [after criticizing **him**][in order to surprise _]?

    b.   *Who will you hire _ [without interviewing **him**][if John recommends _]?

This asymmetric pattern suggests that these sentences have the structures shown in (74)-(77), with an intermediate trace of *wh*-movement appearing in each of the three possible positions. The unacceptability of (73) is a consequence of the OPG; there is no position for an intermediate trace that could give rise to that pattern.
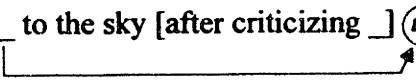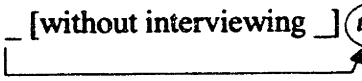
Both adjuncts may contain PGs (70)
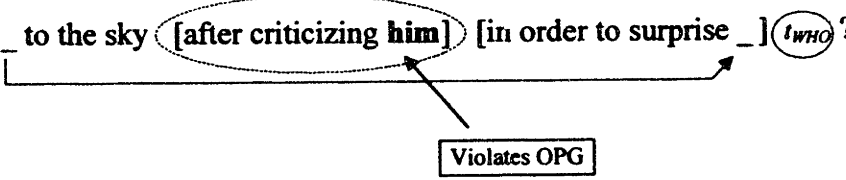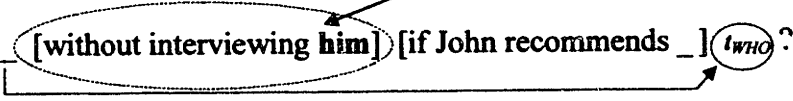
(74) a.   Who did you praise _ to the sky [after criticizing _][in order to surprise _] t_{WHO}?

b.   Who will you hire _ [without interviewing _][if John recommends _] t_{WHO} ?

Both may be devoid of PGs (71):

(75) a.   Who did you praise_to the sky t_{WHO} [after criticizing him][in order to surprise the poor man]?

b.   Who will you hire _ t_{WHO} [without interviewing him][if John recommends him]?

If *only one* of the two adjuncts contains a PG, it must be the innermost (72), (73):

(76) a.   Who did you praise _ to the sky [after criticizing _] t_{WHO} [in order to surprise him]?

b.   Who will you hire _ [without interviewing _] t_{WHO} [if John recommends him]?

(77) a.   *Who did you praise _ to the sky [after criticizing him] [in order to surprise _] t_{WHO} ?

Violates OPG

b.   *Who will you hire [without interviewing him] [if John recommends _] t_{WHO} ?

It is possible in principle—although rather unlikely—that examples like (73) are ruled out by an unexplained prohibition against the sequence [adjunct ...*(no PG)*...][adjunct ...PG...]. But we can demonstrate that such a generalization would be false. Recall the examples from section 2.2 that

involved PGs in both the embedded and the matrix clause, licensed by a single long-distance extraction. One such example is repeated below as (78).

(78)   John's the guy they [said they'll [hire _ [if I criticize _ publicly]]][in order to get me to praise _]

Now consider (79), which differs minimally from (78) in having a pronoun in the first adjunct instead of a PG. However, (79)a seems perfectly acceptable, contrasting with (73)a-b and more minimally with (79)b (in which both adjuncts are construed as modifying the same vP).

(79) a.   John's the guy they [said they'll [hire _ [if I criticize him publicly]]][in order to get me to praise _]

     b.   * John's the guy they'll [hire _ [if I criticize him publicly][in order to console _]]

The acceptability of (79), and the contrast with examples like (73)a-b, follows from the claim that long-distance movement targets the vP in *each* clause. Since the modifiers in (79) are adjoined to two separate vPs (one in the embedded and one in the matrix clause), it is possible for the movement through the embedded clause to target a position below *Adjunct-1* and the movement through the matrix clause to target a position immediately above *Adjunct-2*. The structure underlying this sentence is illustrated in (80).[36]

---

[36] Compare this structure with (47) in section 2.2, which differs only in the position of the trace in the lower spec-*vP*.

(80)   The structure underlying (79) *(cf. (47) in section 2.2)*

vP

vP

vP

t_{THEY} said   CP

...they'll   vP

vP

vP

...hire $t_{\text{RELATIVE OP.}}$ $t_{\text{RELATIVE OP.}}$ if I criticize him publicly

Adjunct

Adjunct

O_j in order to get me to praise $\_j$

$t_{\text{RELATIVE OP}}$
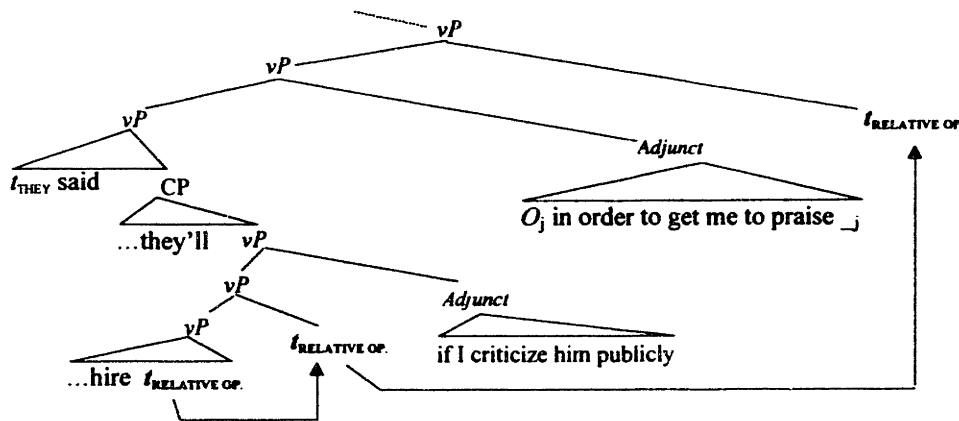
The asymmetric pattern of PGs that has emerged in these cases is rather striking. If PGs were licensed by some path-theoretic property of long-distance dependencies,[37] it would be difficult to explain. The asymmetry seems to reveal the varying positions of an intermediate trace. The presence of this abstract element—although it is unpronounced—yields observable effects via the choice of landing site to which it is assigned in a representation.

3.2.2   Extraposition from *wh*

The second test for the OPG is provided by an additional claim that was defended in Fox and Nissenbaum (1999).[38] F&N argued that relative clause extraposition (from non-subject NPs) marks the position of a covert movement. This movement is therefore just like HNPS, except that only part of the NP—the relative clause—is pronounced in the shifted position. If extraposition from a VP-internal *wh*-phrase behaves similarly—in this case, marking the position of the intermediate trace in spec-vP—we would expect contrasts similar to those above.

---

[37] Cf. the various proposals discussed in section 1.1 (Kayne 1983, Longobardi 1984, Pesetsky 1982, Sag 1983, Gazdar et al. 1985, Richards 1997, Steedman 1997).

72

As above, the results appear to bear out the prediction. Extraposition of a relative clause from a moved *wh*-phrase, *past* a clausal adjunct, forces the appearance of a PG. This is shown by the acceptability of the (a) examples of (81)-(82) as compared with the (b) and (c) examples.

(81) a.   Who did you praise _ [in order to impress _][that teaches literature at Oxford]

   b.   *Who₁ did you praise_ [in order to impress *him*][*that teaches literature at Oxford*]

   c.   *Who₁ did you praise _ [in order to surprise *me*][*that teaches literature at Oxford*]

(82) a.   ?What film₁ would you see_ [if I could get tickets for _][that John recommended]

   b.   *What film₁ would you see_ [if I could get tickets for *it*][*that John recommended*]

   c.   *What film₁ would you see_ [if it doesn't rain] [*that John recommended*]

On the other hand, an extraposed relative clause that appears *to the left* of a clausal vP-adjunct, as in (83)-(84), does not allow a PG but is acceptable without one.

(83) a.   *Who₁ did you praise_[that teaches literature at Oxford][in order to impress _]

   b.   *Who₁ did you praise_[that teaches literature at Oxford][in order to impress him]*

   c.   *Who₁ did you praise_[that teaches literature at Oxford][in order to surprise me]*

(84) a.   *What film would you see_[that John recommended][if I could get tickets for_]

   b.   What film₁would you see_[that John recommended] [if I could get tickets for it]

   c.   What film₁would you see_[that John recommended] [if it doesn't rain]

---

[38] That work appears in a slightly revised form as chapter 4 of this dissertation.

The pattern in (81)-(84) would be explained on the assumption that the extraposition is marking the site of an intermediate trace, together with the Obligatory PG generalization (63). And it would be quite unexpected otherwise.

3.2.3 Antecedent-contained deletion

The final argument for the Obligatory PG generalization (63) has to do with an asymmetry that arises when a raised *wh*-phrase embeds an *antecedent-contained deletion* (ACD). Each of the *wh*-phrases in (85)a,b contains a deleted vP (an ellipsis) that is interpreted as identical to an overt antecedent vP (Sag 1976 and much subsequent work) from which the *wh*-phrase was in turn extracted. The antecedent vPs in (85) are modified by adjuncts which contain PGs, and in both cases the ellipses are ambiguous in a way that we might expect. Specifically, the antecedent for the ellipses can be either the whole vP including the adjunct [$_{vP}$[$_{vP}$ ...][$_{adjunct}$ ...]], or just the portion of the vP that excludes the adjunct:[39]

(85)   **Parasitic gap in the adjunct ⇒ ambiguous ellipsis**

    a.   Which of the books that I ASKED you to ＿did you [[file ＿][without (actually) reading ＿]]?
           = "...*that I ASKED you to file*"
      or = "...*that I ASKED you to file without reading*"

---

[39] These two interpretations are apparently easiest to elicit by means of different prosodic structures. The narrow ellipsis (without the adjunct) is greatly aided with a pronunciation in which the adjunct is set off as a separate intonational phrase, with a slight pause after the main verb phrase: "...did you file, without actually READING?" whereas the wider ellipsis seems most natural when the main verb phrase and adjunct are run together in the same intonational phrase.

b. How many books that he wanted to __ did John [[file _][without (actually) reading _]]?

= "...*that he wanted to file*"

or = "...*that he wanted to file without reading*"

When there is *no* PG in the adjunct, however, the ellipsis has only one interpretation—the adjunct cannot be part of the ellipsis:

(86)   **No Parasitic gap in the adjunct** ⇒ *only narrow ellipsis*

a. Which of the books that I ASKED you to__ did you [[file _][without (actually) reading them]]

= "...*that I ASKED you to file*"

≠ "...*that I ASKED you to file without reading them*"

b. How many books that he wanted to __ did John [[file _][without (actually) reading them]]?

= "...*that he wanted to file*"

≠ "...*that he wanted to file without reading them*"

The sentences in (86) are quite unnatural except under the interpretation indicated (which is easiest to elicit by means of the prosodic pattern described in footnote 39). Similar cases where the vP-adjunct contains neither a gap nor a pronoun pattern with the latter. Only the 'narrow' ellipsis (excluding the adjunct) is available:

(87) a. Which of the books that I ASKED you to __ did you file _ without talking to the boss first
= "...*that I ASKED you to file*"

≠ "...*that I ASKED you to file without talking to the boss first*"

b. How many books that he wanted to __ did John file _ without talking to the boss first
= "...*that he wanted to file*"

≠ "...*that John wanted to file without talking to the boss first*"

This asymmetry looks very much like another instance of an environment in which a PG is obligatory. In particular, the missing readings of (86)-(87) are apparently unavailable because a PG is for some reason necessary in order to license the wide ellipsis $[_{vP}[_{vP} ...][_{adjunct} ...]]$.

While these facts do not follow from the Obligatory PG generalization alone, they would follow if we could motivate an additional assumption: namely that the intermediate trace in all of these constructions needs to be a full copy (in the sense used in section 2.2.1).

(88)   **Supplementary assumption:** The intermediate traces in (85)-(87) are full copies.

What would be striking about this additional assumption is that ACD is well-known for a nearly opposite property. That is, the parallelism condition on ellipsis forces (at least the lowest of) the copies left by movement to be a 'simple' trace (Fox 1995,2000a).[40]

Nonetheless, (88) together with the Obligatory PG generalization would explain the asymmetric pattern of ACD interpretation. The structure of the vPs in (85) and (86) would (by the Obligatory PG generalization) have to be (89)a and (89)b respectively.

(89) a.
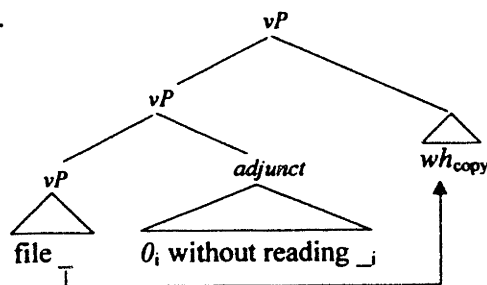
b.
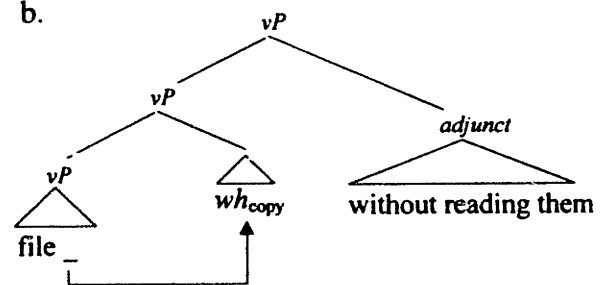
Given the supplementary assumption (88) that the intermediate positions in these structures need to be full copies of the *wh*-phrase containing the ACD, the result will be differing options for the choice of antecedent for the ellipsis. The parallelism condition dictates that the antecedent must

be free of any full copy containing the ellipsis. In the structure containing the PG (89)a, parallelism can be satisfied by either choice of antecedent vP, since neither one contains the intermediate copy. This is illustrated in (90)a.

(90) a.
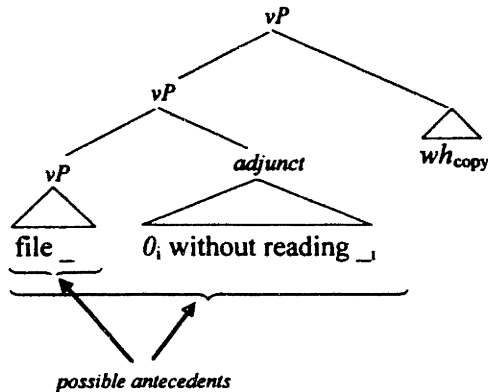
b.
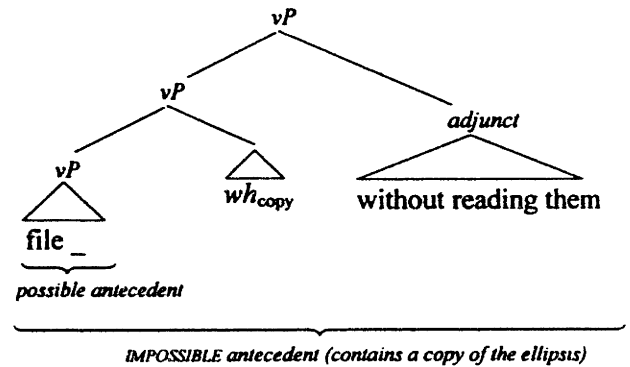
In the structure *lacking* a PG (89)b, however, the only choice of antecedent that would satisfy parallelism is the smaller one, not containing the adjunct (illustrated in (90)b). The larger candidate, which contains the adjunct, also contains the intermediate copy of the *wh*-phrase. Hence the ellipsis in this structure could not be interpreted as parallel to that antecedent.

This account of the asymmetric pattern of (85)-(87) allows us to view it as another case of obligatory parasitic gap, falling together with Larson's generalization and the other cases discussed in this section. As attractive as that would be, however, the account is suspect on the grounds that it requires the curious supplementary assumption (88), as noted above. If we could independently corroborate that (88) is correct and discover its explanation, we would have an extremely powerful argument for the Obligatory PG generalization and the hypothesis that PGs are licensed by a successive-cyclic vP-step of movement.

---

[40] See Sauerland 1998, Merchant 2000 for arguments that this view has to be modified somewhat.

I would like to propose that we can go part way toward such an explanation. While I don't know of a principle that would derive (88), I will argue that it reduces to the following generalization, for which independent evidence can be given:

(91)   **ACD generalization:** Relative clauses with ACD need to be present in the copy that is in a spec-position of the antecedent vP.

If (91) can be shown to hold, then (88) is just an instance of this generalization. The independent evidence for (91) has to do with Condition C. As noted in section 2.2.1, A-bar movement normally bleeds Condition C when the potentially offending r-expression is embedded in a relative clause (Freidin 1986, Lebeaux 1988). This fact is demonstrated by the acceptability of co-reference in (92).

(92)   [Which of the books that John, PROMISED you] did he$_i$ (actually) bring?

This fact entails that all of the copies left by the movement in (92) (i.e. the copy at the tail of the chain and the intermediate copy in spec-vP, if it exists) are interpreted as simple traces, not containing the relative clause:[41]

(93)   [Which of the books [that John$_i$ PROMISED you]] did he$_i$ actually $\langle t \rangle$ [bring $\langle t \rangle$]

However, consider (94). This sentence differs from (92) only in having ACD where (92) has an ordinary relative clause gap. Yet co-reference seems much more unnatural in (94). For some reason the movement doesn't bleed Condition C.

---

[41] I have put the intermediate spec-vP copy on the left in this and the subsequent examples, departing from the convention used in previous examples. As noted above, the linear ordering of this copy and its target is irrelevant for present purposes.

78

(94) ??[Which of the books that John$_i$ PROMISED he would__] did he$_i$ actually bring?

This fact, which to my knowledge has not been noted in the literature, would follow from (91). That generalization requires that a full copy containing the ACD be present in the intermediate spec-vP position, since the vP in question is the antecedent for ellipsis. But as shown in (95), a full copy in the intermediate position would leave an r-expression in the scope of the subject pronoun, yielding a Condition C effect:

(95) [Which of the books [that John$_i$ PROMISED he would_]] did he$_i$ actually ⟨which of the books that John$_i$ PROMISED he would_⟩ [bring ⟨t⟩]
$$\underbrace{\qquad\qquad\qquad}_{\text{Antecedent vP}}$$

That this is the correct sort of account for the contrast between (94) and (95) is suggested by the following controls. (96) is a standard control for Condition C; it is the same as (95) but with the positions of the r-expression and the pronoun reversed so that no violation would result from reconstruction. As expected, co-reference is perfectly natural.

(96) [Which of the books that he$_i$ PROMISED he would__] did John$_i$ actually bring?

The contrast in (97) further supports the contention that the ACD generalization (91) prevents the movements in these examples from bleeding Condition C.

(97)a. [Which of the books that John$_i$ ASKED you to__] did you remind him$_i$ to bring?

  b. ??[Which of the books that John$_i$ PROMISED he would__] did you have to remind him to bring?

The acceptability of (97)a is predicted by (91). Movement in this sentence is long-distance, targeting two vP landing sites. Since one of the vP targets is higher than the indirect object

pronoun *him*, (91) can be satisfied without violating Condition C. As shown in (98) none of the copies in the scope of the pronoun is full.

(98)   [Which of the books that John$_i$ ASKED you to__] did you 〈which of the books that John$_i$ ASKED you to__〉 [remind him$_i$ to 〈 *t*〉 [bring 〈 *t*〉]

Antecedent vP

However, the consequence is that the matrix vP is the only possible antecedent for the ellipsis; by (91), the vP in (98) whose specifier is a full copy needs to be the antecedent. This is indeed the only natural interpretation of (97)a. A structurally similar counterpart, (97)b, is pragmatically restricted to the reading on which the *lower* vP is the antecedent. Once again, the condition C violation re-appears—exactly as we would expect given the LF (99) that is mandated by the ACD generalization (91).

(99)   [Which of the books [that John$_i$ PROMISED he would__]] did you ... remind him$_i$ to 〈which of the books that John$_i$ PROMISED he would__〉 [bring 〈 *t*〉]

Antecedent vP

These examples give us reasonable grounds for supposing that (91) or a similar generalization holds. While the generalization does not appear to follow from principles known to be involved in the licensing of ACD, the facts seem to me robust and would be difficult to explain without it. I will therefore tentatively assume that it is correct.

Given this assumption and the facts that motivated it, we end up with another strong argument for intermediate traces in spec-vP (which may be full or partial copies as determined by independent factors). Moreover, as noted earlier, the assumption of (91) gives us a powerful argument for the Obligatory PG generalization (63), as these two generalizations were shown to explain the asymmetric pattern of interpretation in (85)-(87).

In fact, we are now in a position to further corroborate that explanation by means of a control example modeled on (97)a. Consider the contrast between (86)b, repeated below, and (100).

(86) b. How many books that he wanted to __ did John [[file _][without (actually) reading them]]?

= "...*that he wanted to file*"

≠ "...*that he wanted to file without reading them*"

(100) How many of the books that he promised he would__ did John [(IN FACT) remember to file
_ without reading them]?

Both of these examples have vP-adjuncts without PGs. Yet (100), unlike (86)b, allows an interpretation in which the adjunct is part of the ellipsis. I argued above that the unavailability of this reading for (86) follows from (91) together with the Obligatory PG generalization; the latter requires that the intermediate trace be positioned structurally below the adjunct, and the former requires that it be a full copy. The result is that the parallelism condition on ellipsis is only satisfied if the adjunct is not part of the antecedent. But the prediction is different for (100). This sentence, like (97)a, involves movement through two clauses, each of which provides a vP landing site. As a result, (101) is a possible LF:

(101) [How many of the books that he promised he would __] did John ⟨**How many of the books that he promised he would__**⟩⟨remember...to ⟨ *t* ⟩ [file ⟨ *t* ⟩][without reading them]]⟩

Antecedent vP

In this LF, only the copy in the higher of the two spec-vP positions is full. By (91), the higher vP qualifies as a possible antecedent for the ellipsis. Since this vP contains the adjunct, the adjunct can (and must) be interpreted as part of the ellipsis.

The upshot of this discussion is, first, that a generalization governing the LFs of ACD constructions has emerged (namely (91)). Second, given this generalization a rather intricate and

surprising set of predictions resulted, yielding evidence for the existence of traces in spec-vP that seems difficult to refute, and bolstering the claim that the presence or absence of a PG diagnoses the structural positions of these traces.

.

4.   Summary of main results

The facts discussed in this chapter allow us to draw some interesting conclusions about the properties of grammar. First, the paradigms that emerged from the tests in section 3.2 support quite strikingly the hypothesis that long-distance movement targets vP successive-cyclically, and that it is this intermediate trace that is crucial for PG-licensing. If movement to spec-CP were alone sufficient, the deviant examples in (73) and (83)-(84), and the missing readings of (86)-(87) would be a mystery. Second, these considerations provide a strong argument that the movement relation is correctly characterized as a relation between discrete positions in a chain, rather than as a geometric property of paths in a tree structure. Path-theoretic conceptions of movement proposed by Kayne 1984, Pesetsky 1982, and Richards 1997, and the similar slash-feature percolation conception advocated by Gazdar, Klein, Pullum and Sag 1985, do not have a way of predicting the asymmetries discussed in this section. The asymmetries suggest that the precise position of the intermediate trace is what matters: an adjunct along the path of a long-distance dependency *cannot* contain a PG unless it can compose semantically with the predicate of the intermediate movement. Finally, this intricate and surprising pattern of facts strengthen the already solid case for Chomsky's null operator hypothesis, a crucial assumption underlying the predictions with which this chapter has concluded.

An equally important result of the discussion so far is the emergence of a potentially useful technique for detecting invisible aspects of structure. Pronunciation and word order cannot

give us direct evidence about the structural positions of empty categories in non-thematic positions—let alone their existence. But I have tried to demonstrate that we can detect the effects of intermediate traces by indirect means, via the presence or absence of parasitic gaps in vP-modifiers. In the next chapter we will put this tool to use in a different way: We will investigate the properties of another type of invisible position in a chain, namely those positions created by covert movement.

Appendix 1.   Subject parasitic gaps

Evidence that subjects with PGs always reconstruct

(a.) Variable binding from the raised position

If subject QPs need to reconstruct in order to license a PG, then we would expect a variable that can only be bound from the raised position to create a conflict with a PG in the subject. This expectation is borne out in (1). (1a) is an acceptable instance of a PG contained in a QP that has undergone surface A-movement over the raising predicate *appear*. (1b) is the test case: the PG becomes unacceptable when the QP subject containing it needs to bind a variable from the raised position. (1c) is a further control, showing that the QP is able to bind a variable from its raised position if it doesn't contain a PG. (Example (1d) shows that binding by a PG subject is no problem if the variable can be bound from the reconstructed position)

(1) a.   Sue's the kind of person that [everyone who talks to _] appears to my colleagues to like _

   b. * Sue's the kind of person that [everyone who talks to _]$_i$ appears to his$_i$ colleagues to like _

   c.   Sue's the kind of person that [everyone I know]$_i$ appears to his$_i$ colleagues to like _

   d. *cf.* Sue's the kind of person that [everyone who talks to _]$_i$ appears to want his$_i$ colleagues to meet _

84

(b.) Condition C violations resulting from reconstruction

Obligatory reconstruction should also create Condition C effects.[1] (2)-(3) bear out this prediction. The (a) examples are the control cases, with the r-expression and the co-indexed pronoun positioned such that reconstruction would not violate condition C. The test cases are the (b) sentences, and as predicted they are very much degraded in acceptability. The (c) sentences are identical to those in (b) except that the pronoun is embedded so as not to c-command the reconstructed r-expression, providing a further control to show that Condition C is the relevant factor.

(2) a. Mary's the one that [his$_i$ constant criticism of _] seemed to John$_i$ to have (finally) upset _

  b. * Mary's the one that [John's$_i$ constant criticism of _] seemed to him$_i$ to have (finally) upset _

  c. Mary's the one that [John's$_i$ constant criticism of _] seemed to [his$_i$ mother] to have upset _

(3) a. That's the kind of film that [people who recommend _ to her$_i$] usually strike Mary$_i$ as liking _ for the wrong reasons

  b. * That's the kind of film that [people who recommend _ to Mary$_i$] usually strike her$_i$ as liking _ for the wrong reasons

  c. That's the kind of film that [people who recommend _ to Mary$_i$] usually strike [her$_i$ husband] as liking _ for the wrong reasons

(c.) Scope interactions with modals

The first two tests were grammaticality judgment paradigms. The examples in (4)-(5) make use of ambiguities created by QP subjects and modal verbs. Ordinarily such sentences are

---

[1] On the assumption that binding conditions are sensitive to scope reconstruction, as expected under Chomsky (1993), and argued for empirically by Fox (1999,2000a) and Romero (1997).

ambiguous between a meaning where the surface scope relation is preserved, and an inverse scope interpretation resulting from optional subject reconstruction. (4a) illustrates the ambiguity by means of two disambiguating follow-up clauses. (4a) may be felicitously followed up by (i) in a context where *anyone* from the department would be sufficient to achieve the desired result, whereas professors from other departments would not sufficiently draw attendance if they taught the topic. This is the interpretation that results from subject reconstruction below *needs*; the interpretation can be paraphrased as "It needs to be the case that someone from our department teaches the topic..."

(4) a.   This is the topic that someone in our department needs to teach...
           (i) ... if we want people to show up   *(subject takes narrow scope)*
           (ii)... if the guy ever wants tenure   *(subject takes wide scope)*

If, on the other hand, (4a) is followed up by (ii), the sentence is disambiguated toward a wide scope reading for the subject. With this meaning, the sentence cannot be paraphrased as above, but only as "There's someone in our department such that he needs to teach this topic..."

A parasitic gap in the subject disambiguates such sentences, allowing only the narrow scope reading:[2]

(4)b.   This is the topic that someone who's (just) written about _ needs to teach _ (if we want people to show up).

   c.   * This is the topic that someone who's (just) written about _ needs to teach _ (if the guy ever wants tenure)

---

[2] The contrast in (4) is not due to the so-called "specific" interpretation of the indefinite in (4c). If we further embed the indefinite under the scope of a universal quantifier, as in (i)-(ii), the indefinite in (ii) can lose its specific character yet the contrast remains:
(i)   This is the topic that every dean thinks someone who's (just) written about _ needs to teach _ (if we want people to show up).
(ii) *This is the topic that every dean thinks someone who's (just) written about _ needs to teach _ (if the guy ever wants tenure)

The same logic is used in (5). (5a) has two versions. Disambiguating follow-up sentences are provided, and in addition each of the two interpretations for the sentence comes with a characteristic intonation pattern, very crudely reflected by upper- and lower-case letters. The narrow scope (i) states that "it must happen that no one leaves," and the wide scope (ii) merely asserts that there is no particular individual who must leave.

(5) a.   (i)   No one must LEAVE. (If anyone does, there will be a severe penalty) *(narrow scope)*

      (ii)   NO one MUST leave. (But everyone should feel free to)   *(wide scope)*

It is important to note that the narrow scope reading of (5a) is the more marked reading. It is therefore all the more surprising that this is the only reading that is available when a PG is put in the subject, as shown by the contrast between (b) and (c). This is so no matter what intonation pattern is used. Once again, the facts provide striking support for the prediction that subject PGs are licensed only under reconstruction.

(5) b.   John's the guy who no one that insulted _ must talk to _ (or he'll be really upset) *(narrow)*

    c.   # John's the guy who NO one that insulted _ MUST talk to _ (but everyone should feel free to)   *(wide)*

(d.) Interactions with other scope-bearing predicates

The same test as above can be replicated with other kinds of predicates besides modal auxiliaries. The indefinite subject of (6) can take either narrow or wide scope with respect to *likely*. The narrow scope version (6a) is felicitous in a context where *you* refers to an individual who is an inside candidate for the job (and thus has a good chance at getting it). The wide scope (6b) is only felicitous in a context where *you* is not an inside candidate (and thus has a poor

chance). Here, as in (5), the two meanings have characteristic pronunciations (indicated crudely by capital letters).

(6) a. That's the job that you've got a decent shot at _ because an inside candidate is LIKEly to get _ *(Presumption: "you" are an inside candidate)*

b. That's the job that you shouldn't even bother with _ because an INSIDE CANdidate is likely to get _ *("You" are not an inside candidate)*

Again, using the condition set up in (6), we can test the reconstruction hypothesis by putting a parasitic gap inside the QP. And as the hypothesis predicts, the sentence is disambiguated in favor of narrow scope reading:

(7) a. That's the job that you've got a decent shot at _ because an inside candidate for _ is LIKEly to get _

b. * That's the job that you shouldn't even bother with _ because an INSIDE CANdidate for _ is likely to get _

(e.) **Anti-c-command condition must be defined with respect to thematic positions:**

Extraction is marginally possible from the experiencer argument position of the raising predicate *strike* (as in 8a). But this extraction is completely unacceptable with subject PG if the subject has raised past the site of the gap (8b). On the other hand, a PG in a raising subject is licensed if the extraction is from a position *lower* than the subject's thematic position (8c).

(8.) a. ?Which bureaucrat did you say that friends of mine occasionally strike _ as a nuisance?

b. * Which bureaucrat did you say that friends of _ occasionally strike _ as a nuisance?

c. ?Which bureaucrat did you say that friends of _ occasionally strike you as nicer than _?

These facts are explained on the assumption that the subject PG is licensed under reconstruction. In 8b, reconstruction would place the subject lower than the site of the licensing gap—a violation of the anti-c-command condition. (I.e. the licensing movement does not pass through any projection of which the subject is in a specifier.)

To control for the possibility that 8b is ruled out for some reason having to do with the marginality of extracting the dative argument, consider 9. This example seems to share the somewhat marginal status of 8a. The conclusion must be that in principle, extraction of the dative argument is capable of licensing a subject PG—as long as the subject receives its thematic role in a higher *vP* (a *vP* targeted by the licensing movement).

(9.)   ? Which bureaucrat do friends of _ say that you occasionally strike _ as a nuisance?

Appendix 2. *HNPS* is *HNPS:*
Arguments against *'light-predicate raising'*

Larson (1988a) proposed an alternative analysis of HNPS in which, rather than movement of a DP as I have been assuming, the construction involves raising of a predicate to the left of a stationary DP. Larson argued that a consequence is that PGs are predicted to appear when an predicate-adjunct (formed by null operator raising) undergoes 're-analysis' with a transitive verb, as in (2)

(1)   John filed without reading a recent article about global warming.

(2)   **the structure Larson (1988a) claimed for (1)**
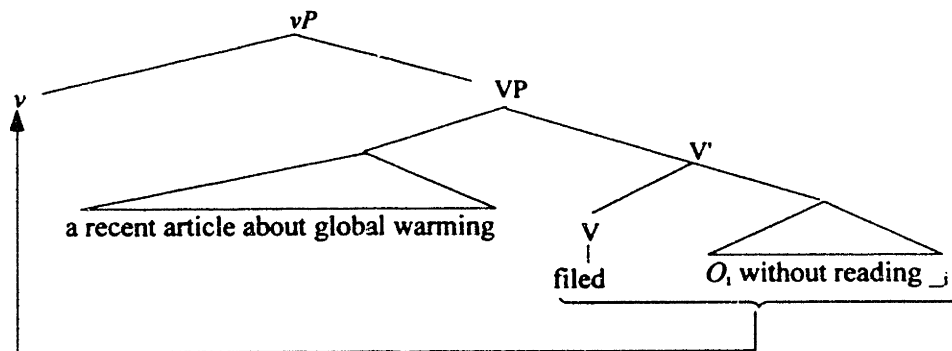
We saw in sections 2 and 3 that PGs can appear in an adjunct that modifies an embedding verb phrase, even if the licensing extraction is from the embedded clause. This was demonstrated with rationale clauses and conditional *if*-clauses, as in (3)-(4), which can only be construed as modifying the matrix *v*Ps.

(3)a.I [claimed that I liked _][in order to get you to rent _] that movie with Fred Astaire and
        Audrey Hepburn

   b. #I [liked _][in order to get you to rent _] that movie with Fred Astaire and Audrey
        Hepburn

(4)a. They'll [deny t! ᴌt they hired _][if we criticize _ publicly] the politician associated wi⁏⁏ the paramilitary.

   b. #They [hired _][if we criticize _ publicly] the politician associated with the paramilitary

Larson 1988a would apparently have to claim that a sentence like (3)a has the unlikely structure

(5), with the V' *claim that you liked* composing with the adjunct to assign a thematic role to *what*

movie.

(5)   **the structure Larson (1988a) would have to claim for (3)a**
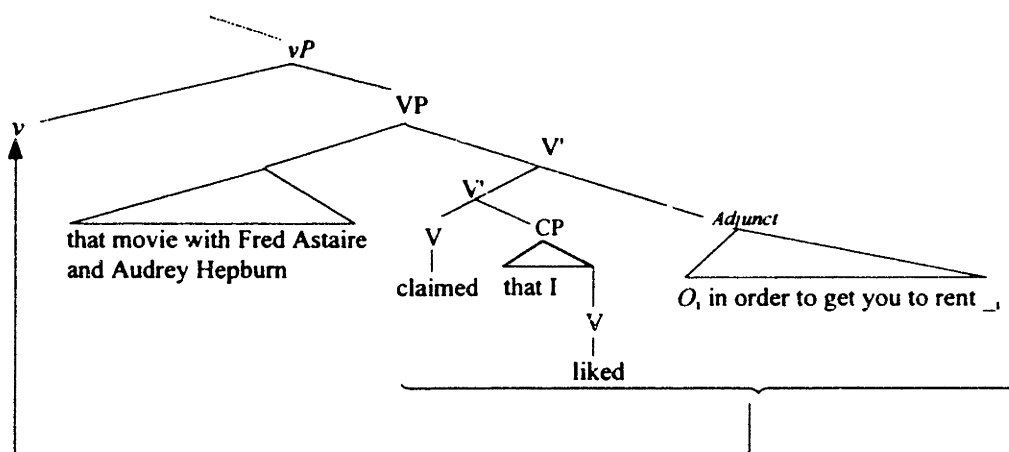
The claim that an embedding verb and its finite complement undergo reanalysis would be

suspect enough, but the structure is implausible for another reason. Given Larson's assumption

that there is no trace of the direct object after the verb that assigns its thematic role, the

embedded CP is uninterpretable—the clause contains a transitive VP lacking a direct

object—hence should not be able to compose with the embedding verb *claim*.[3] Larson's analysis

---

[3]   Alternatively, Larson might claim that there *is* a trace after the embedded verb—with an additional null operator raising in the main clause to take scope over the matrix V' as in (i). This structure, however, has something of an *ad hoc* flavor. The main clause is not an island for extraction, so it is unclear why a null operator would be invoked rather than ordinary extraction of the NP (as claimed by the traditional analysis of HNPS as movement of the DP.)

   (i)   [vp that movie with Fred Astaire and Audrey Hepburn [[v *O* claim that I liked *t*][*O* in order to get you to rent *t*]]

apparently predicts that HNPS should only license a PG if the adjunct directly modifies the verb from which the licensing 'gap' receives its thematic role.

The alternative view, with the adjunct a modifier of the full (clausal) vP, would not have to postulate this. The 'heavy NP' is assigned its thematic role as the direct object of *like*, and raises (perhaps successive-cyclically) targeting a position at the higher vP:

(6)   **Long-distance HNPS**
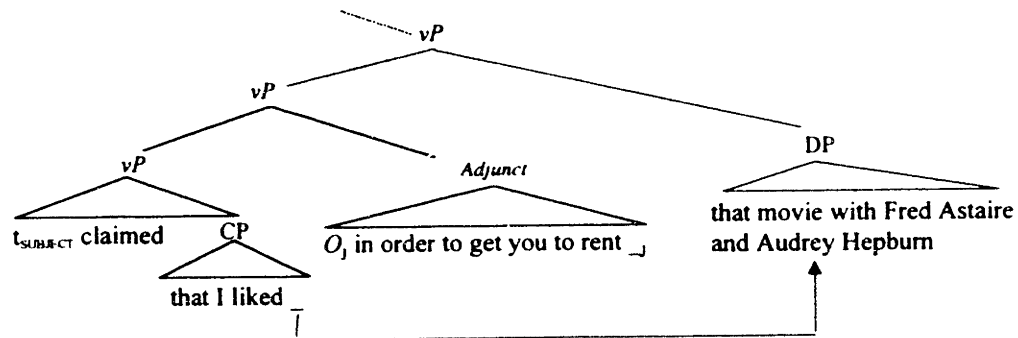
While this structure is a counterexample to the (unexplained) generalization that HNPS doesn't apply long-distance, it does not suffer from uninterpretability and hence is a more plausible candidate than (5).

Examples such as these, which apparently involve long-distance HNPS, therefore present a powerful challenge to Larson's (1988a) analysis of HNPS as 'light-predicate raising.'

Further cases are given below:

(7)   John's the guy they [said they'll [hire _ [if I criticize _ publicly]]][in order to get me to praise _]

(8)

(9)a. Mary [claimed that she liked _][in order to get me to see _] that movie with Fred Astaire and Audrey Hepburn

  b. They [said they'll [hire _ [if I criticize _ publicly]]][in order to get me to praise _] the man who rejected my proposal

These examples, too, are difficult to reconcile with Larson's *Light-predicate raising* theory. An additional control is provided by (10), which shows that nothing is wrong in principle with the sequence [adjunct 1—no pg][adjunct 2—pg]:

(10)   ? I claimed I would [hire _ *after talking to the dean*], [in order to get you to trust _], *the guy who was wearing fancy shoes*

(11)

Furthermore, other movements that target a position below the subject also license PGs. This fact casts doubt on Larson's claim that HNPS involves V' reanalysis.

(12)   short scrambling in German (Martin Hackl, p.c.)

   a.   ...weil der Hans   das Formular₁ [ohne vorher _ auszufüllen] _₁ abgeschrieben hat
           because the H.   the form   without first   out-to-fill   copied   has
           '...because Hans has copied the form without first filling (it) out '

   b.   ...weil der Hans   [ohne vorher es/ *_ auszufüllen]   das formular abgeschrieben hat
           bec. the H.   without first it   out-to-fill   the form   copied   has

(13)   short scrambling in Hindi (Rajesh Bhatt, p.c.)

       John-ne kaun-sii/koii kitaab [binaa _ paRhe] phe~k dii
       J-Erg which/some   book without it   reading throw
       'John filed which/some book without reading (it) '

(14)   quantifier movement in Icelandic (Jonsson 1996)

   a. ?Þeir hafa   ymsar bækur₁   lesið _₁ [án þess að gagnryna _ harðlega]
        they have various books   read   without criticizing   severely

   b. * Þeir hafa lesið ymsar bækur [án þess að gagnryna _ harðlega]
        they have read various books without criticizing   severely

(15)
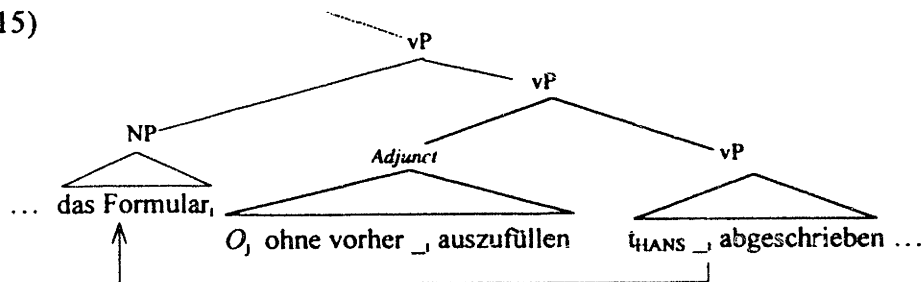
Long-distance scrambling to a matrix *vP* position licenses a PG in a matrix adjunct:

(16)   Hindi (Rajesh Bhatt, p.c.)

Mohan-ne   vo kitaab [binaa __/??use paRhe] kahaa ki   John pasand karegaa
Mohan-Erg that book [without _ / it   reading] said   that John like-will
'Mohan [that book][without reading _] said that John will like _'

(17)   Farsi (from Vahedi 1994)

man in ketaab raa   [ghabl az   in-ke _ be-xaan-am] fekr-kardam ke Ali _ nevešte
I   this-book-SPECIFIC [before   this-that _ SUJ-read-1s] thought   that Ali _ has written
'I [this book] [before reading _] thought that Ali has written _'

Chapter 3. An apparent 'Y-model effect':
Covert movement and parasitic gaps

In the previous chapter I argued that parasitic gap (PG) constructions, and more generally, sentences with clausal vP-modifiers, provide a useful means of detecting unpronounced structure. In particular, I argued that an investigation of these environments provides evidence for intermediate traces of successive-cyclic movement. The typically free alternation between vP-adjuncts that contain PGs and those that do not, I tried to show, is a reflex of the availability of more than one landing site for the intermediate movement (namely above and below the adjunct, respectively). Once the choice of landing site is fixed, so is the presence or absence of a PG.

This chapter will make use of those results to investigate the structures created by *covert* movement. The focus of the investigation will be a generalization about the distribution of PGs that has resisted explanation, namely that they are not licensed by covert movement. Since the earliest research on PGs at the close of the 1970s, it has been taken for granted that *overt* dependencies—those involving overtly dislocated phrases and their gaps—are necessary to license them. As Engdahl observed in her seminal work on PGs (published as Engdahl 1983), "It appears to be the actual presence of a real gap that licenses a parasitic gap and not just the presence of a *wh*-phrase" (1983:22). Engdahl illustrated this generalization with the contrast in (1):

(1) **Engdahl's generalization:** *PGs are not licensed by covert movement*

    a.   Which article₁ did you file __₁ [without reading __pg]

    b.   * Who filed *which article* [without reading __pg]

I will show that Engdahl's generalization is only partially correct. The failure of covert movement to license PGs in the standard paradigms, I will argue, reflects a failure of the experimental paradigms rather than a property of covert movement. I will show that once a appropriate paradigm is used, covert movement does in fact license PGs. In addition to reversing a longstanding empirical generalization, this result bears on fundamental questions concerning the architecture of grammar. In particular, it provides strong evidence for both the existence of covert movement and for a model of grammar that does not rely on levels of representation other than those accessed by external cognitive systems—an 'interface theory' in the sense of Chomsky 1993. PGs have remained an obstacle to such a theory, because their apparent limitation to overt movement environments has been difficult to characterize without reference to a level of s-structure. Several recent attempts have been made to account for Engdahl's generalization without the stipulation of an s-structure condition (Kim and Lyle 1995, Nissenbaum 1998a,b, Nunes 1995). The facts I will present constitute an empirical demonstration that they are licensed at LF.

In addition, I will argue that the different behavior of overt and covert movement with respect to PG licensing (i.e. the fact that Engdahl's generalization is largely correct) supports a traditional claim about the organization of grammar: namely that the difference between overt and covert movement itself reduces to a derivational sequencing imposed by the point at which *spellout* applies (2). This specific claim was part of the model proposed in Chomsky 1993 (3) and has its roots in the Y-model that was motivated by the s-structure hypothesis.[1]

---

[1] In Chapter 4 I will give an argument (based on joint research undertaken with Danny Fox) that this picture cannot be correct; a different model of grammar that resolves the conflict will be motivated in Chapter 5. The central proposal will be to embed the claim (2) in a cyclic theory of spellout, with the result that 'Y-model effects' largely

(2)   **The difference between overt and covert movement is due to sequencing:**
Operations that precede *spellout* are overt, while those that follow it are covert

(3) a.   **The 'Y-model' (as revised in Chomsky 1993)**
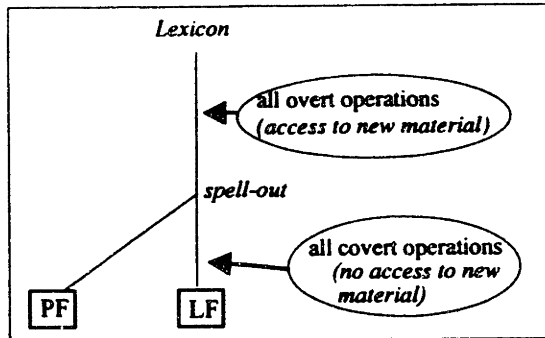
```
          Lexicon
            |
            |<──── (all overt operations
            |       (access to new material))
            |
  spell-out |
     \      |
      \     |<──── (all covert operations
      [PF] [LF]    (no access to new
                    material))
```

The argument will have the following form. If it could be shown that the configuration in which PGs are licensed can only be derived by *overt* movement, we would have an explanation for Engdahl's generalization. The first step of the argument (section 1) consists of motivating a general condition on movement: namely, that second (and subsequent) movements to a single projection do not extend the tree, but rather 'tuck in' as argued by Richards (1997). A consideration of multiple *wh*-constructions in Bulgarian will illustrate the 'tucking in' property of movement. I will argue that this condition blocks movement from forming the structural configuration needed for licensing a PG (cf. Chapter 2), on the grounds that the vP-step of movement would be required to tuck in below the adjunct. Consequently, that configuration can only be derived in the 'reverse' order, with movement targeting the vP *before* insertion of the adjunct (to the position between the mover and its target.

Crucially, the second of these two steps is necessarily overt: insertion of the adjunct feeds pronunciation. If this is the only derivation that licenses PGs, it would follow from the Y-model

___

carry over. For the remainder of the discussion in this chapter I will keep to more familiar picture and continue to speak in terms of the 'Y-model.'

that covert movement is incapable of licensing PGs; a covert operation cannot be followed by an overt one.

This account makes an interesting prediction for cases of multiple extraction. I will show in section 2 that a closer examination of the Bulgarian multiple-*wh* facts suggests an unexplored consequence of the *Tucking-in condition*. An ordering puzzle noticed by Bošković (1995), I will argue, reveals that additional movement possibilities are created when the structure provides more than one position below which a phrase could tuck in. The PG-licensing configuration provides such an environment. Once that configuration has been formed, a *subsequent* movement should be able to raise past the adjunct as long as it can tuck in below the phrase that moved first, forming the structure [$_{vP}$ *WH-1* [$_{vP}$ *WH-2* [$_{vP}$ adjunct [$_{vP}$...t_1...t_2...]]]]

A predicted consequence, given the Obligatory PG generalization defended in Chapter 2, is that two PGs will be forced to appear in the adjunct (one licensed by each raised XP). Since the Y-model prohibits only the first of these two movements from being covert, the account leads us to expect that a second PG can be licensed by a covert movement. Once we look for such cases, we do indeed find PGs licensed by *wh*-in-situ:

(4) a. ?Which senator$_1$ did you persuade _$_1$ to borrow *which car$_2$* [after getting an opponent of _$_{pg1}$ to put a bomb in _$_{pg2}$]?

    b. * *Which senator$_1$ did you persuade _$_1$ to borrow which car$_2$ [after putting a bomb in _$_{pg2}$]*

(5) a. ?Which kid$_1$ did you give *which candy bar$_2$* to _$_1$ [without first telling a parent of _$_{pg1}$ about the ingredients in _$_{pg2}$]?

    b. * *Which kid$_1$ did you give which candy bar$_2$ to_$_1$ [without looking at the ingredients in _$_{pg2}$]*

1.   Deriving Engdahl's generalization

In Chapter 2 we saw evidence that PGs are licensed by a local DP that has raised to an outer vP-specifier position (just) above the predicate-adjuncts that contain them. The licensing DP can be an overt category—as in HNPS—or it can be the intermediate trace of a long-distance movement. In either case, the null operator structure containing the PG composes semantically with the vP to form a modified predicate of the raised DP. We can refer to this configuration as a 'modified derived-predicate configuration' (or 'modified-predicate configuration' for short):

(6) **Modified-predicate configuration:**
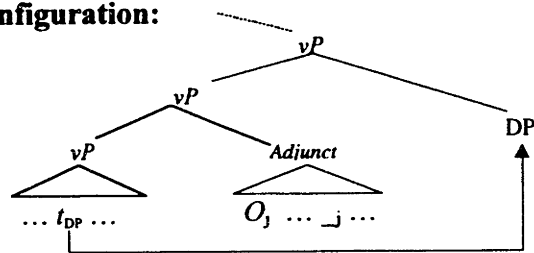
If it could be demonstrated that this configuration can only be derived by overt movement of DP, we would derive Engdahl's generalization that only overt movement licenses PGs. To the extent that this generalization is correct—as it appears to be for at least standard cases—a theory of grammar ought to explain it. The goal of this section is to argue that Engdahl's generalization is explained by a general constraint on movement that forces the modified-predicate configuration to be derived in the overt syntax. Evidence for this constraint comes from a consideration of multiple interrogatives in Bulgarian.

1.1   Bulgarian multiple-*wh* questions and the *Tucking-in* condition

Bulgarian is a multiple-*wh*-fronting language: all *wh*-phrases appear at the left edge of the interrogative clause (7). Moreover, the ordering of the two *wh*-phrases is rigid. The one whose base position is higher must appear first, as show by the unacceptability of (8).

(7)   Kogo kakvo e   pital Ivan   (Examples from Bošković 1995, cited in Richards 1997)
      whom what AUX asked Ivan
      '*Who did Ivan ask what?*'

(8)   ?*Kakvo kogo e   pital Ivan
      what   whom AUX asked Ivan
      ? * '*What did Ivan ask who?*'

Richards (1997), following Rudin (1985), interpreted the rigid order of the *wh*-phrases as evidence for the Superiority condition. On the assumption that the *wh*-phrases all target C^0 (forming multiple specifier positions of CP), *superiority* reduces to a locality condition on movement driven by a single attractor—a requirement that the closest *wh*-phrase must be attracted first.[2] Richards proposed that the parallel order of *wh*-phrases and their gaps is

---

[2] N. Chomsky (p.c.) points out that, although superiority effects are expected given locality of attraction, we would want to see independent support for their existence, given that the observations about English that originally motivated the condition (cf. Chomsky 1973) are problematic. Specifically, it is a fact that multiple-*wh* questions always require prominence on both (all) *wh*-phrases. Yet phonological prominence is generally sufficient for "d-linking exceptions" to superiority (cf. Pesetsky 1987, in press).

The following may provide the basis for an independent argument that superiority exists (as it is predicted to). Multiple *wh*-constructions allow focus-prominence to fall on other constituents in addition to the wh-phrases. Thus (i) is perfectly natural.

(i)   WHO do you think persuaded WHO [(foc) to talk to MARY]?

However, it seems to me that (for whatever reason) when superiority is violated, the 'skipped over' *wh*-phrase must bear the nuclear accent—everything to its right needs to be phonologically reduced or de-accented:
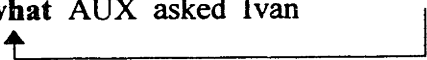
(ii) WHO do you think WHO persuaded to talk to Mary?
(iii) *WHO do you think WHO [(foc) persuaded to talk to MARY]?

This seems true no matter how wide or narrow the domain of focus. Similarly:

(iv) WHO did you persuade to give WHAT [(foc) to the man in the red SHIRT]
(v)   *WHAT did you persuade WHO [(foc) to give to the man in the red SHIRT]

If this observation is correct, it shows that violations of superiority require something extra that is not required generally of multiple-*wh*, and therefore have a more limited distribution.

explained by superiority (=attract closest) together with another assumption, namely that the

second movement 'tucks in' below the position of the first:

(9)   Kogo$_1$ kakvo$_2$ e   pital Ivan ... _$_1$ _$_2$ (Modified from Pesetsky *in press*)
        whom what AUX asked Ivan

(Second movement does not extend the tree)

Richards supported the observation that secondary movements tuck in with facts from a number

of other languages; it appears to be a phenomenon of some generality. One way that we might

state Richards' 'tucking in' generalization is as a preference for non-extending movement:

(10)   **Tucking-in condition (TIC):**
        Movement does not extend the tree if an alternative exists (it must tuck in *below the*
        outermost segment whenever possible).

There are good reasons for stating the TIC as a preference rather than an absolute principle,

although this move obviously raises questions. The *first* movement to spec-CP (movement of

kogo 'whom' in (9), for example) does not tuck in but rather raises past C^0. There are arguably

further cases of movement that are excluded from the TIC: on the assumption that HNPS and

intermediate steps of *wh*-movement target an *outer* spec-vP position (i.e. above the thematic

position of the subject), we would want to know why these movements, too, are excluded from

the TIC. The generalization may be that movement must extend past selected arguments

(complements and thematic specifiers), and that the TIC holds otherwise.[3] Whatever explains the

exceptions to it, cases that are known to fall under the TIC include multiple movements attracted

by a single head.

---

[3] For a speculation about what explains both the TIC and its exceptions, see section 4.

I will argue in the next sub-section that the TIC applies more generally this. Specifically, the TIC requires movement to tuck in not only below an outer specifier but also below an adjunct. Thus the TIC can block extension of the tree even when there is just a single movement.

1.2. 'Tucking in' explains why covert movements don't (normally) license PGs

I tried to show in Chapter 2 that parasitic gaps are licensed in the modified-predicate configuration illustrated in (6). Given the statement of the TIC in (10) it follows that (6) can be created *only* if movement of XP is overt. Suppose the derivation has progressed to the point of merging the vP with an adjunct, and that the next step is raising of an XP to a spec-vP position. Raising of XP past the adjunct will be blocked by the TIC; the XP will be forced instead to tuck in below the adjunct—a position from which it will fail to license the PG.

(11)
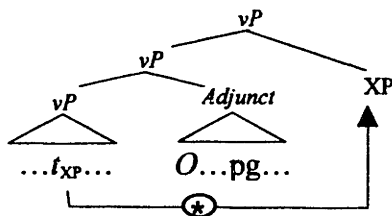
Tucking-In condition (10) prohibits movement of XP above the adjunct

If (10) will always block movement from forming the proper configuration, the question arises how a PG could ever be licensed. An alternative derivation provides an answer: Nothing blocks XP from raising to spec-vP *prior* to merger of the adjunct (12)a. The right configuration can then be created by merging the adjunct *below* XP (12)b.
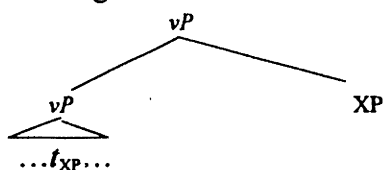
(12) a.   raising of XP   b.   merger of adjunct below XP
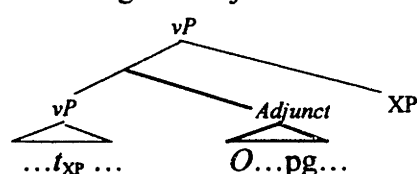
This derivation will work for sentences with overt movement. However, on the assumption that overt operations precede covert ones—an assumption forced by the Y-model—it will follow that an instance of covert *wh*-movement like that in (13) cannot have the derivation in (12). The reason is that merger of the adjunct is an overt operation—it needs to be spelled out—hence it must *precede* wh-movement. And then, as was the case in (11), the TIC (10) will force *what* to tuck in below the adjunct:

(13)   * Who filed *what* [without reading _]

(14)

CP
  who
        vP
      vP
    vP   *Adjunct*   ⟨*what*⟩
  ...filed *what* O_j without reading t_j
                        (*)
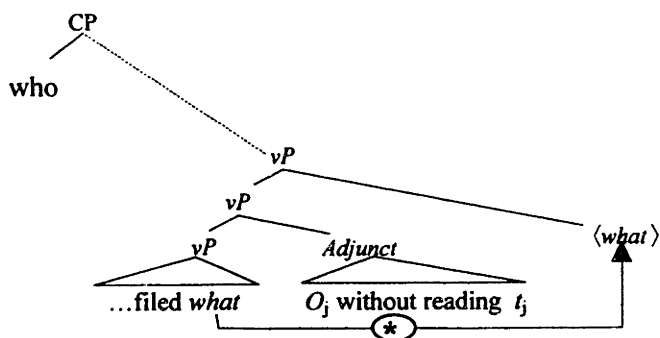
Consequently covert movement targeting vP cannot form the modified-predicate configuration. This interaction between a constraint on movement (the TIC) and a constraint on derivational sequencing (the Y-model) suffices to derive Engdahl's generalization without any reference to s-structure. If this account for the generalization could be independently supported, it would be preferable to accounts that rely on s-structure, since it assumes that PGs are interpreted by means of the same principles that enter into interpretation of other antecedent-gap structures. In the next section I will give an empirical argument that this is the correct way to understand Engdahl's generalization. The argument is based on a derivation that—if the explanation is correct—is predicted to yield a counterexample.

2. Predicting the cases where Engdahl's generalization fails to hold

I observed in the last section that a very general statement of Richards' *Tucking-in* condition would prohibit tree-extending movements past adjuncts and (non-thematic) specifiers. In this section I will provide further support for this way of understanding the TIC by showing that an interesting prediction is borne out. Namely, Engdahl's generalization is predicted to break down in a restricted domain of cases involving multiple *wh*-movement from inside vP.

2.1 A Bulgarian word order puzzle, and a simple solution

Bošković (1995, 1997) noticed an interesting word order puzzle in Bulgarian. While the order of two *wh*-phrases is rigid, the order of *more* than two *wh*-phrases is not. When a Bulgarian multiple interrogative contains three *wh*-phrases, Wh_2 and Wh_3 may be freely ordered.

(15) a.   [Wh_1 Wh_2 Wh_3 ...t_1...t_2...t_3]

    b.   [Wh_1 Wh_3 Wh_2 ...t_1...t_2...t_3]

(16) a.   Koj   Kogo   kakvo   e   pital   (examples from Bošković 1997)
        who   **whom**   **what**   AUX   asked

    b.   Koj   Kakvo   kogo   e   pital
        who   **what**   **whom**   AUX   asked
        'Who asked whom what?'

A number of proposals have been advanced to account for this fact. Bošković (1997) argued that only the highest of the *wh*-phrases moves to spec-CP; the others adjoin to a lower projection and are not constrained by Superiority. Richards (1998), following Brody (1995) proposed instead that a general property of grammatical dependencies allows constraints like Superiority to be ignored for second and subsequent movements (his Principle of Minimal Compliance). These

proposals have in common an assumption that the lesson from Bošković's word order puzzle is that the Superiority condition constrains only the first movement, and not the others.

However, we need not make that assumption. Another solution—one that does not assume Superiority is ever relaxed—is already provided by Richards' TIC if it is stated as in (10). Recall that the rigid ordering of two *wh*-phrases follows from Superiority and TIC because the second mover always tucks in below the first moved phrase. But notice that the *free* ordering of wh_2 and wh_3 would also follow from these two conditions. Superiority will always force wh_2 to move second, and TIC will force it to tuck in below wh_1. But the *third* mover—wh_3—will have two options. It can tuck in below both wh_1 and wh_2; alternatively, it can sandwich in between wh_1 and wh_2. Neither of these options would extend the tree. The two possibilities are illustrated in (17)b,c:

(17) a.   first two movements of (16)a, b
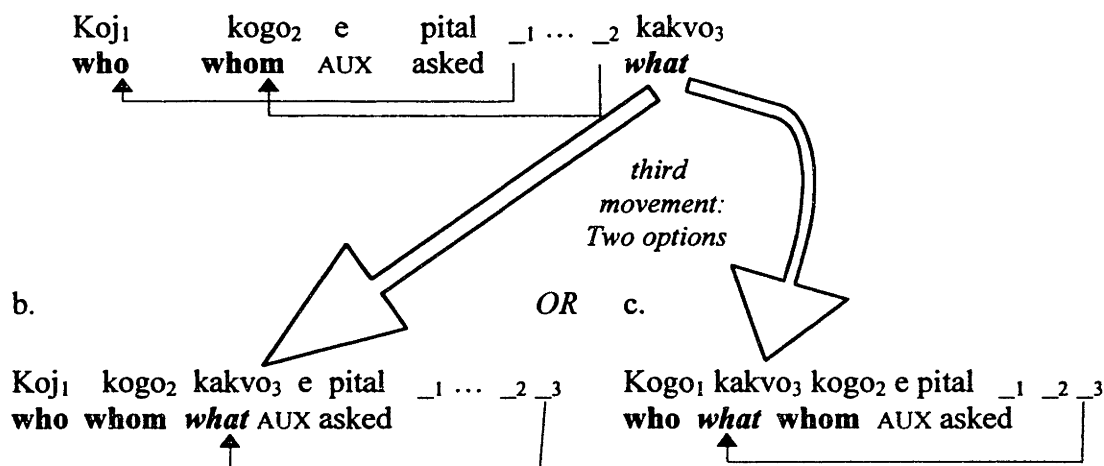
2.2   A 'Bulgarian strategy' for multiple parasitic gaps in English

In section 1, I argued that the TIC (10) underlies two very different phenomena: the parallel order of *wh*-phrases and their gaps in Bulgarian double interrogatives, and Engdahl's

generalization that covert movements don't license PGs. However, in the last sub-section we saw that the same condition predicts a case where the rigid ordering is relaxed in Bulgarian. I will now show that in exactly the same way, the TIC predicts a break-down of Engdahl's generalization. In short, it predicts that Engdahl's generalization will fail when a 'Bulgarian strategy' is possible.

Consider the possible derivations in the case where there are *two* wh-phrases internal to a vP modified by an adjunct—both of which will raise to form intermediate spec-vP positions (one of them overtly). Suppose the overt *wh*-movement precedes merger of the adjunct (to form a modified predicate configuration), as in (18)a. The derivation should have *two* possible continuations, parallel to Bulgarian triple-*wh*-questions. The *wh*-in-situ will be able either to tuck in below the adjunct, or to raise past it to a position below the outer specifier. These options are illustrated in (18)b,c.

(18) a.   First: Overt movement of *Wh*1 and then merger of the adjunct (just like (17)a)

b.   *Wh*2 tucks in below the adjunct   OR   c.   *Wh*2 tucks in below *Wh*1, but above the adjunct

The step illustrated in (18)c is the continuation of interest. If this derivation is chosen, then the adjunct will be positioned in between *two* moved phrases and the vP from which they have raised. In Chapter 2 I argued that a vP-adjunct positioned below a raised DP, as in (19), will obligatorily contain a PG. This generalization is repeated as (20).

(19)   **Modified-predicate configuration in which PGs are obligatory:**

(20)   **Obligatory PG generalization (OPG):**
A movement that targets vP will force the appearance of a PG inside a vP-modifier adjoined below the landing site.

According to (20), the adjunct in (18)c is forced to contain two PGs. Recall that the OPG was argued in Chapter 2 (section 2.1) -o reduce to the fact that the lowest vP segment is interpreted at LF as a derived predicate—in the case of (18)c, a two-place predicate. The adjunct in this structure therefore has to be a two-place predicate (formed by raising of two null operators), to allow predicate modification. We have already seen relatively acceptable examples of the kind of sentence predicted by this derivation, repeated below as (21)a,b. The structure underlying (21)a is shown in (22).

(21) a. ?Which senator$_1$ did you persuade $_1$ to borrow *which car$_2$*
                            [after getting an opponent of $_1$ to put a bomb in $_2$]?

   b. ?Which kid$_1$ did you give *which candy bar$_2$* to $_1$
                            [before having a word with $_1$ about the ingredients in $_2$]?

(22)

The relative acceptability of sentences like (21) would be hard to explain without the set of assumptions I have been making, in particular (i) that *wh*-in-situ undergoes covert movement; (ii) the movement is successive-cyclic, targeting vP; (iii) the resulting vP is interpreted as a derived two-place predicate that can be modified by a two-place predicate-adjunct; and (iv) PGs are licensed at LF.

Independent confirmation that multiple extraction from vP is able to create such structures is provided by Bulgarian. Like English, Bulgarian has PGs, as shown by the paradigm in (23); the gap inside the adjunct in (23)a is parasitic on a gap in the main clause (hence the unacceptability of (23)b) which cannot c-command it ((23)c).

(23)　　**PGs in Bulgarian:**　　　　　　　　(Roumyana Izvorski, p.c.)

a.　Koja　kola vze　nazaem _ bez　　da　　　　　vurnesh _?
　　which car　took on-loan _ [without SUBJUNCTIVE return _]
　　'Which car did you borrow _ [without returning _]'

b.　* Koja kola　putuva do Washington bez　　da　　　　popravi _
　　which car　travel　to Washington [without SUBJUNCTIVE fixing _]
　　* 'Which car did you travel to Washington [without fixing _]'

c.　* Koj　covek vze　nazaem kolata mi　sled kato　vidjah
　　which person took on-loan car　my　after when saw-1sg
　　* 'Which person borrowed my car [after I saw _]'

Bulgarian, being a multiple-*wh*-fronting language, thus provides a means of corroborating the existence of structures like (22). When a second *wh*-phrase is extracted from the VP, an additional PG is licensed in the adjunct, as shown in (24).

A second *wh*-phrase licenses a second PG   (Roumyana Izvorski, p.c.)

(24)   Koj senator$_i$ koja kola$_j$ vidjahte _$_j$ da otkradva _$_j$ bez da   mozhete da ubedite _$_j$ da vume _$_j$
         which sen. which car saw-2pl _   steal-3sg _ [without can-2pl convince-2pl _ retum-3sg _]

         'Which senator which car did you see (him) steal (it) [without succeeding to convince (him) to return (it)]'

Other than the fact that the second *wh*-movement is overt in the Bulgarian (24) and covert in the English (21), the structures underlying these sentences are identical. In the case of Bulgarian the existence of the second PG is not surprising. Given the above assumptions, neither should the existence of the second PG in English be surprising.

Returning to the two possible derivations illustrated in (18), recall that an alternative choice for the landing site of the covert movement is predicted to exist as well, reflected in (17)b. If this derivation is chosen instead, the adjunct will not be interpretable with two PGs. However it will be interpretable with just one PG, licensed by the overtly raised *wh*-phrase[4]:

(25) a.   ?Which senator$_1$ did you persuade _$_1$ to borrow *which car$_2$* [after talking to _$_1$ for an hour]

     b.   ?Which kid$_1$ did you give *which candy bar$_2$* to _$_1$ [in order to impress _$_1$]?

---

[4] Semantic composition would be straightforward. While the lowest vP-segment would be a two-place predicate, the outermost argument position (the lambda-abstract over the gap left by the second movement) would be bound immediately by the intermediate trace of *wh2*, leaving a one-place predicate.

On the other hand, there should be no way to license a single PG associated with the *wh*-in-situ. This is predicted by the TIC; the necessary structure would require the covert movement to raise to the outermost spec-vP position, extending the vP. This prediction, too, seems to be borne out:

(26) a. * Which of you _ persuaded that senator to borrow *which car$_2$* [after putting a bomb in$_2$]

     b. * Which of you gave that kid *which candy bar$_2$* [without first telling him about the ingredients in _$_2$]?

Interestingly, it seems that the order of the PGs has to mimic the surface order of the two *wh*-phrases, suggested by the contrast between (21) and (27).This is exactly what is predicted given that the empty operator movements in the adjunct will be subject to the same constraints (Superiority and TIC) as the movements in the matrix.

(27) a. * Which car$_1$ did you lend_$_1$ to *which senator$_2$* [after getting an opponent of _$_2$ to put a bomb in _$_1$]

     b. * Which kid$_1$ did you give *which candy bar$_2$* to _$_1$ [without mentioning the ingredients in _$_2$ to a parent of _$_1$]

Notice that (21)b,(27)b involve a superiority violation in the matrix, licensed by d-linking (cf. Pesetsky 1987, in press). The fact that (21)b is comparatively acceptable, in contrast to examples like those in (27) in which superiority would have to be violated inside the adjunct, apparently shows that superiority is strictly enforced in null operator movement. This is unsurprising, given the generalizations that are known to govern exceptions to superiority. Among other conditions that have to be met are a distinctive prosodic pattern with a prominent pitch accent on the 'skipped over' *wh*-phrase. This prosodic pattern is unavailable with a null operator, for obvious reasons.

A further prediction is that the 'extra' PG must be in the same island as the 'first' one. This follows from principles of semantic composition. The vP is a two-place predicate; if each of the adjuncts is only a one-place predicate, as in (28), predicate modification will fail and the structure will be uninterpretable.

(28) a. * Which senator₁ did you persuade _₁ to buy *which car₂* [after talking to _₁][without fixing _₁ first]

   b. * Which kid₁ did you give *which candy bar₂* to _₁ [without talking to _₁][in order to get rid of _₂]?

Moreover, if the second of two vP-adjuncts contains just one PG, it must be licensed by the *overt* movement. It cannot be associated with the *wh*-in-situ. This is parallel to the one-gap asymmetry that we saw in Chapter 2 (section 3.2.1):

(29) ? Which paper₁ will you file _₁ in *which cabinet₂* [without telling the author of _₁ how to open _₂]
   [unless I write out a label for _₁]
   *[unless I fix the drawers of _₂]

Finally, sentences like (30) provide evidence for covert pied-piping. The second PG in this example is associated with the entire bracketed DP containing the *wh*-in-situ, not the *wh*-phrase itself. PGs licensed by overt *wh*-movement are similarly licensed by a pied-piped category (31).

(30) ? Which senator₁ did you persuade _₁ to borrow [*which person's₂ car*]₃
   a.   ...[after getting an opponent of _₁ to put a bomb in _₃]
   b.   *?...[after getting an opponent of _₁ to shake hands with _₂]

(31) a. [Which person's₁ car]₂ did you borrow _₂ [without cleaning _₂ afterward]

   b.   *...[without thanking _₁ afterward]

I conclude this section with a few more examples of 'extra' PGs licensed by *wh*-in-situ:

(32) a. ?Who did you talk to_about reviewing *which article* [after showing a colleague of_ several examples in _]

  Who _ talked to you about reviewing which article [after discussing several examples in _]

  Who did you talk to _ about reviewing which article [after showing a colleague several examples in_]

  b. ?Who did you invite _ to *which department* [in order to introduce _ to people who work for_]

  Who _ invited you to which department [in order to introduce you to people who work for _]

  Who did you invite _ to which department [in order to introduce him to people who work for _]

  c. ?Which book₁ did you give _ to *which student* [without first showing _ to friends of_]

  Who _ gave a book to which student [without first showing it to friends of _]

  Which book did you give_to which student [without first showing it to friends of _]

  d. ?Which actor did they assign *which role* to _ [without even asking _ if he wanted _]

  Which actor _ was assigned which role [without even saying he wanted _]

  Which actor did they assign which role to _ [without even knowing if he wanted _]

  e. ?Who did you put _ in *which office* [before talking to _ about the furniture in _]

  Who _ put people in which office [before talking to anyone about the furniture in _]

  Who did you put _ in which office [before talking to anyone about the furniture in _]

3. Multiple overt extractions in English and multiple PGs

It is possible to give independent evidence that the Obligatory PG generalization will force two PGs to appear in an adjunct that modifies a vP which has been targeted by two movements. The independent evidence comes from cases that differ from those of the previous section, in that the first and second extractions are both overt.

3.1 HNPS coupled with *wh*-movement licenses two PGs

In (33) a *wh*-phrase has been extracted from the vP, and another DP has undergone HNPS past the adjunct. The two overt movements license two PGs in the adjunct.

(33)   ?Which book did Smith find _ on top of _, after putting a copy of _ next to _, the table in the corner?

(34)
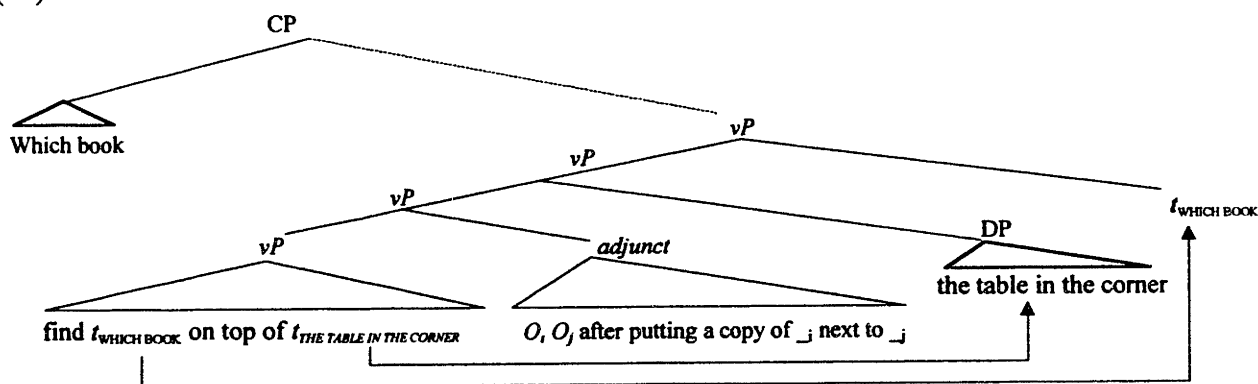
It is possible to demonstrate that this example does not involve Right-node-raising (RNR). Whatever the proper analysis of RNR is, it is known to be immune from islands; RNR also has the property that the two domains from which 'extraction' has take place (i.e. which share the apparently shifted NP) need to have parallel structure and contrastive prosody. Thus compare:

(35) a. **RNR:**

What book did Smith find a chapter of next to, after putting a review of next to, the lamp beside his desk?

b. **RNR is immune to islands:**

Which book did Smith find a CHAPTER of next to, after talking to [the person who put a REVIEW of next to], the table in the corner of the room?

(36) **Example (33), like HNPS and unlike RNR, is constrained by islands:**

* Which book did Smith find on top of, after giving a review of to [the person who was standing next to], the table in the corner of the room?

I argued in the previous section that structures differing from (34) only in that the second extraction is covert will have two obligatory PGs (by the OPG (20)). However, where the second movement is covert, the only way we have of detecting whether the landing site is above or below the adjunct is the presence or absence of the second PG. HNPS provides an independent way of controlling for this; in (33) the word order gives more direct evidence that the landing site of the second movement is past the adjunct. Therefore cases like (33) are predicted to differ from multiple *wh*-movement in requiring two PGs. The unacceptability of (37)a,b bears out the prediction.

(37) a. * Which book did Smith find on top of, after reading a review of, the table in the corner?

b. * Which book did Smith find on top of, after wiping with a sponge, the table in the corner
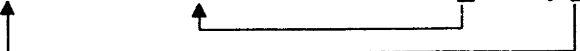
3.2 Pesetsky's *Volvo*-sentences

Pesetsky (1982) made several interesting observations concerning multiple overt extractions in English. The cases Pesetsky discussed involve *wh*-island violations (unlike the cases above where one of the extractions is HNPS), and have at best marginal acceptability. One of
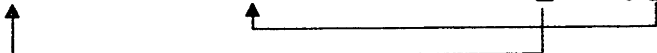
Pesetsky's discoveries about such cases is that the dependencies can only be interpreted as 'nested' rather than 'crossing'. Examples (38) and (39) illustrate this point:

Pesetsky's generalization: the *Path Containment Condition* (PCC)

(38) a. ??This Volvo is one car that I know who we can convince _ to buy _

b. # Trent Lott is one senator that I know which car we can convince _ to buy _

(39) a. ??Which car did I say that it was TRENT LOTT who we should convince _ to buy _

b. # Which senator did I say that it was GORE'S VOLVO that we should convince _ to buy _

Pesetsky's generalization gives us the effect that in (38) the *wh*-phrase with embedded interrogative scope, and in (39) the operator of the embedded cleft, must bind the higher of the two gaps. In both cases it is the lower of the gaps that is bound from outside the *wh*-island. In chapter 5 I will argue that Pesetsky's generalization is a predicted consequence of principles of attraction and the cycle. For now, all that is important is the observation that it is a very strong effect. While the (a) examples of (38)-(39) have the marginal status of island violations, the (b) examples are uninterpretable (except on the nonsensical PCC-obeying readings under which cars are understood as being convinced to buy people).

Of interest in the present discussion is another fact that Pesetsky discovered about these environments. Multiple overt extractions of this sort license multiple-PGs (as do the cases above)—and the order of the PGs must replicate the order of the licensing gaps. This is exactly

what we observed above with ordinary multiple-*wh*-movement (27) where the second movement

was covert.[5, 6] The second of the two *wh*-movements to the vP [*convince _₁ to buy _₂*] must be

associated with the *second* PG:

(40)　??Which car did you say that it was TRENT LOTT who we should convince _ to buy _
　　　a.　…[after getting an opponent of _ to put a bomb in _]
　　　b.　*…[after getting a former owner of _ to vote for _]

(41)　??This Volvo is one car that I know which senator we can convince _ to buy _
　　　a.　…[after getting an opponent of _ to put a bomb in _]
　　　b.　*…[after getting a former owner of _ to vote for _]
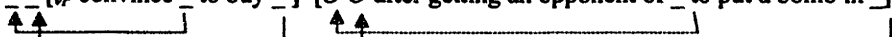
We now have an additional prediction that can be tested. If we change (40)-(41) slightly so that

the adjunct contains only one PG, an asymmetry (of the type that we saw in Chapter 2, section

3.2.1) ought to emerge. The second of the two *wh*-movements to the vP [convince…] cannot

license a single PG.

　　This prediction is tested in (42)a-b. Both are double extractions with a single-PG adjunct.

In (42)a the PG is licensed by the highest gap; in (42)b the PG ought to be licensed (for

pragmatic reasons) by the lower gap. But (42)b cannot be interpreted this way—to the extent that

---

[5] I have modified Pesetsky's examples by putting the PGs in an adjunct rather than a subject; his observation carries over to these versions. I thank N. Richards for reminding me of Pesetsky's *Volvo*-sentences and of pointing out to me their relevance for the discussion in this chapter.

[6] Richards (1997) argued that sentences like (40)-(41) provide evidence against the null operator theory of PGs. The basis for Richards' argument is an assumption that the (b) examples are ruled out because of PCC violations. Richards observed that the null operator hypothesis would require an extra stipulation to ensure that the (b) examples violate the PCC internal to the island. The assumption that the PGs are bound by the overt antecedents requires no such stipulation; both pairs of gaps (main clause and adjunct) enter into PCC-obeying dependencies only in the (a) sentences.

　　However, we might now be in a better position to understand what is going on in these cases: given TIC (even under Richards' formulation), all instances of multiple movement to the same projection (i.e. the null operator movements internal to the island, as well as the successive-cyclic movement to the matrix vP) are *expected* to create nested dependencies, as in (i). Pesetsky's contrast then is due to a fact about predicate modification: the two-place predicates pass up their open argument positions in the same order.

(i)　　　…[vP _ _ [vP convince _ to buy _] [O O after getting an opponent of _ to put a bomb in _]]

the sentence is interpretable (the double-extraction is somewhat marked to begin with) the adjunct seems to imply that a bomb should be 'put in' Senator Lott. (Alternatively, a low attachment for the adjunct is interpretable with the reading in which the PRO subject is controlled by Trent Lott (i.e. ...[*to borrow _ after putting a bomb in _*])).

(42) ?Which car did you say that it was TRENT LOTT who we should convince _ to borrow _
     a.   ... [after getting an opponent of _ to put a bomb in _]   *2 PGs*
     b.   ... [after talking to _]   *1 PG (first movement)*
     c.   ...# [after putting a bomb in _]   *1 PG (second movement)*

(43) ?This Volvo is one car that I know which senator we can convince _ to buy _
     a.   ... [after getting an opponent of _ to put a bomb in _]   *2 PGs*
     b.   ... [after talking to _]   *1 PG (first movement)*
     c.   ...# [after putting a bomb in _]   *1 PG (second movement)*

Structures underlying the vPs in (40)-(41) and those in (42)-(43) are shown below, in (44) and (45) respectively.

(44)
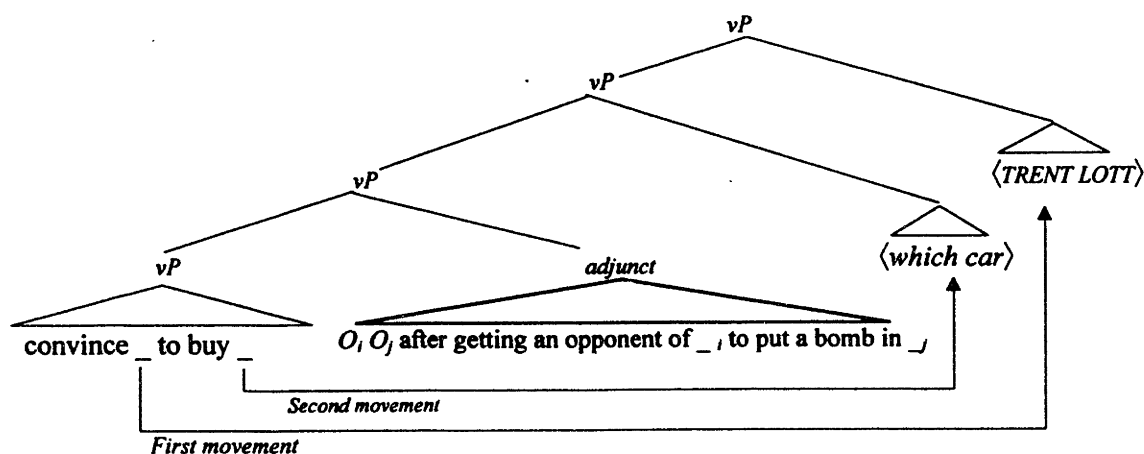
(45)
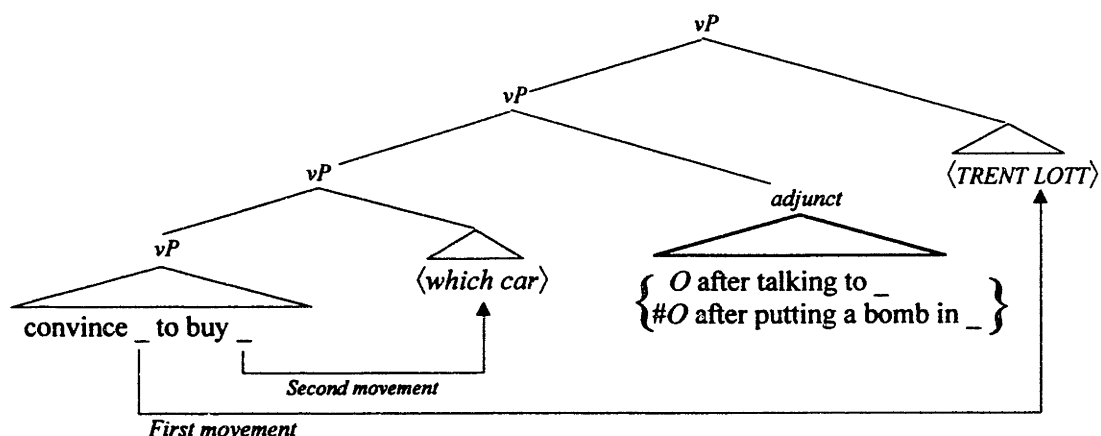
4.   QR and parasitic gaps

So far the discussion of PGs and covert movement has been limited to multiple-*wh*-movement

(and more generally multiple successive-cyclic A-bar movement). Consider similar cases in

which an overt instance of *wh*-movement is followed by QR. Sentences like (46)a appear to be

marginally acceptable, contrasting in an expected way with examples of comparable complexity

in which there is no PG associated with the overt movement ((46)b).

(46) a. ??Which car did you persuade each republican senator to borrow _ [after getting an
        opponent of _ to put a bomb in _]?

   b. * Which car did you persuade each republican senator to borrow _ [after getting an
        opponent of _ to put a bomb in it]?

Apparently a PG can be licensed by QR as well as by covert *wh*-movement. Importantly, the

order of the PGs in (46)a is *reversed* from the order of licensing movements. This is the opposite

of what we found in multiple-*wh*-questions. When it is QR that licenses a second PG the

examples seem to degrade if the order of PGs is the same as the order of licensing movements.

Compare (47) with (46)a:

(47)   * Which senator₁ did you persuade _₁ ʰʰ borrow each rental car₂ [after getting an opponent of _₁ to put a bomb in _ⱼ

Notice further that the only natural interpretation of (46)a seems to be the reading in which the cars vary with the Republican senators. In other words, the universal quantifier needs to take scope over the *wh*-phrase when there are two PGs in the adjunct. This is not a consequence of the fact that the quantifier *each* is used in the sentence (although it is true that *each* has a strong tendency to take wide scope). Sentences like (48) seem perfectly natural with either wide or narrow scope for the quantifier.[7]

(48)   Which car did you persuade each Republican senator to borrow _ last week?

The addition of PGs to an otherwise ambiguous sentence like (48) (as in (46)a) thus seems to disambiguate in favor of inverse scope. This fact looks like it is somehow related to the ordering of PGs in the adjunct. It is difficult to test whether the relation is merely coincidental, because of the apparently fixed ordering of gaps in the adjunct. Nonetheless, the order of the PGs tells us something about the kind of derived predicate that the matrix vP has to be; the open argument positions have to come in the right sequence in order for predicate modification to yield an appropriate interpretation. As we saw from the cases involving multiple-*wh*-movement, the operator O_1 that binds the first PG in the adjunct must be associated with the structurally higher of the two licensing phrases in spec-vP. Hence the vP in a sentence like (46)a should have the structure indicated in (49), with the universal quantifier outscoping the vP-trace of *wh*-movement.

---

[7] Narrow scope is facilitated if the universal is taken to distribute over a covert temporal operator; the reading can be paraphrased as "Which car did you persuade each senator to borrow *on a different day* last week"

120

(49)

persuade [each senator] to borrow $t_{WHICH\ CAR}$ $O_i\ O_j$ after getting an opponent of $\_j$ to put a bomb in $\_j$

⟨each sen.⟩ $t_{WHICH\ CAR}$
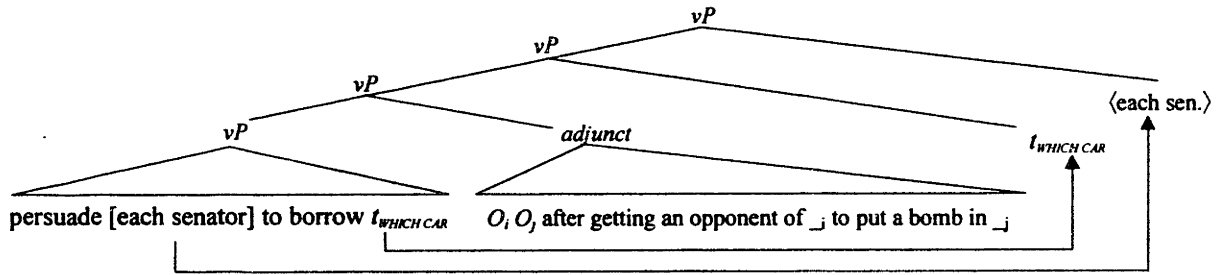
4.1 Tucking In and the scope economy condition

This consequence is interesting in that it appears to challenge a central tenet of the discussion up to now: namely the TIC (10). Covert raising of the quantifier, which is sequenced after the overt *wh*-movement (by the Y-model), has to extend the tree in order to form (49). However it would clearly be wrong to say that QR differs from *wh*-movement in that it is not subject to the TIC. If that were the case then we would expect QR to freely license PGs, which it does not (cf. (46)b). To the extent that QR can license an interpretable PG in cases like (46)a, it seems to be governed by the same generalization that governs other instances of covert PG-licensing: there needs to be an additional PG licensed by an overt movement in the same domain.

I will not make a concrete proposal concerning this apparent contradiction. However I will offer a speculation. The speculation is based on a hunch that there is a relationship between the obviation of *TIC* in (46)a/(49) and the disambiguation toward wide scope; the pair-list reading is (somehow) related to the universal taking scope over the intermediate position of the *wh*-phrase. This might make sense within a theory of interrogatives in which the *wh*-phrase is interpreted as a kind of existential quantifier that is reconstructed into the clause over which an interrogative operator takes scope (cf. Baker 1970, Rullmann and Beck 1998).

The tentative speculation is that the obviation of *tucking in* in (49) is licensed by the *scope economy condition* proposed by Fox (2000a). *Scope economy* could conceivably give us a way of thinking about the Tucking-in condition itself: movements are as local as possible,[8] but locality can be obviated when a longer movement would yield a distinct interpretation. A second *wh*-movement in a given domain can't extend past the first, because extension would yield an interpretation that is truth-conditionally non-distinct from the more local tucking in. Raising past a predicate-adjunct in order to license a PG evidently would not constitute sufficient grounds for licensing the longer movement—just as QR cannot violate scope economy solely to license a bound variable (Fox 2000a). On the other hand, raising a second *wh*-phrase to the position above an adjunct *would* be licensed if the structure that results would put two *wh*-phrases in a structurally adjacent relation (argued by Higginbotham and May 1980 to be a necessary configuration for the semantic rule of 'absorbtion' which treats a pair of *wh*-phrases as a single operator that binds two variables). Consequently a *wh*-phrase could target the position above the adjunct only if it simultaneously tucks in below another *wh*-phrase. Finally, a universal quantifier (as in the above examples) can extend past a *wh*-phrase, on the assumption that the resulting scope-inversion is involved in the 'pair-list' reading.

If something along these lines is correct, it would answer two questions that have been lurking in the background. The first is why tucking in doesn't apply to the first *wh*-movement to spec-vP, which evidently raises past the subject. The reduction of tucking in to *scope economy* (if it could be achieved) would answer this question (at least for quantificational elements): a less

---

[8] The notion that Tucking In reduces to a locality requirement was part of Richards' (1997) original conception of the condition. I refrained in sections 1 and 2 from stating TIC as a locality condition largely to allow the alternative account for Bošković's puzzle that I proposed. If the speculation above is on the right track (i.e. that TIC reduces to Fox's (1995,2000a) scope economy condition), the Bulgarian triple-*wh*-data do not obviously follow. One would want to investigate whether there are subtle differences in interpretation of the two orders.

local movement (i.e. past the internal subject) is allowed so that a propositional constituent can be targeted. The second question that would be answered by the *scope economy* view of tucking in has to do with examples like *Who filed which book [without reading _][in order to clear his desk]*. The PG in the first adjunct is uninterpretable, in spite of the fact that the *wh*-in-situ could raise past that adjunct without violating TIC (due to the presence of the second adjunct). The ungrammaticality of this derivation would be explained if *scope economy* is what allows a *wh*-phrase to raise past an adjunct. Raising past the adjunct to tuck in below another *wh*-phrase is allowable (so that 'absorbtion' is possible). But raising past one adjunct to tuck in below another adjunct would not yield an interpretation that is distinct in the relevant sense.

5.   Conclusions

This chapter presented new facts that topple a long-standing assumption about parasitic gap constructions: it was shown that in a well-defined class of cases PGs are licensed by *wh*-in-situ. These facts constitute fairly strong evidence for the existence of covert phrase movement; they also imply that previous accounts for the supposed inability of covert movement to license PGs cannot be correct. Moreover, the supposed difference between overt and covert movement with respect to PG licensing has been problematic for attempts to eliminate the need for s-structure. By showing that the apparent difference is an illusion, this obstacle has been removed, contributing to the now growing body of evidence for an 'interface model'.

   We also saw evidence that covert movement has three important properties in common with overt movement (in addition to the ability to license PGs). Both overt and covert movement are subject to a condition that requires tucking in whenever possible; both target vP successive-cyclically; both induce pied-piping of larger categories.

Finally I have argued that the explanation for the failure of standard cases of covert movement to license PGs provides evidence for the 'Y-model'. From the sequencing of overt before covert operations that this model imposes, together with the independently motivated 'tucking in' property of movement, it follows that the configuration in which PGs are licensed can only be created in the overt syntax. That configuration must be derived first by movement of the licensing DP to spec-vP, then by merger of the adjunct immediately below. Since merger of an adjunct is an overt operation (it is pronounced), the immediately prior movement to spec-vP must also be overt.

The exceptional environments in which covert movement is able to license a second PG supports this explanation. A predicted derivation involving multiple movements to spec-vP allows a second (covert) movement to raise past the adjunct without violating the 'tucking in' generalization—the covert movement can merge below the outer specifier but above the adjunct. We saw that in exactly in those environments an adjunct was able to contain two PGs, mirroring cases of multiple overt extraction.

The appendix to this chapter will provide further evidence both for covert movement of phrases and for an 'interface theory' that does not invoke a level of s-structure. In a different way, chapter 4 will likewise provide a evidence for covert phrase movement as well as for an interface theory. However, Chapter 4 will also yield a conclusion that seems to be at odds with a central result of this investigation: namely that the Y-model needs to be replaced with an alternative that allows covert operations to follow overt ones. I will explore the conflicting results in Chapter 5 and suggest an alternative model that is consistent with the evidence for and against the Y-model.

Appendix.   Covert movement and Condition A

The contrast in (1) illustrates the well-known fact that movement feeds Condition A of the binding theory. (1)b is perfectly natural despite the fact that the matrix subject cannot bind a reflexive in the position of the gap (as shown by (1)a).

(1) a. * Mary knows John is looking at a picture of herself

 b. Mary knows which picture of herself John is looking at _

It is generally assumed, however, that covert phrasal movement (if it exists) cannot feed Condition A.

(2)   * Mary knows which man was looking at which picture of herself

(3) a. The students asked what stories about each other the teacher had heard

 b. * The students asked which teacher had heard what stories about each other

(4) a. **s-structure of (2):**

 Mary knows [which man]$_i$ _$_i$ was looking at [which picture of herself]$_j$

 b. **A possible LF of (2) (formed by raising of _wh_-in situ):**

 Mary knows [which picture of herself]$_j$ [which man]$_{i\_j}$ was looking at _$_j$

(5)   **An alternative LF of (2)**

 Mary knows [which man]$_i$ [which picture of herself]$_{j\_i}$ was looking at _$_j$

The postulation of the alternative LF raises several important questions: (i) Would that structure (5) fail to provide the locality required for binding by the matrix subject? (ii) Is (5) the LF of

(2)? And (iii) Is Condition A a property of LF (or, similarly, of the cognitive system that interprets LFs)?

I will argue that (i), (ii) and (iii) should all be answered in the affirmative. The 'Tucking in' property that is held by movement has created the illusion that ostensible cases of covert movement don't feed Condition A.

Given the Tucking-In condition (TIC) argued for in this chapter, covert movement of *wh-in-situ* tucks in below the *wh*-phrase in spec-CP:

s-structure:

SUBJECT$_i$ knows [$_{CP}$ WH-1$_j$ _$_j$ ... [WH-2 ... **anaphor**$_{*i/j}$...]$_k$ ...]

LF (after covert raising of WH-2):

SUBJECT$_i$ knows [$_{CP}$ WH-1$_j$ [WH-2 ... **anaphor**$_{*i/j}$...]$_k$ _$_j$ ... _$_k$ ...]

If this is the LF of (2), then (2) does not provide an argument against either covert *wh*-movement or the view that Condition A applies at LF. Even on these two assumptions, the reflexive would not necessarily be bound by the matrix subject. In the next section, we will see direct evidence that it would not be bound in this configuration. Assuming this to be the case, (2) is simply not the right test.

A better test might have the following structure:

s-structure:

WH-1$_i$ _$_i$ said [$_{CP}$ that SUBJECT$_j$ ... [WH-2 ... anaphor$_i$...]$_k$...]

LF:

WH-1$_i$ [WH-2 ... anaphor$_i$...]$_k$ _$_i$ said [$_{CP}$ that SUBJECT$_j$ ... _$_k$ ...]

(6) a.   Who thinks Mary was looking at which picture of himself?

   b. * Who thinks Mary was looking at a picture of himself?

(7)   * Who thinks Mary took which picture of himself?

(8) a.   How many (of those) pictures did Mary already agree to buy?

    Has a natural reading which asks for the number of particular pictures that Mary has agreed to buy

   # How many (of those) pictures did Mary already agree to paint?

   b.   How many pictures did Mary agree in principle to buy?

    There need not be any particular pictures that she has agreed to buy

   How many pictures did Mary agree in principle to paint?

(9) a.   Who thinks Mary (already) agreed to buy how many pictures of herself?

   b. i) Who thinks Mary agreed (in principle) to buy how many pictures of herself?

    ii) Who thinks Mary agreed to paint how many pictures of herself?

(10) a.   Who thinks Mary (already) decided to buy how many pictures of himself?

   b. * Who thinks Mary decided in principle to buy how many pictures of himself?

    * Who thinks Mary decided to paint how many pictures of himself?

The 'Baker' ambiguity:

(11) a.   Which boy asked where we can buy which pictures of his mother?

   b.   Which boy asked how many people bought which pictures of his mother?

only wide scope:

(12) a.   Which boy asked where we can buy which pictures of himself?

   b.   Which boy asked how many people bought which pictures of himself?
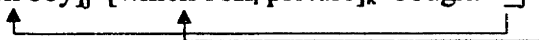
wh-in-situ and ACD

(13) a. Which teacher told John and Bill to read which of the books that each other's advisors had recommended?

b. * Which teacher told John and Bill to read which of the books that each other's advisors asked her to?

c. Which two **teachers** told John to read which of the books that each other's T.A.s asked them to?

Recall that Bulgarian wh-movement obeys a *superiority* condition (Rudin, Richards):

(14) a. Meri znae [koj]ⱼ [kakvo]ₖ kupi   _ⱼ _ₖ

M. knows [who]ⱼ [what]ₖ bought _ⱼ _ₖ

'Mary knows who bought what'

b. * Meri znae [kakvo]ₖ [koj]ⱼ kupi   _ⱼ _ₖ

M. knows [what]ₖ [who]ⱼ bought _ⱼ _ₖ

* 'Mary knows what who bought'

In the Bulgarian equivalent of (2) the movements are overt—so we know that the structure is like (5). Apparently the contrast is fairly noticeable: a reflexive contained in a "second" *wh*-phrase cannot be bound by the matrix subject, while a reflexive contained in a "first" *wh*-phrase can be.

(15) a. * **Meri**ᵢ znae [koe momĉe]ⱼ [koja **svoja**ᵢ snimka]ₖ kupi   _ⱼ _ₖ

Maryᵢ knows [which boy]ⱼ [which **refl**ᵢ picture]ₖ bought _ⱼ _ₖ

'Mary knows which boy bought which picture of himself/*herself'

b. ??Meri$_i$ znae   [koj svoj$_i$ brat]$_j$ [koja snimka]$_k$ kupi   _$_j$ _$_k$

Mary$_i$ knows [which refl$_i$ brother]$_j$ [which picture]$_k$ bought _$_j$ _$_k$

'Mary knows which of her(self's) brothers bought which book'

If the ordering of *wh*-phrases is reversed (which D-linking permits), a similar pattern emerges:

(16) a. ??Meri$_i$ znae [koja svoja$_i$ snimka]$_k$ [koe momĉe]$_j$ kupi   _$_j$ _$_k$
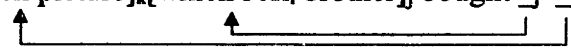
M$_i$ knows [which refl$_i$ picture]$_k$ [which boy]$_j$ bought _$_j$ _$_k$

'Mary knows which picture of herself which boy bought'

b. *Meri$_i$ znae   [koja snimka]$_k$ [koj svoj$_i$ brat]$_j$ kupi   _$_j$ _$_k$

M$_i$ knows [which picture]$_k$[which refl$_i$ brother]$_j$ bought _$_j$ _$_k$

'Mary knows which book which of her(self's) brothers bought'

Further examples from Bulgarian illustrate the same pattern:

(17) a. ?Meri$_i$ znae [koja svoja$_i$ snimka]$_j$ [na koe momĉe]$_k$ se haresva _$_j$ _$_k$

M$_i$ knows [which refl$_i$ picture]$_j$ [to which boy]$_k$ refl like-3sg _$_j$ _$_k$

(refl + like3sg = pleases)

'Mary knows which picture of herself pleases which boy'

b. ??Meri$_i$ znae [na koe momĉe]$_k$ [koja svoja$_i$ snimka]$_j$ se haresva _$_j$ _$_k$

M$_i$ knows [to which boy]$_k$ [which refl$_i$ picture]$_j$ refl-like-3sg _$_j$ _$_k$

'Mary knows which boy which picture of herself pleases'

(18) a. * Meri_i znae [na koe momče]_k [koja svoja_i snimka] dade   __j __k

    M_i knows [to which boy]_k [which refl_i picture] gave-2sg __j __k

    'Mary knows which boy you gave which picture of herself to'

  b. ? Meri_i znae [koja svoja_i snimka]_j [na koe momče]_k dade   __j __k

    M_i knows [which refl_i picture]_j [to which boy]_k gave-2sg __j __k

    'Mary knows which picture of herself you gave to which boy'

  c. ?? Meri_i znae [na koj svoj_i brat]_k [koja snimka]_j dade   __j __k

    M_i knows [to which refl_i brother]_k [which picture]_j gave-2sg __j __k

    'Mary knows which of her(self's) brothers you gave which picture to'

  d. * Meri_i znae [koja snimka]_j [na koj svoj_i brat]_k dade   __j __k

    M_i knows [which picture]_j [to which refl_i brother]_k gave-2sg __j __k

    'Mary knows which picture you gave to which of her(self's) brothers'

QR and Antecedent-contained deletion create another feeding relation

(19)   **S-structure:**
      SUBJECT [_{VP}...IND.OBJ. [_{QP}...**anaphor**..._{VP}⟨ellipsis site⟩]_j]

    LF:
      SUBJECT [_{QP}...**anaphor**..._{VP}⟨ellipsis site⟩]_j [_{VP}...IND.OBJ.__j]

(20) a.   I gave John and Ted *cabernets* that each other's wives couldn't stand

    b.   John and Ted gave me *cabernets* that each other's wives couldn't stand

    c. ??I gave John and Ted the (same) *cabernets* that each other's wives did

    d.   John and Ted gave me the (same) *cabernets* that each other's wives did

(21) a.   I sent Mary the pictures of herself that you liked so much

    b.   Mary sent me the pictures of herself that you liked so much

    c. *?I sent Mary the pictures of herself that you told me to

    d.   Mary sent me the pictures of herself that you told her to

(control to show that no occurrence of a reflexive may be interpreted above its antecedent)

(22) a. * Which picture of herself that she bought from John does he think Mary hates

    b.   Which picture of herself that she bought from John does Mary think he hates

    c.   Which picture of her mother that she bought from John does he think Mary actually hates

(23) a.   How many books that each other's friends gave John for his birthday did his parents suspect he wouldn't read?

    b. * How many books that each other's friends gave John for his birthday did he suspect his parents hadn't even read?

    c.   How many books that each other's friends gave him for his birthday did John suspect his parents hadn't even read?

Chapter 4. An apparent '*anti-Y-model effect*': Covert movement and extraposition from NP[1]

This chapter argues that "covert" operations like Quantifier Raising (QR) can precede "overt" operations. Specifically we argue that there are overt operations that must take the output of QR as their input. If this argument is successful there are two interesting consequences for the theory of grammar. First, there cannot be a "covert" (i.e. post-spellout) component of the grammar. That is, what distinguishes operations that affect phonology from those that do not cannot be an arbitrary point in the derivation ("spellout") before which the former apply and after which the latter do; all syntactic operations apply in the same component (henceforth 'single component grammar'). Second, there must be some alternative means for distinguishing "overt" from "covert" operations. One such alternative, which we can call the 'phonological theory of QR', was suggested by Brody (1995), Bobaljik (1995), Pesetsky (1998), Groat and O'Neil (1994). These authors proposed that the distinguishing property has to do with principles of the syntax-phonology interface. Assume that movement is a copying operation with phonology targeting one copy in a chain for pronunciation. The distinction between "overt" and "covert" movement, these authors suggest, is this: "overt" movements are the result of phonology targeting the head of a chain for pronunciation, while "covert" movements result from phonology targeting the tail of a chain. We will tentatively adopt this phonological theory of QR (but see footnote 5).

Notice that these conclusions, if they are right, conflict with a central result of Chapter 3. A seemingly strong argument for the Y-model emerged from the investigation of parasitic gaps

---

[1] This chapter is the product of collaborative research with Danny Fox. What follows, with only very minor revisions, is our joint paper *Extraposition and scope: a case for overt QR*, published as Fox and Nissenbaum (1999).

in that chapter. In particular, the sequencing of overt before covert movements that is imposed by the Y-model played a crucial role in the explanation for Engdahl's generalization that covert movement (typically) does not license PGs. For now we will put this conflict aside; in Chapter 5 it will be brought to the fore, and will serve as the starting point for a theory of derivation which seeks to reconcile the apparently conflicting evidence for and against the Y-model. The proposal will entail that neither the Y-model nor the 'phonological theory' of covert movement is correct. However, for the sake of clarity of exposition, this chapter will keep to the more familiar description of covert movement in terms of the 'phonological theory'.

The argument that covert operations sometimes precede overt operations is based on extraposition from NP. Specifically, we argue that certain instances of extraposition result not from movement of the extraposed material but rather from QR of an NP and subsequent merger of an adjunct phrase. Phonology will determine that the NP is pronounced in its pre-QR position. But the late-inserted NP-adjunct is not present in the pre-QR position—it can only be pronounced in the position in which it was merged into the structure. QR followed by merger of an adjunct which is "overt" is impossible if covert operations apply after spell-out, hence the consequences for the architecture of the grammar noted above.

We start this paper with a well-known puzzle: extraposition seems to violate a robust generalization about movement, namely that adjuncts cannot be extracted from NP. A possible resolution for this puzzle is provided by the assumption that extraposition is not a unified phenomenon. Adjunct extraposition does not involve movement of the adjunct, hence does not violate the constraint. We argue in the remainder of the paper in favor of a resolution of this sort. Specifically, we argue that adjunct extraposition is derived by post-QR merger of the adjunct. Extraposition of complements, by contrast, is derived in a traditional manner, i.e. by movement

of the complement.[2] The argument is based on two observations. First, the "extraposed constituent" in adjunct extraposition—in contrast to complement extraposition—shows no sign that it has been moved (and every sign that it has not). Second, in adjunct extraposition and not in complement extraposition, the NP with which the extraposed constituent is associated shows every indication that it has undergone QR.

1. Extraposition from NP—a puzzle

Consider the paradigm in . This paradigm illustrates a well-established restriction on movement: a complement can be extracted from NP (1)a, and an adjunct cannot (1)b.

(1) a.   Of whom did you see [a painting t]?

b. *??From where/*??By whom did you see [a painting t]?

Extraposition from NP doesn't seem to obey this restriction, as exemplified in (2) (noted by Culicover and Rochemont 1990, 1992).

(2) a.   We saw [a painting t_i] yesterday [of John]$_i$.

b.   We saw [a painting (t_i)] yesterday [from the museum]$_i$.

c.   We saw [a painting (t_i)] yesterday [by John]$_i$.

2. The proposal—post-QR merger of adjuncts

The fact that an adjunct can be "extraposed" from an NP is puzzling under the assumption that extraposition uniformly involves movement of the "extraposed constituent" (EC). However, as

---

[2] For reasons of space, this paper will not deal with extraposition from subject NPs, which has somewhat different properties from the cases of extraposition that we investigate here. We discuss subject extraposition in Fox

Culicover and Rochemont point out, this fact is not puzzling if the assumption is abandoned. Consequently, Culicover and Rochemont suggest that extraposition never involves movement of the EC. In this paper we argue for an alternative resolution of the puzzle. Specifically, we argue that complement extraposition *is* derived by movement of EC—a movement which is consistent with the extraction fact noted in (1)—and that adjunct extraposition has a totally different derivation for which the constraint is irrelevant.

We propose that adjunct extraposition is derived by two different operations, the first covert and the second overt. First the NP with which the EC is associated (the "source NP") undergoes covert movement (QR) to a position (in this case to the right) in which it can be interpreted, and then the EC is adjoined to the source NP. This is illustrated in (3).[3]

(3) We saw a painting yesterday by John.

a.
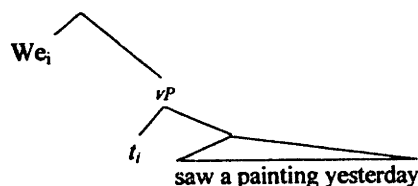
b. *QR* ('covert')

---

and Nissenbaum (in progress).

[3] Something needs to be said about how an LF such as (3)c is interpreted. Assume for the purposes of this chapter, along the lines of Fox (2000a,2000b), that the copy at the trace position is interpreted as a definite description: *the painting (identical to) x*. As a result, (3)c will receive a straightforward compositional interpretation as (i). For an alternative proposal see Sauerland (1998).

(i) [A painting by John] λx we saw [the painting (identical to) x]

c. *adjunct merger* ('overt')
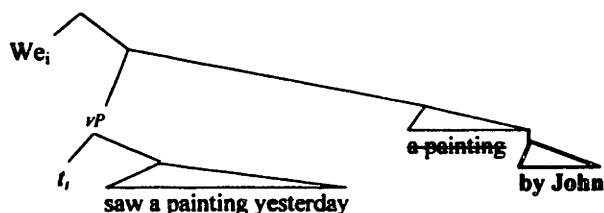
A derivation along these lines was proposed for overt *wh*-movement by Lebeaux (1988). The extension to covert movement is straightforward under the phonological theory of QR. (For related but different proposals see Guéron and May 1984 and Reinhart 1991.)[4]

In the remainder of this paper we will present various arguments in favor of the derivation in (3) for adjunct extraposition. If these arguments are successful, they will strongly support a single component grammar in which covert QR can precede overt merger of an adjunct. Furthermore, the arguments will support the phonological theory of QR, which provides an alternative to a covert component in accounting for the invisibility of QR.[5]

---

[4] Gueron and May's proposal shares with ours the property that the adjunct merges into the structure at the position in which it is pronounced. Similarly, they propose that the source NP raises to the position of the EC, in order (under their view) for the former to govern the latter. Thus the predictions for scope of the source NP, which we spell out and test below, are also implicit in Gueron and May, although G&M do not attempt to confirm them. G&M likewise do not spell out the nature of the movement of the source NP, specifically how it simultaneously satisfies the requirement that a moved constituent c-command its trace, and the standard sisterhood condition for semantic composition of an NP with its complement or adjunct. Finally, G&M's proposal (like that of Culicover and Rochement) does not distinguish between adjunct and complement extraposition, and therefore fails to predict the range of asymmetries that we find and investigate in this paper. Reinhart's proposal for elliptic conjunctions is very similar to G&M's proposal for extraposition. However, she argues that her proposal should not extend to extraposition. We don't have space to discuss her proposal here, but we hope to have something to say about it in Fox and Nissenbaum (in progress).

[5] If our proposal for extraposition is correct, a single-component grammar is virtually forced. The arguments in favor of the phonological theory of QR, however, are strong only inasmuch as this theory provides an alternative to a separate covert component in allowing for the invisibility of QR. Another potential alternative to a model with a separate covert component, which is consistent with our proposal, would abandon the assumption that there is a single point of spellout. Assume that there are many instances of spellout, each one updating a previously computed PF. A theory along these lines will be developed in Chapter 6.

3. Prediction for Scope

Our proposal that adjunct extraposition is derived by the steps shown in (3) makes an immediate prediction: the source NP must have wider scope than its surface position.[6] Specifically, we predict that (4) should hold:

(4)   **Adjunct-extraposition marks scope:**
        When an extraposed constituent (EC) is an adjunct, the scope of the source NP will be at least as high as the attachment site of EC.

To see a case which bears out this prediction, consider (5). These examples exploit a peculiar property of 'free choice' *any*, namely that it must appear in the scope of some modal operator like *look for* or *would*. This property is illustrated in (5)a, where *look for* must outscope *any*; there is no interpretation which requires that there be a particular thing that the speaker was looking for. If (4) is correct then an adjunct extraposed from a source NP headed by 'free choice' *any* will yield an unacceptable result whenever the attachment site of the adjunct is higher than the modal licenser of *any*. Hence the unacceptability of (5)b is predicted. The EC appears to the right of an adverbial that modifies *look for*, signaling that the scope of the source NP must be at least that high—outside the scope of its licenser.

(5) *'Free choice'* any *is licensed in the scope of the verb look for.*

   a.   I was looking very intensely for anything that would help me with my thesis.

   b. * I was looking for anything very intensely that will/would help me with my thesis.

---

[6] This prediction doesn't necessarily follow from QR alone, given the general possibility for scope reconstruction. However, the prediction does follow from the combination of QR and late merger of an adjunct as we have proposed. If an adjunct is present only at the head of a chain, and if scope reconstruction results from interpreting only the tail of a chain (i.e. the head of the chain is deleted at LF), then late merger will block reconstruction; the adjunct would not be interpretable as a modifier of the source NP. Exactly these considerations are needed independently (as pointed out by Fox 1999) to account for correlations between scope reconstruction and binding theory.

c. I was looking for something very intensely that will (likely) help me with my thesis

d. I would buy anything without making a fuss that will/would help me with my thesis.

(5)c and (5)d are control cases. An EC outside the scope of *look for* is in principle allowable as long as the source NP isn't required to have narrow scope (5)c. (In fact, the source NP in (5)c can *only* have wide scope; the sentence would be false if there is no particular thing that the speaker was looking for.) And extraposition is allowed in principle even from a source NP headed by 'free choice' *any*, as long as the EC does not appear outside the scope of the modal licenser of *any*. This is shown by (5)d. There, the EC appears to the right of an adverbial modifier of the main VP; the modal is the auxiliary verb *would*, which is (at least under one available structure) higher than the site of extraposition. Hence QR of the source NP does not bring it out of the scope of its licenser in this case.

These facts suggest that the correlation predicted in (4) is correct: extraposition of an adjunct marks wide scope for the source NP. The correlation would be quite unexpected under the traditional view of extraposition as movement of the EC, but is exactly what is predicted if adjunct extraposition is derived by the steps in (3). This result is replicated in a range of other tests correlating the scope of source NPs with the surface position of ECs. (A fuller paradigm is given in Fox and Nissenbaum, in progress. A few more examples are provided in section 5.7.

4. Complements vs. adjuncts—further predictions

So far we have considered only cases of adjunct extraposition, and provided evidence that in such cases extraposition signals that QR has taken place. Our proposal makes additional predictions, but in order to test these we need to cover some background relating to the interaction of movement and binding theory, and the consequences for late merger. (6)a

illustrates a general property of A-bar movement, namely that it doesn't bleed Condition C of the binding theory. The pronoun *he* in (6)a cannot be co-referent with *John*. From the perspective of Condition C, it looks as if the wh-phrase is in its trace position and *he* c-commands the r-expression *John*. Under the copy theory of movement, this is just what is expected: since movement leaves a copy of the wh-phrase, the pronoun c-commands *John* in the lower copy. However, (6)b is perfectly natural under the co-referent interpretation. Sentences like this, in which the r-expression is in an adjunct rather than a complement, are well-known exceptions to the generalization that A-bar movement doesn't bleed Condition C (see van Riemsdijk and Williams 1981, Freidin 1986).

(6) a.   ??/*[Which book about John's$_i$ library] did he$_i$ read?

     b.   [Which book from John's $_i$ library] did he$_i$ read?

Lebeaux (1988) proposed an explanation for this contrast, which Chomsky (1993) modified to render consistent with his proposal that binding theory applies at LF. While (6)a is ungrammatical due to the presence of *John* in the lower copy of the wh-phrase, (6)b is acceptable because it has an alternative derivation, illustrated in (6)b'. The lower copy of the wh-phrase merges into the structure without the adjunct modifier (6)b'.i). After wh-movement brings the wh-phrase out of the scope of the pronoun (6)b'.ii), the adjunct containing *John* is merged into the structure, adjoining to the higher copy of the wh-phrase (6)b'.iii).

(6)b'   i.   he$_i$ read [Which book]

         ii.   *wh-movement* --> [**Which book**] did he$_i$ read [Which book]

         iii. adjunct merger --> [Which book **from John's$_i$ library**] did he$_i$ read [Which book]

Lebeaux argued that the late-merger component of such a derivation is impossible on principled grounds for NP-complements (the Projection Principle)[7] hence the unacceptability of (6)a.

The derivation in (6)b' is exactly parallel with the derivation that we have proposed for adjunct extraposition (3), and thus supports it. But we are now ready to specify two conclusions that our proposal leads us to. First, given the prohibition against adjunct extraction from NP (section 2) we can conclude that the derivation involving QR followed by late merger is the *only* derivation possible for adjunct extraposition.

Second, Lebeaux's explanation for the contrast in (6) leads us to opposite conclusions for complement extraposition. Specifically, complement extraposition must have a derivation that does not involve QR and late merger (given the Projection Principle). However, we have already seen (section 5.2) that complement extraposition can be derived by (rightward) movement of the EC. In other words, adjunct extraposition can be derived only by QR of the source NP and late merger of the EC, while complement extraposition can be derived only by rightward extraction of the EC from the source NP.

From these two conclusions we derive the following pair of predictions:

(7) **Further** Predictions:

    a.   Indications that the EC has undergone rightward movement from the source NP will be detectable if the EC is a complement but not if it is an adjunct.

    b.   Indications that the source NP has undergone QR will be evident if the EC is an adjunct but not if it is a complement.

---

[7] The projection principle states that the theta criterion must be satisfied at every level of representation. Consequently an argument of a head must be merged with the head at D-Structure; hence there can be no late (post movement) merger of arguments. Alternatively, if we assume (with Chomsky 1993) that a copy of the restrictor in A-bar movement is interpreted in the trace position, then the prohibition against late merger of arguments would be an immediate consequence. If the restrictor contains a noun which needs an argument, it would not be interpretable with the argument absent.

5. Testing whether the Extraposed Constituent moves

In this section we will attempt to show that the EC behaves like a moved constituent in complement extraposition but not in adjunct extraposition as predicted in (7)a. The properties of movement that we will investigate relate to definiteness, Condition C, coordination and parasitic gap licensing.

5.1 Definiteness

Consider the pair in (8). This pair illustrates the well-known fact that extraction of NP is slightly marked when the NP is definite (see Fiengo and Higginbotham 1980).[8]

(8) a. Who did Mary see [a (good) picture of t]?

    b. ??Who did Mary see [the (best) picture of t]?

From this perspective EC in adjunct extraposition behaves like it has not been extracted out of the source NP ((9)a, (10)a). Complement extraposition, by contrast, shows the definiteness restriction that one would expect under the assumption that EC is extracted. (Compare (9)b to (9)c and (10)b to (10)c.)

(9) a. I saw the (best) picture yesterday from the museum.

    b. ??I saw the (best) picture yesterday of the museum.

    c. I saw a (very good) picture yesterday of the museum.

---

[8] The definiteness restriction holds only under the "absolute reading" (Szabolcsi 1986), in which the definite description refers to the best picture in the class of pictures of some individual x (bound by *who*). Szabolcsi argues that under other interpretations there is no real definite description.

(10) a.   I heard the same rumor yesterday that you were spreading.

b.   ??I heard the same rumor yesterday that you were quitting.

c.   I heard a similar rumor yesterday that you were quitting.

5.2 Condition C

As discussed in section 5.5, standard cases of movement are not expected to bleed Condition C (given the copy theory of movement). However, it has been known since Taraldsen (1981) that adjunct extraposition does not meet this expectation:

(11) a.   I gave him$_i$ a picture yesterday from John's$_i$ collection.

(Cf. ??/*_I gave him$_i$ a picture from_ John's$_i$ _collection yesterday._)

b.   I gave him$_i$ an argument yesterday that supports John's$_i$ theory.

(Cf. ??/*I gave him$_i$ an argument that supports John's$_i$ theory yesterday.)

c.   I told you that he$_i$ will accept the argument when you and I last spoke that I presented to John$_i$ yesterday.

(Cf. *_I told you when you and I last spoke that he$_i$ will accept the argument that I presented to John$_i$ yesterday._)

This fact is not puzzling under our hypothesis that adjunct extraposition does not involve movement of EC but rather late merger of the type proposed by Lebeaux (1988). As we saw in section 5, Lebeaux proposed late merger to account for the cases in which overt movement appears to bleed condition C. The same reasoning should hold for covert QR. Furthermore, our proposal makes an additional prediction: complement extraposition, which does involve movement, should be unable to bleed Condition C. This prediction appears to be borne out:

(12) a. ??/*I gave him$_i$ a picture yesterday of John's$_i$ mother.

b. ??/*I gave him$_i$ an argument yesterday that this sentence supports John's$_i$ theory.

c. * I said that he$_i$ would accept the argument when we met that what we presented to John$_i$ yesterday is correct.

5.3 Coordination

The behavior of extraction in coordination is distinctive and can serve as an additional test for movement. Extraction of a constituent is possible out of coordination only if it occurs across the board (ATB). In this section, we will see that displacement is attested ATB in complement extraposition but not in adjunct extraposition. This will provide further evidence that the EC is a moved constituent in complement extraposition but not in adjunct extraposition.

Consider the pairs in (13)-(14). The (a) sentences involve ATB complement extraposition and are acceptable, as predicted.

(13) a.   I wanted to [present an argument__] and [discuss evidence__] very badly that what John told me is right.

b. *I wanted to [present an argument__] and [discuss evidence__] very badly that John told me about.

(14) a.   I wanted to [read a book__] and [understand an article__] very badly about the museum we visited last year.

b. *I wanted to [read a book__] and [understand an article__] very badly from the library we visited last year.

The (b) sentences, by contrast, involve adjunct extraposition. Adjunct extraposition is impossible ATB given our hypothesis that adjunct extraposition involves QR of the source NP rather than rightward movement of the EC; there is no NP that can move ATB and be modified by the late inserted adjunct. (In section 7.2 we will discuss the properties of QR in coordination and see that

adjunct extraposition is possible in exactly those environments that allow the source NP to move by QR.)

5.4 Parasitic Gaps

Finally consider the following pairs:

(15) I presented an argument__ before having evidence__

        a.   that what you told me is right.

        b.   *that you told me about.

(16) I read a book__before reading an article__

        a.   about John.

        b.   *from John's library.

The (a) sentences show that complement extraposition licenses Parasitic Gaps and therefore suggest that complement extraposition is derived by movement of the EC. The (b) sentences show that adjunct extraposition cannot license Parasitic Gaps, thus suggesting that it is derived in some other manner, as we have hypothesized.

6. Testing whether the Source NP undergoes QR

In this section we turn to the second prediction stated in (7). We will attempt to show that the source NP behaves as if it has undergone QR in adjunct extraposition but not in complement extraposition. The properties of QR that we will investigate relate to scope and the behavior of quantifiers in coordination.

6.1 Scope of the source NP

The most obvious reflex of QR (plus late merger of an adjunct, which blocks scope reconstruction; see footnote 6) is the relative scope of the NP undergoing movement with respect to some other operator. The prediction is that adjunct extraposition should—but complement extraposition should not—signal wide scope for the source NP. We have already seen (in section 3) evidence for the first half of the prediction, that an adjunct EC sets a lower bound for the scope of the source NP. Evidence of this sort is repeated below as (17)a.

(17) a. * I looked for any clue very desperately that the detective might have overlooked.

   b.   I looked for any clue very desperately that the detective might have overlooked important evidence.

The deviance of this sentence results from the fact that, on the one hand 'free choice' *any* has to have scope narrower than the modal verb *look for*, but on the other hand extraposition of the adjunct marks scope which is wider. As we saw earlier, this sentence should be acceptable if the source NP weren't required to undergo QR. What we haven't seen yet is (17)b. This example stands in sharp contrast to the unacceptable (17)a. It differs only in that the EC in (17)b is a complement of the source NP rather than an adjunct. The fact that it is perfectly acceptable shows that the EC in complement extraposition—as opposed to adjunct extraposition—does not set a lower bound for the scope of the source NP.

Further examples illustrating this point are shown in (18)-(19). The (a) examples involve adjunct extraposition and the (b) examples involve complement extraposition. Consequently we predict that the source NP will be required to have scope over *look for* in (a) but not in (b). To see that this prediction is borne out, let's focus on the contrast in (18). (18)a is true only in a situation in which there is a particular picture from John's factory that the speaker was looking

for. It cannot be true when the speaker's search would be satisfied by any picture from John's factory; e.g. it would be false if the speaker was merely interested in finding out about the quality of film used and to this end is looking for a sample. (Compare this with *I looked very intensely for a picture from John's factory*, which could be true in this situation.)

(18) a. I looked for a picture very intensely from John's factory.

ヨ> *look for*, * *look for* > ヨ

b. I looked for a picture very intensely of John's factory.

ヨ> *look for*, *look for* > ヨ

(19) a. I looked for a picture very intensely by this artist.

ヨ> *look for*, * *look for* > ヨ

b. I looked for a picture very intensely of this artist.
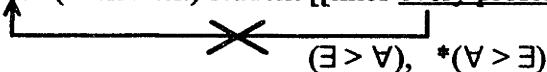
ヨ> *look for*, *look for* > ヨ

The source NP in (18)b, by contrast, *can* have narrow scope with respect to *look for*: it could be true under scenarios parallel to the two described above.

6.2 QR in co-ordination

In section 6.3 we looked at a property of coordination that served as a diagnostic for overt movement of the EC. We saw that complement extraposition shows this property and adjunct extraposition doesn't. We will look in this section at a different property of coordination, one that can serve as diagnostic for covert movement of the source NP. In this case we expect the exact opposite: adjunct extraposition should show this other property while complement extraposition shouldn't.

The property in question was discovered by Ruys (1992). It is well-known that QR in general obeys the coordinate structure constraint (Lakoff 1970, Rodman 1976). This is illustrated by (20), in which the object cannot move by QR over the subject—out of only one of the two conjuncts. Hence, the sentence is limited to the interpretation in which the subject has wide scope.

(20)   A (#different) student [[likes every professor] and [hates the dean]]
        $(\exists > \forall)$, $*(\forall > \exists)$

What Ruys noticed, however, was that there is a specific environment in which QR does not appear to obey the CSC:

(21)   A (different) student [[likes every professor$_i$] and [wants him$_i$ to be on his committee]]
        $(\exists > \forall)$ $(\forall > \exists)$

In (21), unlike (20), *every professor* can have scope over the subject, indicating that QR was able to take place out of the first conjunct alone. Ruys observed that if the second of two conjuncts contains a variable, the QP in the first conjunct is allowed to scope out if (and only if) it is going to bind this variable. The relevant generalization for QR can be stated as (22):

(22) QR of a QP out of a conjunct A (in a structure A & B) is possible iff QP binds a variable in B (Ruys 1992).

We can use (22) as a diagnostic for QR of a source NP. Consider the facts in (23) and (24). The contrast between the (a) and (b) sentences exactly parallels the contrast between (20) and (21) and follows from (22), under the assumption that the source NP undergoes QR in adjunct extraposition. In order for the source NP to undergo QR out of the first of two conjuncts, it must have a variable to bind in the second conjunct.

(23) a. * I wanted to [present an argument__] and [talk about these consequences] very badly that John told me about.

   b. ? I wanted to [present an argument$_i$__] and [talk about its$_i$ consequences] very badly that John told me about.

(24) a. * I wanted to [read a book__] and [meet this author] very badly from the library we visited last year.

   b. ??I wanted to [read a book$_i$__] and [meet its$_i$ author] very badly from the library we visited last year.

These facts are extremely surprising under the view in which the EC undergoes movement in adjunct extraposition. Not only is this movement impossible across-the-board (as we saw in section 6.3), it can occur in violation of the CSC in the (b) sentences in *exactly* the environment in which QR of a different constituent (an NP) is able to circumvent this constraint.

In the (c) sentences, given below, we see that complement extraposition is different in exactly the expected way, on the assumption that complement extraposition involves movement of the EC rather that QR of the source NP. As we saw in section 6.3, movement of the EC—being overt—is possible only when it occurs in the normal across-the-board manner. There is no reason why a variable in the second conjunct (bound by the source NP) would facilitate non-ATB movement of the EC.

(23) c. * I wanted to [present an argument$_i$__] and [talk about its$_i$ consequences] very badly that what John told me is right.

(24) c. * I wanted to [read a book$_i$__] and [meet its$_i$ author] very badly about the museum we visited last year.

7. Conclusions

Throughout this chapter we have seen evidence that complement extraposition shows properties of movement of the EC, whereas adjunct extraposition doesn't. The evidence was

drawn from the restriction on movement from adjuncts as opposed to arguments (section 2), the Definiteness restriction on movement (section 6.1), Condition C (section 6.2), the Coordinate Structure Constraint (section 6.3), and parasitic gap licensing (section 6.4). Consequently we need a different derivation for adjunct extraposition. A phonological theory of QR, together with Lebeaux's late-merger proposal, provides us with this derivation. Adjunct extraposition is the result of post QR merger of an adjunct. This proposed derivation leads us to predict that adjunct extraposition would show properties of QR of the source NP—a prediction which is borne out in the investigation of scope (sections 4, 7.1) and the peculiar behavior of QR in coordination (section 7.2).

The post-QR merger of "extraposed" adjuncts is a case of an overt (i.e. pronounced) operation following a covert (i.e. silent) movement. Such an ordering is impossible under the traditional Y-model of the grammar, for which we apparently saw evidence in Chapter 3. To the extent that our arguments here are successful we need an alternative model, in which syntax intersperses pronounced operations with silent ones. The next chapter will seek to reconcile the evidence for and against such an interspersal.

Chapter 5. Syntax and the *Single-Cycle:*
The nature of covert movement and the architecture of grammar

From the investigations reported in Chapters 2-4 we have reached some tentative conclusions about the kinds of structures that are interpreted at the interface level of LF, as well as some (even more tentative) conclusions about how such structures are derived. This chapter will examine apparent inconsistencies among those results and attempt to reconcile them by means of some proposals about sentence pronunciation embedded in a theory of cyclic computation of structure.

With regard to the conclusions about the structures themselves, the results from all of the studies appear to be consistent and mutually reinforcing, thus providing several converging strands of evidence for them. All three studies support an 'interface theory' (in the sense of Chomsky 1993); two among the remaining phenomena that have appeared to argue strongly for a level of s-structure (PGs and locally bound anaphora) can be shown to result from conditions that hold at LF (or in the interpretive component(s) that access LF). A stronger argument is provided by the study on extraposition from NP, which yields the conclusion that there cannot be a level of s-structure.

To summarize the other major findings:

(i)   There are operations that move phrases covertly.

(ii) Covert movement (just like overt movement) is subject to a condition that requires phrases to tuck in when possible.

(iii) Covert movement (just like overt movement) is successive-cyclic.

(iv) Successive-cyclic movement targets vP as well as CP.

On the other hand, various conclusions that we've reached about properties of *derivation* appear to be more problematic. First, and most troubling, the conclusions that were drawn from the case studies in chapters three and four are mutually incompatible. The study on parasitic gaps provided evidence for the sequencing of overt before covert operations that is predicted by the Y-model. Specifically, the assumption of free ordering between merge and move operations in the *overt block* was shown to allow an explanation for the fact that overt but not covert movement freely licenses parasitic gaps. Although movement cannot target adjunction structures—the mover must *tuck in*—an overt movement can circumvent this requirement, by raising before the adjunct is merged into the structure (below the moved phrase). But covert movement, being in a separate block of operations that cannot be followed by merge, does not have this option. This picture was supported with new facts showing that, in exactly the (limited) environments where covert movement is predicted to raise past an adjunct, it too licenses PGs.

The study on extraposition from NP, however, yielded the opposite conclusion: the Y-model's organization of overt and covert operations into separate blocks was shown to be incompatible with an explanation for the observed correlation between extraposition and the scope of the associate NP. Chapter four (Fox and Nissenbaum 1999) argued that this correlation is explained by the assumption that extraposed adjuncts are merged into the structure (an overt operation) after QR of the NP (a covert one). Since the adjunct is merged to the post-QR position, it can be neither pronounced nor interpreted at the tail of the chain. This ordering of operations, incompatible with the Y-model, was shown to account not only for the obligatory wide scope of the NP but also the fact that adjunct extraposition does not yield *copy-theory* induced violations of Condition C. The conclusion was bolstered by evidence that extraposition

of complements shows neither of these effects, but instead has all the hallmarks of extraction (of the complement) from the NP—including the ability to reconstruct to the position of the copy.

Other troubling questions arise from the conclusions that we've made about the nature of derivation. For instance, the theory of PGs—despite relying on cyclic structure building—appears to be at odds with the most straightforward notion of the cycle. The evidence that was found for intermediate positions in chains was taken to result from the successive-cyclic property of movement. Crucially, we saw evidence that both overt and covert movements target these successive-cyclic positions. As Brody (1995) and Chomsky (1998,1999) have observed, however, under the view that structure is built incrementally by successive applications of merge and move, we don't expect separate cycles for overt and covert blocks. The simpler theory holds that there is one cycle for all movements.

Moreover, the Fox/Nissenbaum theory of extraposition from NP crucially relied on a merge operation that is countercyclic in the extreme: merger to an NP embedded in a DP which is in turn embedded in a vP. This, too, is (obviously) at odds with the notion of cyclic structure formation. We want to see if such an operation has a place in a theory in which structure is built by successive operations (i.e. a cyclic theory): what allows it, how it is constrained. Given the evidence for these countercyclic operations, we hope to discover that they are very tightly constrained by natural and independently motivated principles.

In sum, the following problematic issues have arisen from the studies reported in this dissertation. The aim of the chapter that follows is to resolve them.

(i) The studies on PGs and extraposition from NP yielded conflicting evidence—for and against the Y-model, respectively.

(ii) The account of PGs apparently requires separate cycles for overt and covert operations.

(iii) The account of adjunct extraposition apparently requires a rejection of the cycle altogether.

I hope to show that a theory in which spellout applies on successive cycles will eliminate these problems and pave the way for an understanding of key properties of syntactic derivation. After a review of background assumptions about cyclicity (section 1), I will sketch the outlines of a cyclic theory of spellout (section 2). I will argue for principles that predict the appearance of Y-model effects on each cycle for operations that are 'triggered', i.e. movement and merge of arguments. I will show that adjunct merge, in contrast, is not governed by the cycle and should therefore give rise to 'anti-Y-model effects.' In section 3 I will argue that the cyclic theory allows a principled answer to the questions raised by anti-Y-model effects and countercyclic operations. Section 4 will show that the theory has some surprising empirical consequences concerning restrictions on extraposition. Finally I will suggest some further consequences in section 5, including an account of *wh*-island effects in English and an explanation for the absence of *wh*-island effects in certain multiple-*wh*-fronting languages.

1.   Preliminaries: the cycle

A natural basis for the notion of the syntactic cycle is the incremental formation of structure by successive applications of *merge*. Under this view of grammar, cyclic rule application is a reflex of the fact that phrase structures are built by successively combining smaller units to form larger ones, and that other operations apply in tandem with these combinations. A given instance of *merge* may in turn trigger other grammatical operations (*move, agree,* etc.); a derivation then consists of a sequence of repeated steps: accessing the objects to be combined, applying *merge* and other syntactic operations, computation of stress and other aspects of pronunciation.

An assumption that is almost universally made about the operation of the cycle[1] which has been reinforced by empirical study over many years is that it operates from the bottom up. This necessitates, in a theory of the cycle based on incremental structure formation, a stipulation that structure grows 'outward' rather than 'inward'. Chomsky 1993 stated a strong version of this assumption, requiring that structure-building operations apply at the root and thus satisfy an *Extension Condition*.

However, we have seen empirical results from chapter 3 that suggest the *Extension Condition* is too strong. While those results are consistent with, and provide support for, a cyclic model in which structure grows outward, they argue that some operations apply *near*—but not *at*—the root.[2] Specifically, some movements (e.g. secondary *wh*-movements) are forced to 'tuck in', targeting a position just below the root. This non-extending movement explains the quite restricted distribution of parasitic gaps licensed by covert *wh*-movement, which typically cannot target a vP-adjunction structure, and would also account for the failure of covert movement to feed Condition A in cases where tucking in would not create a local enough relation with the potential antecedent. Similarly, some vP-adjuncts (those that contain PGs) need to merge below the root, to form the appropriate predicate-modification structure after a movement has targeted the vP. The assumption of this non-extending merge operation (together with the assumption of free ordering of overt operations) was argued in chapter three to explain the free distribution of PGs licensed by overt movement.

An alternative assumption which would partially capture the 'bottom-up' property of cyclic structure-formation, while also allowing these non-tree-extending operations, is that

---

[1] But cf. Phillips 1996, Richards 1999 for an interesting exploration of a model that rejects this assumption.

[2] I'll postpone for the moment consideration of the countercyclic merge operation—well below the root structure—that we saw evidence for in chapter 4 (extraposition).

154

operations are *driven by some property of the head* of the root structure. This assumption would

enforce the result that structure is added close enough to the root to satisfy the relevant property

of the head, allowing sufficient flexibility for 'tucking in'. And it would pre-empt operations that

add structure to a *non-root* projection in order to satisfy properties of *its* head. A proposal along

these lines was made in Chomsky 1995. If operations are driven by properties of lexical items

(*merge* by selectional requirements and *move* by properties of attracting heads), we can state the

assumption in terms of the need for a triggering head to satisfy its selectional/attracting features

before computation proceeds any further. A version of this stipulation is stated as (1):

(1) **Operations are driven by the head of the root projection**
Properties of the selector/attractor must be satisfied before a new selector/attractor is
accessed to drive further operations.[3]

Two important consequence of (1) are worth noting immediately. First, (1) says nothing about

insertion of adjuncts. The fact that it refers specifically to operations that are driven by properties

of a head (attraction of a mover and selection of an argument) leaves open the possibility that an

adjunct—which is neither selected nor attracted—may be inserted to an embedded projection.

This is potentially a desirable result given the evidence from extraposition that such operations

are sanctioned by the grammar. However, we will want to investigate this consequence carefully

to make sure that it does not conflict with other important results (such as Y-model effects in the

licensing of PGs). We will return to this issue below.

The second point worth noting is that, while (1) forces all obligatory (feature-driven)

operations to apply on a given cycle before the next cycle starts, it does not impose any ordering

---

[3] This statement is similar to the version proposed in Chomsky 1998:132 [53]: "Properties of the probe/selector α must be satisfied before new elements of the lexical subarray are accessed to drive further operations." As N.

between feature-driven and *non*-feature-driven operations on a cycle. Specifically, argument selection and attracted movements may be freely interspersed with (optional) insertion of adjuncts within the cycle. (1) just requires that the operations triggered by a head be completed by the time a higher head is merged to the structure. This, too, is a desirable consequence given the evidence from PG constructions that vP-adjuncts are present before covert movement targets the vP (a Y-model effect).

One final assumption is (possibly) needed in conjunction with (1) to enforce the 'bottom-up' cycle. As Chomsky (1998) points out, a condition like (1) by itself does not block 'top-down' structure building *within* a cycle. For example, it does not rule out a derivation in which a head that selects two arguments merges with its external argument *below* its complement, or a derivation in which extraction from the complement tucks in below the complement. These derivations are illustrated in (2), where *H* is the selecting/attracting head, and *spec* is the external argument or raised element.

(2)   [H complement]   ⇒   [[[spec] H] complement]

Thus (1) by itself may not sufficiently enforce 'outward' formation of structure.[4] Chomsky 1998 proposed an additional condition that rules out merge below a complement by stipulating that the relation between the projecting head and its *sister* cannot be altered by any structure-building operation. This condition, a version of which is stated in (3), allows tampering at the structural

---

Chomsky (p.c.) points out, a modification is needed if uninterpretable features of a head H are involved in head movement of H on the next cycle. See below.

[4] It is possible that these derivations would be ruled out by interpretive principles, without stipulation. In the case of merge of external argument at the bottom, for example, the interpretive component would simply treat the arguments as reversed, assigning a (possibly) deviant interpretation (or no interpretation at all). It is less clear that (2) would be ruled out where *spec* is extracted from the complement. A possibility is that the trace would not be

'edge' of a projection—i.e. tucking in below specifiers and adjuncts, or inserting adjuncts below

specifiers—but would not allow anything to be merged between the head and its complement.

(3)   **Sisterhood is Unbreakable**
        All structure-building operations preserve sisterhood and c-command relations
        involving the head of the category that projects.[5]

If this additional condition on rule application holds, then structure is built 'bottom-up' not only

from one cycle to the next (enforced by (1)) but within each cycle as well.

Again, notice what operations are allowed and disallowed if (1) is supplemented with (3).

(3) prevents any phrase—including an adjunct XP—from merging to a head that already has a

sister. But (3) does *not* disallow adjunction to an embedded category *if that category* projects, as

in (4).

(4)   [Det [NP]]   ⇒   [Det [$_{NP'}$[NP][**ADJUNCT**]]]

While the merge operation in (4) breaks up the sisterhood relation that initially held between *Det*

and its complement *NP,* that sisterhood relation is not the relevant one under (3)—it is the

category *NP* that projects in the operation (to form the new node *NP'*), not the category *DP*. And

the c-command relation between *Det* and the head *N* of the projecting category remains

unchanged by the merge operation, as does the sisterhood relation holding between *N* and its

complement.[6]

---

bound, in which case the chain would fail to receive an interpretation. However this depends on assumptions about
how the interpretive rule for chains (i.e. predicate abstraction) is defined.

[5] Cf. Chomsky 1998:136-7 [59], which states that "[for all syntactic objects α, γ and R=sisterhood/c-command,]
given a choice of operations applying to α and projecting its label L, select one that preserves R(L, γ)"

[6] As formulated, (3) would disallow the step in (4) if N has no complement, at least on the 'bare phrase-
structure' (Chomsky 1995) assumption that N with no complement does not project to N' or NP. One might assume
that N always has an implicit argument, hence N is a syntactically complex object (NP) even when the complement

To summarize the assumptions reviewed so far, we have adopted a model of structure building which states a 'bottom-up' cycle in the form of (1), with the chief consequence that all operations that will be triggered by a given head must apply before any other head is merged into the structure. A secondary consequence, which will be of importance in the remainder of the chapter, is that insertion of adjuncts is not governed by the cycle. Adjuncts can be inserted during the course of a cycle (interspersed with operations that are triggered by the head), and they can be added post-cyclically. If (1) is supplemented with the additional condition (3), this consequence still holds, with the (irrelevant) exception that XP adjuncts cannot be merged to a projecting head that already has a sister (as in (2)).

1.1   Successive-cyclic movement

Successive-cyclic movement sits uneasily in a theory of grammar in which movement is driven strictly by features of attracting heads; we want to say in some sense that the final landing site (e.g. the interrogative C^0 in long-distance *wh*-movement) is the attracting element, and that the intermediate landing sites along the way are only incidentally targeted. Moreover, in a cyclic theory (in the sense of (1)) the ultimate landing site is not even in the structure at the point of derivation where the moving phrase targets the intermediate landing sites. Successive-cyclicity suggests that some notion of free movement co-exists with obligatory feature-driven movement; phrases optionally target positions that end up being intermediate. One way of importing free movement into a feature-driven theory is by pinning the optionality on the attracting heads (Chomsky 1998,1999). I will assume that *move* may be driven by the need to delete

---

isn't overt. This assumption is natural for relational N (*picture, friend,* etc.), but stipulative for other cases (*man*). I will leave this for now as an unresolved issue.

158

uninterpretable features of an attractor, or by an optional "extra merge" property of a functional head, as stated in (5).

(5)   **Extra Merge:** The *extra merge* property of a head *H* can be satisfied by selecting one or more phrasal categories in the c-command domain (the sister) of *H* and merging each phrase selected to a new projection of *H* (leaving a copy in the original position).

Given the availability of free movement driven by optional *extra merge* features of attractors, it is necessary to say something about how such movement is constrained. I will assume that the extra merge feature can in principle attract any phrase (perhaps only those that share some property of the attractor), and that interpretive conditions applying at interface levels will restrict the choice of final landing site. For instance if a phrase selected by *extra merge* happens to be a *wh*-phrase, it will ultimately have to end up in the specifier position of either an interrogative clause or a relative clause (in which case it yields an interpretation of the clause as a predicate).

A straightforward implementation, using mechanisms similar to those that Chomsky (1998/99) proposed for successive-cyclic A-movement, would be to assume that *wh*-phrases have some uninterpretable feature *F* that makes them eligible for attraction by heads (C and *v*) that also have *F*. The final landing site (e.g. interrogative C) deletes *F*; a succession of non-final attractors (non-interrogative C, *v*) attracts phrases with *F*, but leaves *F* undeleted. We could speculate further, for the sake of explicitness (although nothing here rests on the specific choices of this implementation), that *F* is *Mood*,[7] a property of final landing site C with the values [+/–*Interrog*] (with –*Interrog* possibly having sub-variants [+/–*Imperative*, etc.]). The *wh*-feature (*F*) on DP would have to match and agree with either [+*Interrog*] (for interrogative clauses) or

---

[7] Thanks to N. Chomsky (p.c.) for this suggestion and for very helpful discussion.

[*–Interrog*] (for relatives)—which means the *wh*-feature is *F* with no inherent value (hence is uninterpretable). Intermediate steps of movement are driven by C and *v* that attract but don't delete the *wh*-feature. We would then have to say that every non-final C, *v* has unvalued *F*, which, in the case of *v*, might make sense (since vP is propositional but without specified mood). In the case of embedded non-interrogative C, it also seems not unreasonable to suppose that the mood property need not be inherently valued—these heads receive a *mood* value under agreement with the predicates that select them (with verbs like *'say'* assigning [*–interrog*] and verbs like *'know'* ambiguously selecting either variant). These assumptions at least seem to make the right distinctions. Final landing sites for *wh*-phrases will be those with inherently valued *mood*: interrogative clauses, and those non-interrogative clauses *that are not selected* (i.e. relatives and adjuncts), hence do not enter into an agreement relation with a selecting predicate.[8]

Non-*wh*-phrases can also end up in the specifier positions licensed by *extra merge*. I will assume that non-*wh*-phrases ending up in non-interrogative *extra merge* positions are interpreted as focused (or in some cases topicalized).[9] Thus while such movement is free it has an interpretive consequence: for instance an NP undergoing HNPS (driven by the *extra merge* property of *v*) is subject to a focus interpretation; phrases ending up in the spec position of a non-interrogative C^0 are interpreted as topicalization structures. These tentative assumptions are summarized in their most general form as follows:

(6) a. Final landing sites for *wh*-movement are provided by heads that delete the *wh*-feature (interrogative or non-selected C).

---

[8] D. Pesetsky (p.c.) points out an interesting empirical problem for this implementation: namely that embedded complementizers in Irish show morphological agreement with *wh*-phrases that have undergone successive-cyclic raising (cf. McCloskey, to appear). This suggests that F on a *wh*-phrase can be inherently valued. I will put this problem aside for now.

[9] Possibly such cases, like *wh*-phrases, depend on the choice of final landing site. Cf. Rizzi 1997, Belletti 1999. Throughout, I have abstracted away from possible proliferation of functional structure at the left periphery that these investigators have argued for at length.

b.   Intermediate landing sites for *wh*-movement are provided by heads that attract but do not delete the *wh*-feature (*v*, selected non-interrogative C).

c.   Non-*wh*-specifiers ending up in *extra merge* positions are subject to topic/focus interpretation.[10]

I will assume, given the evidence for successive-cyclicity of covert *wh*-movement (chapter 3), that covert movements (while arguably subject to further constraints like scope economy) fall under the cycle. By assumption, then, covert movement is mechanically driven by properties of attracting heads and is constrained by (1) and possibly (3). Specifically, I assume that *wh*-in-situ, quantificational DPs and other XPs that are pronounced *in situ* can undergo covert phrasal movement, targeting categories whose heads have the *extra merge* property (5).

The assumption of a unified cycle for overt and covert operations—i.e. the rejection of the Y-model—raises the critical question of what distinguishes the two. One answer which is possible in principle is that there are no covert syntactic operations that involve movement of phrasal categories. This is the answer that is assumed in Chomsky (1998,1999); a covert operation is simply an agreement relation between probe and goal, without movement of the goal. This answer would explain why such phrases are pronounced *in situ*—the *in situ* position is the head of the chain. But it cannot be correct if we take seriously the evidence for covert phrase movement that has been investigated in detail in chapters 3-4.

Another possible answer is that the chains formed by overt and covert movement are treated equally by the syntax, and that (as yet unspecified) principles of the syntax-phonology interface determine for each chain whether it is pronounced at the head, at the tail, or in an intermediate position. That is the answer that was tentatively assumed in chapter four (Fox and

Nissenbaum 1999; cf. also Brody 1995, Bobaljik 1995, Pesetsky 1998, Groat and O'Neil 1994). But this answer leaves unexplained the Y-model effects that we found in chapter three. I will therefore assume that it is incorrect.

An alternative possibility which I would like to explore here is that the correct theory of the unified cycle predicts the existence of Y-model effects. Specifically, the theory that we are after will have the property that it predicts just those Y-model effects that block *post-covert-movement* insertion of a vP-adjunct (which would overgenerate PGs)—but at the same time it will not rule out the cases of *post-covert-movement* insertion of NP-adjuncts that yield extraposition. In the sections that follow I will sketch out such a theory and show that it generates a number of surprising empirical consequences.

2. Toward a cyclic theory of spellout

Under both the Y-model and the single-output alternative assumed in chapter 4 spellout was an all-at-once operation, submitting the fully formed syntactic structure to the phonology. (Cyclic effects of phrasal phonology were therefore not subsumed under the syntactic cycle). Under the unified cycle, each instance of spellout updates phonological representations computed on earlier cycles, submitting a larger piece of structure to the phonology to re-compute stress, syllabification and the like.

The basic idea that I would like to suggest is that Y-model effects largely carry over to a model of grammar in which the *spellout* operation is repeated on successive cycles. To show this will require being explicit about several (empirically warranted) assumptions. Any theory of

---

[10] Furthermore, if covert movement is subject to a *scope economy* condition that prohibits truth-conditionally inert operations with no phonological effect (as argued by Fox 1995/2000a), then (6)c is, in practice, a constraint on

sentence pronunciation has to take account of the fact that chains are (typically) assigned a pronunciation in just one position. The Y-model implicitly assumed a very simple principle for determining which position of a chain is pronounced:

(7)   **Pronunciation of chains:**
        A syntactic chain is spelled out in exactly one position—the highest position possible.

Since some movements are postponed until after spellout under the Y-model, the highest position possible turns out to be the tail of the chain.[11] Single-output models (in which LF is the representation accessed by the phonology) require a more complicated procedure for determining which position of a chain is pronounced. I will take the simpler principle (7) to be the correct one. We will see that it is very easy to maintain (7) under the cyclic model, and that there may be some empirical advantage in doing so.

2.1 Spellout applies to the internal domain

Under cyclic spellout, exactly like the Y-model, the interaction of (7) with the sequencing of operations is the deciding factor in determining which movements are overt and which ones are covert. If spellout applies before a movement has taken place on a given cycle, the chain will be assigned a pronunciation in the pre-movement position. And if spellout is delayed until after the movement, (7) will determine that the chain is pronounced in the higher position.

However, if (7) is to allow overt successive-cyclic movement, one further principle is needed that was not required under the Y-model. Clearly, if spellout applies to a structure after an initial step of movement has formed a chain α_1,α_0 as in (8)a below, then (7) will determine

---

overt categories in extra-merge positions.

[11] Or, in principle, an intermediate link (i.e. the highest link present before spellout).

that the chain is pronounced in the position of α_1, the highest position possible. Since a chain is pronounced in only one position, further steps of movement will not be able to alter the pronunciation already assigned to the chain. So even if spellout applies on a subsequent cycle to the structure (8)b, the chain—already having been assigned a pronunciation in the position α_1—will not be assigned a new pronunciation in the position α_2.

(8) a. $[\alpha_1 [...\alpha_0...]]$

     b. $[\alpha_2 ...[\alpha_1 [...\alpha_0...]]]$

This unwelcome result would prevent any chain formed by successive-cyclic movement from being pronounced in a position higher than the first step, assuming that movement and spellout are governed by the same cycle.

The problem would disappear if we could show that spellout applies not to the full structure formed on a cycle, but to a designated sub-part of the structure. For instance, if spellout applies to the inner brackets (bold type) in (8)a rather than to the entire structure, then (7) will determine only that the chain will *not* be pronounced in the position of the tail α_0 (since that is not the highest position at the time of spellout). The decision whether to pronounce the higher position α_1 would be postponed until the subsequent cycle: if the next spellout applies after further raising has formed (8)b, then the chain will not be pronounced in the intermediate position α_1 either.

If spelling out a designated sub-part of the structure on each cycle is what allows overt successive-cyclic movement, we would hope to find evidence for a principle by which the relevant sub-part is selected. One very simple and consistent algorithm that would achieve this would be to always select the *internal domain of the projection* (i.e. the sister of the head). This

164

algorithm makes some sense given the assumption that the spellout operation—like other operations—is triggered by some property of (certain) lexical items (perhaps just those heads whose projections can be domains of successive-cyclic A-bar raising). We then define the operation as applying to the *sister* of a head endowed with the *spellout* property, as in (9).[12]

(9) *Spellout* **applies to the internal domain on each cycle**

The *spellout* property of a head H is satisfied by applying rules of phonology to the sister of H.

Syntactic operations that feed the spellout operation on a given cycle can therefore have phonological consequences. For example, suppose an operation moves a phrase α to spec-HP (as in (10)a) before H satisfies its spellout feature. Movement of α prevents the chain thus formed from being spelled out by H. Only the sister of H is spelled out ((10)b)—and that domain doesn't contain the (head of the) chain. By (7), then, the tail is not assigned a pronunciation.

(10) **Overt movement:**

    a. Raising: $[_{HP}\alpha... H[...\alpha...]]$

    b. Spell out the internal domain: $[_{HP}\alpha...H[\underbrace{"...\langle\alpha\rangle..."}_{\textit{domain of spellout}}]]$ ($\langle\alpha\rangle$ unpronounced)

On the other hand, once a domain has undergone phonological rules, the (narrow) syntax cannot do anything to change the pronunciation (phonetic representation) that has been assigned by the phonology.[13] (Further phonological operations can of course alter the pronunciation, for

---

[12] This principle would necessitate a further assumption about spellout of the root structure: for instance the postulation of a root head R whose complement is the CP. This idea could be seen as related to Ross' (1967) proposal that sentences are embedded under silent performatives. (Cf. also Karttunen (1977). I thank N. Chomsky (p.c.) for drawing this connection.

[13] The acceptance of (9) as the basis for cyclic spellout is therefore radically at odds with the proposals of Bresnan (1971,1972). Bresnan argued that certain alternations in the stress pattern involving relative clauses, questions, etc. would be explained if the phonological rule assigning nuclear stress applied on each syntactic cycle,

instance by re-computing the stress pattern. However phonological rules (typically) cannot add, delete or rearrange syntactic phrases.[14]) So if spellout applies prior to raising of α to spec-HP ((11)a), α will be—irrevocably—assigned pronunciation in the pre-raising position. Nothing prohibits further operations (such as attraction of α) after spellout has applied, spellout being but one property of the head H that needs to be satisfied at some point during the cycle. Thus if H has the extra merge property as well, α can raise as the next step ((11)b). But the movement will be "covert", i.e. pronounced at the tail.

(11) **Covert movement:**

a. Spellout the internal domain: $[_{HP}...H["...\alpha..."]]$

$$\underbrace{}_{domain\ of\ spellout}$$

b. Raising: $[_{HP}\langle\alpha\rangle...H["...\alpha..."]]$ ($\langle\alpha\rangle$ unpronounced)

It should be clear that these assumptions, and in particular (9), predict Y-model effects in spite of the fact that spellout applies repeatedly in the course of a derivation. The crucial notion that this view of cyclic spellout has in common with the Y-model is that anything left in the domain of spellout when it applies will be 'caught' there by the pronunciation—even if it subsequently undergoes movement as in (11). Under the Y-model the domain of spellout is the fully formed structure, with spellout applying just once. By limiting the domain to a designated

---

and if the prominence level assigned to a phrase could be carried with it under movement. I thank D. Pesetsky (p.c.) for pointing out to me the incompatibility between my proposal and Bresnan's.

[14] Although cf. Chomsky 1999 for arguments that there are some 'post-syntactic' re-arrangements of phrases, possibly including head-movement, movement of Scandinavian object shift to its surface position, Chomsky's proposed TH/EX operation, and pure 'stylistic' movement.

subpart of the structure, (9) allows spellout to apply on successive cycles without altering the previously computed phonological representation.[15]

In sections 3 and 4 I will argue that this is the correct characterization of the distribution of overt and covert movement. The chief argument (section 3) will be a demonstration that (9), together with the view of the cycle suggested in section 1, predicts a difference in behavior between movement and adjunct-insertion with respect to Y-model effects. Specifically, it was argued in section 1 that insertion of adjuncts (an operation that involves neither selection nor attraction) is not governed by the cycle. Therefore we might expect the possibility of merging an adjunct to a previously spelled out domain. If a late-merger of this sort were possible, the consequence under cyclic spellout would be that the adjunct would be incorporated into the phonological representation on the next spellout—an 'anti-Y-model effect.' I will show that an additional principle with some independent support would distinguish extraposition (which *must* involve post-spellout insertion) from PG constructions (which cannot). In section 4 I will provide supporting evidence based on further predictions about the distribution of extraposition.

But before going on to section 3 I would like to address two more issues that are important for the discussion. The first concerns the principles governing the distribution of overt and covert operations, and the second concerns successive-cyclicity and its interaction with cyclic spellout. In both cases the assumption that (9) is the correct way of characterizing the spellout operation allows a simple statement of the relevant principles.

---

[15] N. Chomsky (p.c.) points out a possible technical problem with (9), in the case of T agreeing with quirky object NOM. If this implies that NOM has a structural case feature SC which is unvalued until it enters an agreement relation with T, then obligatory spellout of the sister of v would presumably cause the derivation to crash, as SC is unvalued at spellout. For now I will leave this as an unsolved problem requiring further investigation. Related to this is the potentially less serious problem of passive and unaccusative subjects. Notice that (9) strongly commits me to the claim that passive and unaccusative constructions lack v (alternatively that v is defective, lacking the spellout property). However cf. Legate 1999 for some evidence to the contrary.

2.2   Parameterization: a cross-linguistic typology of *wh*-fronting

English is a language that overtly raises exactly one *wh*-phrase per interrogative clause. If a language-learner is equipped with the tacit knowledge that all *wh*-phrases take scope over the interrogative clause (i.e. that multiple specifiers are used in interrogative CPs as in Bulgarian), her task will be to infer the generalization that secondary *wh*-movements are covert. Assuming the model of overt and covert movement sketched above, we can say that the acquisition task is to set a parameter on the sequencing of the *spellout* operation with respect to movements of *wh*-phrases. In order to produce the English generalization, the language-learner might posit the following condition:

(12)   **English *wh*-movement:** Apply spellout after exactly one *wh*-phrase raises to the periphery of an interrogative clause.

A consequence of (12) is that one and only one *wh*-phrase will be assigned a pronunciation at the head of the chain; any others will be 'caught' by the spellout operation that applies after the overt raising, and their raising will then be covert. That is, (12) stipulates that the second step of the derivation (13) obligatorily applies after the first step and before the third step:

(13)   Three steps in the derivation of *"Who bought what"*:

   Step one: *raise* who   [$_{CP}$ who$_i$ C^0 [who$_i$ bought what]]

   Step two: *spellout the internal domain* [$_{CP}$ who$_i$ C^0 ["$_{\langle who \rangle i}$ **bought what**"]]

   Step three: *raise* what   [$_{CP}$ who$_i$ $_{\langle what \rangle j}$ C^0 ["$_{\langle whoi \rangle i}$ **bought what$_j$**"]]

Implicit in (12) is an assumption that operations within a cycle (e.g. movements, spellout) are in principle freely sequenced and that language-learning involves setting parameters that impose

local restrictions on the sequence.[16] The sequencing restriction imposed by (12), a property of interrogative C^0 in English, is consistent with the Borer (1982) hypothesis that parameters are stated as properties of functional heads. The generalizations for *in situ* languages like Chinese and multiple-*wh*-fronting languages like Bulgarian could just as easily be characterized as parameterized properties of interrogative C^0:

(14) a. **Chinese:** Don't apply spellout after any *wh*-phrase has raised to the periphery of an interrogative clause.

b. **Bulgarian:** Don't attract any *wh*-phrase to the periphery of an interrogative clause after spellout has applied.

These generalizations are more difficult to express as parametric properties under the Y-model. The assumption of s-structure allowed them to be stated as simple representational conditions that determine what movements must apply before (and after) that level. However, the proposal to eliminate s-structure (Chomsky 1993/1995) forced a mechanical statement in terms of feature 'strength' that allowed some phrases to violate the principle *procrastinate* by raising before spellout. Some notion of *strength* and *procrastinate* may still be required in the framework of cyclic-spellout that I have sketched in this section. The suggestion of the parameters (12) and

---

[16] Notice that given the assumptions stated so far, (12) is not necessarily all the learner would have to posit. This condition alone would allow overt movement of *what* to pre-final landing sites on earlier phases. For example, (12) imposes no condition on the timing of spellout on the *vP*-cycle, since *vP* is not an interrogative clause. So *what* could in principle raise to a spec-*vP* position prior to spellout of the internal domain of *vP*, yielding (i). And then spellout on the CP-cycle, obeying (12), would yield (ii), * *'Who what bought'*.

(i)   [$_{vP}$ what$_j$ ... v ["bought ⟨what⟩$_j$"]]

(ii) [$_{CP}$ who$_i$ ⟨what⟩$_j$ C^0 ["⟨who⟩$_i$ what$_j$ ... bought ⟨what⟩$_j$"]]

Evidently the learner must also make use of a principle that prevents 'partially overt *wh*-movement', i.e. pronunciation in an intermediate landing site. A way to make sense of the idea would be to assume a general condition that *wh*-phrases 'type' the clauses over which they (overtly) take scope as either interrogatives or predicates. This plausibly relates to the condition (6)a, which would then be understood as a condition on *overt wh*-specifiers (as suggested in footnote 10 for non-*wh*-specifiers, condition (6)c). This suggestion is intended to be in the spirit of Cheng's (1991) 'clause-typing' hypothesis, which would have to be generalized to filter out 'partially overt' chains for relative clauses as well as interrogatives. Thanks to N. Chomsky, D. Pesetsky and S. Iatridou for discussion of these issues.

(14)a,b shows how both could be eliminated, at least for *wh*-movement. Whether the suggestion provides a guideline for a more general theory of overtness/covertness that makes due without *procrastinate* or *strength* will have to remain beyond the scope of this research, but obviously it is a hope that these notions are dispensable without loss of descriptive coverage.

2.3 *'Spell out the internal domain'* subsumes *Phase Impenetrability*

One further potential advantage of defining spellout as in (9) is that it provides a way of relating spellout to the condition that forces successive-cyclic movement, when taken together with recent proposals of Chomsky concerning stages of derivation. Specifically, (9) is a natural way of stating the *Phase Impenetrability Condition* (PIC) of Chomsky 1998.

Noting that several distinct aspects of derivation apparently require statements of local domains, Chomsky proposed that syntactic derivations proceed in stages, where each stage (or *phase*) is defined by accessing an array of terms and using those terms to build a piece of structure. By cyclicity (i.e. (1)), a new array of terms is accessed only after all terms in the original array have been merged into the structure and all requirements of the head have been satisfied.[17] This proposal was based foremost on considerations of economy—local evaluation of comparison sets for choosing *merge* or more complex operations. *Phases* also solved a problem relating to deletion of uninterpretable features (they need to be deleted cyclically but also remain accessible to the phonology). If *spellout* is a property of phases, uninterpretable features can be deleted cyclically and fed into the phonology on each cycle. Finally, the proposal that derivations proceed phase-by-phase allowed these two properties (local evaluation of merge vs. move and local deletion/spellout of uninterpretable features) to be coupled together with the condition that

---

[17] Except (possibly) for those that are involved in attraction of the head at the next phase, as noted in footnote 3.

forces movement to proceed in local steps. The *Phase Impenetrability Condition* (15) stipulates

that the search space for *attract/move* is as local as the structural edge of the previous phase:

(15)   **Phase Impenetrability Condition** (PIC): "In phase α with head H, the domain of H is
not accessible to operations outside α, only H and its edge are accessible to such
operations." (Chomsky 1998: 108 [21])

Phases are thus domains of successive-cyclic movement as well as for lexical access and

spellout. Exactly what projections constitute phases is an empirical question; we have seen

evidence from the effects of successive-cyclic movement that vP is a cyclic domain, in addition

to CP, DP and certain (proposition-embedding) PPs.[18]

Whatever projections turn out to be phases, the PIC imposes the condition that there is

something special about their internal domains (i.e. the sister of the head): extraction from this

domain is impossible once the next phase is reached.[19] Notice that the impenetrable domain

stipulated by PIC is identical to the domain of spellout stipulated by (9). If (9) is correct, it would

explain why this domain is impenetrable, on the assumption that spelling out a domain renders it

"frozen" at the next phase, invisible to further syntactic operations.

---

[18] But not TP—given the apparently obligatory reconstruction of subjects containing PGs reported in the first appendix to chapter two. This result converges with other empirical studies in syntax and phonology which have suggested that not every merging of two objects forms a 'cyclic' domain (in the sense of triggering the full complement of cyclic operations). In word-level phonology the various domains created by concatenating two elements divide into two groups: ordinary cyclic domains formed by stress-changing affixes like *-ic, -al,* and exceptional domains formed by such stress-neutral affixes as *-ing, -ment.* Notably, cyclic and exceptional (non-cyclic) domains may be interspersed, as in *environmentalist* and *fatalistic:* [[cyclic [[environ] –ment] –al] –ist]; [cyclic [[cyclic [fate] –al] –ist] –ic]. Cf. Halle and Vergnaud 1987.

Similarly, the study of syntax has consistently yielded evidence that clauses (CPs) are domains for cyclic movement, and a tacit working assumption has been that other domains are not cyclic in this sense unless evidence is found to the contrary. Akmajian (1975) found evidence of an NP(=DP)-cycle, from the observation of bounding effects in (complement) extraposition from NP embedded within NP. Others have since reported supporting evidence for an NP-cycle (e.g. Jacobson 1979). Additional empirical evidence for a vP-cycle is discussed in Fox (1999,2000a).

[19] This follows a suggestion of Chomsky (1999), to the effect that the structure formed on a given phase is evaluated at the next phase up. This proposal (or something similar) is required in order to allow post-spellout movements *on a current cycle,* as I have proposed (crucially) to be involved in covert movement (cf. (11))

In what follows I will tentatively adopt the hypothesis that derivations proceed in phases more or less along the lines developed in Chomsky 1998/1999.[20] The one aspect of this hypothesis that, together with the view of the cycle and of cyclic spellout sketched above, will play a crucial role is the claim that small arrays of terms are accessed in stages and 'used up' before derivation proceeds further. Specifically I will assume that phases are syntactic objects derived in the manner specified in (16).

(16)   A *phase* is a mapping from an array of terms to a syntactic object:

    a.   A phase begins with the selection of some array of terms[21] that includes a phase head *H* (=C, v, D...)

    b.   The phase is complete when all triggering properties of the head *H* are satisfied.[22]

3.   'Y-model effects' and 'anti-Y-model effects'

We saw evidence in Chapter 3 that movement tucks in below the root, with the consequence that parasitic gap constructions require late merge of the adjunct ((17)a). Those two claims account for the unacceptability of (17)b, on the assumption that late merge of the adjunct cannot follow spellout. However, if the evidence discussed by Fox and Nissenbaum 1999 (Chapter 4) is interpreted correctly then the relative clause in (17)c can and must be merged subsequent to a

---

[20] With departures from Chomsky's specific assumptions noted as they arise.

[21] I will depart from Chomsky 1998/99 in taking the array to include both lexical items and syntactic objects already constructed on previous phases; this will become clear in section 3. Chomsky defined the arrays selected in the step (16)a as Lexical Arrays, including only lexical items.

[22] This follows from (1) (with the caveat noted in footnote 3). Triggering properties are those that trigger operations: selection, attraction (including *extra merge*) and spellout. (16)b is intended to yield the effects of Chomsky's condition that "A phase head *H* is inert when the phase is complete, triggering no further operations" (Chomsky 1998: 107 [17]).

172

covert operation, hence after spellout. To complete the paradox, (17)d ought by the logic of

Chapters 3 and 4 to involve two late-merged constituents. The example is unacceptable; even the

presence of the constituent that apparently *can* be merged after a covert movement (the relative

clause) does not suffice to legitimate the one that apparently *cannot* (the vP-adjunct).

(17) a.   I filed _ [without even reading _] every book that was on the table

    b.   * I filed every book that was on the table [without even reading _]

    c.   I filed every book yesterday [that was on the table]

    d.   * I filed every book [without even reading _][that was on the table]

Let us take the investigation of PGs in chapter 3 to have taught us that (17)a has the following

derivation. Step 4 reflects the proposal of this chapter that *spellout* applies to the internal domain

of the phase (9):[23]

(18)   **Derivation of (17)a**

    step 1: *Form vP*
       [$_{vP}$ I filed every book that was on the table]

    step 2: *Raise DP*
       [$_{vP}$[I filed every book that was on the table] every book that was on the table]

    step 3: *Insert adjunct*
       [$_{vP}$[I filed every book that was on the table]_____every book that was on the table]
              [O_i without reading _$_i$]

    step 4: *SPELLOUT the internal domain*
       [$_{vP}$[I **"filed** ⟨every book that was on the table⟩**"**][without reading _] every book that was on the table]

---

[23] In this and subsequent figures I will include the verb in the domain that is spelled out, even though, as N. Chomsky (p.c.) points out, the verb root is raised to *v* and so would not be spelled out on the vP-phase.

In (17)b-d, on the other hand, spellout applies *before* raising of *every book*, steps 1-3 below. Of the two potential fourth steps, only one is actually possible, namely the one that would result in (17)c.

(19)　　**Derivations of (17)b-d:**

　　step 1:　*Form vP*
　　　[vP I filed every book]

　　step 2: *SPELLOUT the internal domain*
　　　[vP I **"filed every book"**]

　　step 3: *Raise DP*
　　　[vP [vP I **"filed every book"**] ⟨every book⟩]

　　step 4: *Insert adjunct*
　　　[vP [I **"filed every book"**][without reading_]⟨every book⟩ that was on the table]
　　　　　　　　　　　　　　　　　　∧　　　　　　　　　　　∧
　　　　　　　　　　　　　　　　　　‖　　　　　　　　　　　‖
　　　　　　　　　　　　not okay to add　　**okay to add**
　　　　　　　　　　　　post-spellout ((17)b)　**post-spellout ((17)c)**

The aim of this section is to face the question raised by this apparent conflict: Namely, what is the relevant difference between the two possible final steps of (19)? I hope to show that the question has a principled answer that is consistent with the notion of cycle defended in section 1. First I will suggest (section 3.1) that post-spellout merge of the adjunct into a still active phase is prohibited by general principle. This entails that if late merge (i.e. after a phase has been spelled out) is ever possible it will be truly 'countercyclic'—applying on a later phase, into a previously computed syntactic object. I will then argue in section 3.2 that the relevant difference between the illicit and the licit merge in step 4 has to do with linear order. The grammatical merge resulting in extraposition ((17)c) always takes place at the linear edge of the of the vP phase (as well as the edge of the previously computed DP phase). On the other hand, the ungrammatical merge resulting in the illicit PG ((17)b) nestles the adjunct in the middle of a previous phase. I

will suggest that this difference is a principled one. In section 3.3 I will provide independent

support for this *linear edge* constraint on countercyclic merge, with evidence from morphology.

3.1   Adjunct merge to a phase that remains active after spellout

Recall from section 1 that the cycle is driven by properties of selecting/attracting heads (1). Since

merge of adjuncts is neither selected nor attracted, it is not governed by the cycle. The

consequence, in principle, is that late merge of an adjunct should be possible even after a phase

has been spelled out. Under cyclic spellout, the adjunct would be incorporated into the

phonological representation on the next application of spellout. The theory developed in sections

1 and 2 thus allows for 'anti-Y-model effects'. Given this, the first question that needs to be

answered is why (17)b is ruled out by the theory. Specifically, why couldn't (17)b have the

following derivation, with spellout applying *before* insertion of the adjunct:

(20)   **Potential Derivation of (17)b**

step 1:   *Form vP*
[$_{vP}$ I filed every book that was on the table]

step 2:   *SPELLOUT the internal domain*
[$_{vP}$[I **"filed every book that was on the table"**]

step 3:   *Raise DP*
[$_{vP}$[I **"filed every book that was on the table"**] ⟨every book that was on the table⟩]

step 4:   *Insert adjunct*
[$_{vP}$[I **"filed every book that was on the table"**]_____⟨every book that was on the table⟩]
[*O,* without reading _$_i$]

I would like to propose that some instances of adjunct merge are—indirectly—subject to

the cycle. Specifically, assume that the condition (21) constrains the sequencing of spellout on

any given phase, requiring the array that determines the phase (cf. (16)a) to be 'used up' before spellout can apply.

(21)   Apply spellout only if the array is exhausted.[24]

This principle would force adjuncts to be merge before spellout, if they are included in the array. That is, if the array that is selected for the vP phase includes the adjunct, (21) will allow derivation (18) but block the derivation (20). (21) thus partially constrains anti-Y-model effects.

However, there is another way to derive anti-Y-model effects under cyclic spellout. Given that the cycle constrains only move and argument merge, late merge of an adjunct should be possible *even into a syntactic object that was constructed on an earlier phase*—i.e. into a phase that is complete in the sense of (16)b. If the syntactic object SO that was constructed on a previous phase is selected as part of the array of a new phase, along with an adjunct that is merged into SO on the new phase, the late-merged adjunct is predicted to be incorporated into the phonological representation. If (21) holds, a consequence will be that anti-Y-model effects will always be of this type.

I will assume without argument that (21) does hold. The assumption is rooted view of derivations as economical, with each phase a self-contained computational unit—meaning that the derivation is 'blind' to the possibility of spellout on a later phase. Consequently an item cannot be selected in an array unless it is going to be used during the phase determined by that array.

---

[24] Equivalently, in Chomsky's (1998/99) terms, apply spellout only if there is nothing in the workspace left to merge. Recall that Chomsky assumed the *Array* to include only lexical items (a subset of his *Numeration*).

3.2 Adjunct merge into a previously computed phase

Given (21), anti-Y-model effects are blocked unless adjunct merge is 'post-cyclic'—applying inside a syntactic object that was built on an earlier phase. Suppose the array that determines the vP phase includes the items {[$_{DP}$ I], [v], [filed], [$_{DP}$ every book], [yesterday]}, and computation proceeds until the syntactic object SO = [$_{vP}$ [$_{vP}$ I filed every book yesterday] $\langle_{every\ book}\rangle$] is derived, completing the phase. A subsequent phase may include SO in its initial array, along with SO' = [$_{CP}$ that was on the table]. Suppose that on the new phase SO' merges with the NP [*book*] embedded inside SO—specifically, into the raised DP [*every book*] that is embedded in SO. The condition (21) would not prohibit this operation, as long as it precedes spellout on the new phase.

One potential worry about allowing (any) derivations of this sort is that the *spellout* operation is forced to 'look all the way down' into a completed phase in order to incorporate an adjunct merged into an already spelled-out domain. We would thereby (arguably) lose part of the motivation for cyclic spellout: reducing computational load, allowing local mappings from the syntactic to the PF component. Suppose, however, that late merge (i.e. inside a completed phase) is constrained by the generalization (22):

(22)   **Linear Edge Generalization:**
        For any syntactic object SO accessed in an array, merge of new material is possible
        inside SO only at the linear edge.

Cyclic spellout by definition maps new syntactic structure to a PF string, (possibly) incorporating a string (or strings) computed on a previous cycle. Given (22), spellout need not 'look all the way down'; the operation merely has to access a previous phase as a unit and concatenate new material to it. While the syntactic operation merge may apply to a constituent that is deeply

embedded (late-adjunction), (22) allows *spellout* to be constrained quite locally, accessing only the syntactic objects drawn in the current array. Thus, while the merge operation in (23)a inserts an adjunct into a deeply embedded position (to NP embedded in DP embedded in vP), from the point of view of the subsequent *spellout* operation (on the CP-phase) it is no different from (23)b.

(23) a.   [vP [vP I **"filed every book yesterday"**] [DP every [NP[NP book] _____]]]
                                                                    [that **"was on the table"**]

   b.   [_____ [vP [vP I **"filed every book yesterday"**] [DP every book]]]
        never

On the other hand, suppose the new array includes the two syntactic objects shown in (24)a. The merge operation shown in (24)b is ruled out by (22); SO_2 would be merged into SO_1 with material both to its left and to its right.

(24) a.   {... SO_1 = [vP I filed every book yesterday ⟨every book⟩], SO_2 = [O_i without reading _i]...}

   b.   [vP [vP I **"filed every book yesterday"**] _____ [DP every book]]
                                                              [O_i *without reading* _i]

Notice that this commits us to some non-trivial claims about structural representations. Although the DP at the right edge of the completed phase is 'covert' (the chain that it heads was assigned a pronunciation at the tail by spellout on the vP-cycle), the presence of the DP is what is needed to distinguish this case from (23)a. Implicit in the distinction is that the covert DP is a 'real' part of the structure, not just a notation to mark scope. This corroborates a central result of chapters 3-4. Moreover, if (22) plays a role in explaining the deviance of the example, it is important that QR is to the right, a claim that we have independent evidence for if we have interpreted the

extraposition data correctly. Covert elements, then can and do play a role in the determination of linear properties of a structure.

Presumably (22) reduces to a very general principle: the terms that are accessed in computation—whether lexical items or syntactic objects computed on previous phases—are accessed as units that cannot be linearly disrupted.

I will argue in section 3.3 that this principle has independent support from facts about word structure.

3.3   Independent evidence for the *Linear Edge Generalization*

One potential independent source of evidence for the *linear edge* constraint on merging on a subsequent cycle comes from facts about word structure. Pesetsky (1979) discussed a class of morphological bracketing paradoxes involving comparative suffix *-er*. The distribution of this suffix is limited to stems of one metrical foot:

(25) a.   happier, sadder, taller, whiter

b.   *intenser, *furiouser, *impressiver, *irresponsibler

This generalization predicts wrongly that words like *unhappy* should not be able to take the *-er* suffix. Pesetsky proposed a sequence of derivational steps in word-formation (a proposal that became the basis of lexical phonology) that partially solved the problem. Pesetsky suggested that the suffix was attached to the stem first (at which point the metrical structure would allow it), followed by *un-* prefixation to form [un-[happi-er]]. This proposal required that something else be said about the resulting semantic composition problem. However, there is an alternative possibility that capitalizes on Pesetsky's proposal that structure building operations are interspersed with the phonology (in a sense, the original cyclic spellout proposal) which

eliminates the bracketing paradox. If merge is allowed to apply below the root of the structure, as

I have argued, then words like *unhappier* could have the following derivation:

(26)
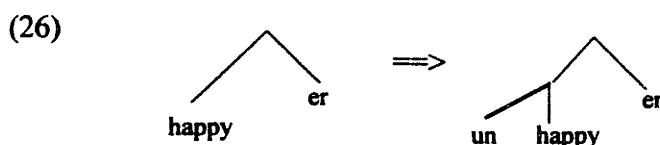
A similar derivation would work for other cases discussed by Pesetsky, such as compounds like

nuclear physicist, higher mathematician. There, the term that is embedded below a suffix (and

would have to be merged late) is not a prefix but rather a word or even in some cases a

compound.[25]

The claim that the non-head of a compound is merged countercyclically in these cases is

supported by their behavior under *-er* suffixation. Along with *trigger-happy* we get *trigger-

happier* but not **trigger-unhappy*. The unacceptability of the latter form indicates that *trigger*

and *happy* must be bracketed together. But then the constraint on *-er* suffixation suggests that

there was a stage of derivation in which *happy* and *-er* were bracketed together:

(27)
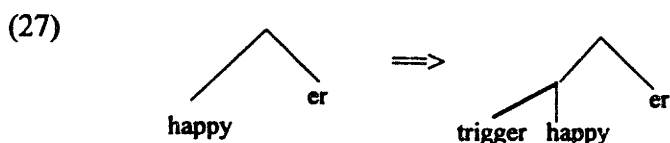
If it is correct that a derivational approach along these lines is the right solution to these

bracketing problems, we have an independent way of testing the *linear edge generalization* on

---

[25] An alternative solution to bracketing puzzles of this kind might hold that phonology applies independently to the compound head that bears the suffix, regardless of the morphological bracketing. But this strategy would not differentiate compounds whose head is on the right from those (rare) ones whose head is on the left, so it predicts that we should get forms like **artist deco* alongside of *pop artist, decorative artist.*

countercyclic merge. Consider verb particle constructions such as *throw away, screw up, spell out*.[26] Pesetsky (1995) observed that these constructions resist agentive nominalization:

(28)   *throw-awayer, *screw-upper (??screwer-upper), *spell-outer, *blow-outer, *throw-up
        (??thrower away, ??screwer up, ??speller out, ??blower out, ??thrower up)

These contrast with minimally differing verb particle constructions in which the particle appears to the left: *out-lie, uphold, upchuck* (a crude synonym for *throw up*). These words seem much more natural with the agentive nominalizer:

(29)   out-lier, upholder, up-chucker

This contrast would make sense under the *linear edge generalization*. At the point in the derivation where the verbal head of the construction has merged with the nominalizing suffix, the *linear edge generalization* distinguishes the forms in (28) from those in (29). In the former, the particle would have to merge into the middle of a previously computed syntactic object (in this case a word), an operation that is disallowed ((30)a). Late merge at the linear edge ((30)b), however, is possible.[27]

(30) a.

b.

---
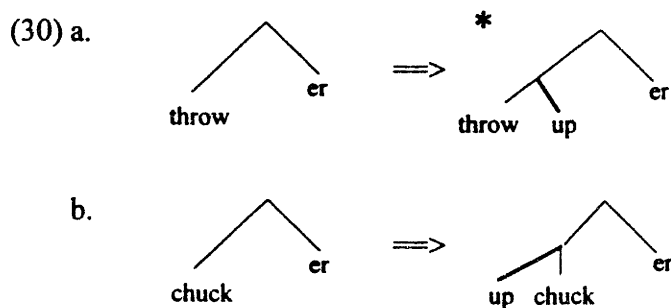
[26] I am indebted to D. Pesetsky (p.c.) for suggesting this test and for helpful discussion of the issues involved.

[27] Notice that if this is the correct analysis of the construction, then we seem to have evidence that inflectional morphology is not formed in the syntax by merging items together: inflected forms of verb-particles would not obey the linear edge generalization *(threw away/thrown away; screwed up)*. This suggests that words are accessed in their inflected forms and are (possibly) subject to something like checking theory of Chomsky 1993.

3.4   Summary

The cyclic theory of spellout in section 2 predicts that movement and argument merge will be subject to Y-model effects. These feature driven operations are forced to apply cyclically by the end of a phase; operations that apply before spellout applies to the internal domain of the phase will have phonological consequences while those that apply after spellout will not. After the phase is complete its head is inert, meaning its head cannot trigger further operations.

Non-feature driven operations—i.e. merge of adjuncts—are predicted to be immune from Y-model effects. These operations are not triggered by any property of the phase or its head. Therefore the cycle does not itself prevent adjuncts from merging indefinitely low inside previously computed syntactic objects. This consequence is desirable given the evidence from extraposition that countercyclic merge is possible (post-QR adjunction of the relative clause). However, we saw that anti-Y-model effects are overgenerated under cyclic spellout: Covert movement is apparently predicted to license PGs, by late merge of the adjunct following the movement.

I have tried to show in this section that the important distinction between these two cases lies in the fact that late merge would be forced to break up a previously computed linear sequence in order to license a PG. The argument had two parts. First I suggested that merge after spellout is prohibited *during* a phase. A requirement that spellout cannot apply on a phase until the array is exhausted dictates that nothing will be left to merge after spellout.

Therefore anti-Y-model effects can only arise by adjunct merge on a subsequent phase, into an already constructed syntactic object that is accessed in the new array. I argued that merge into a syntactic object is constrained to apply at its linear edge. This condition was shown to hold more generally, constraining word-formation as well. The linear edge generalization bars

countercyclic derivation of PGs, but allows extraposition. I will show in the next section that even extraposition is tightly constrained. When extraposition would violate the linear edge generalization, it is impossible.

4.   Further consequences

A family of predictions follows from the *linear edge condition*, as pointed out to me by Danny Fox. The predictions all take the form of the following schema: when there are two movements to the right edge of some XP, the second one will have to tuck in—so extraposition should be possible only from the phrase that moves first.

(31)
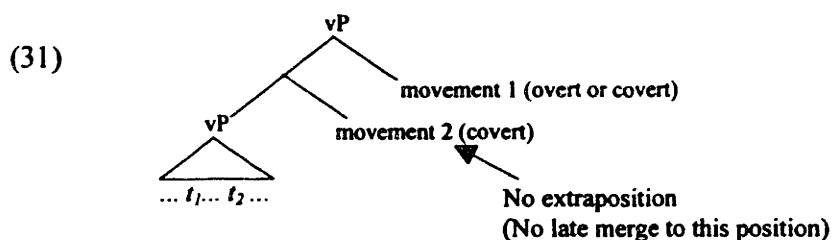
This schema yields an array of specific predictions, some of them fairly intricate and unexpected. First, multiple extraposition to the same domain should be ruled out (at least where both instances of extraposition are of the Fox/Nissenbaum late-merger type). Second, extraposition from *wh*-in-situ is predicted to divide into two classes, based on whether covert movement of the *wh*-in-situ would have to tuck in below some other phrase. Where tucking in is forced—for instance, if an overt *wh*-movement has targeted the vP—no extraposition should be possible from the *wh*-in-situ. On the other hand, in cases where covert movement of a *wh*-in-situ is the first to target a given vP, extraposition should be fine. A third prediction follows from the *linear*

edge condition together with a recent proposal of Fox (2000b) concerning the derivation of constructions with antecedent-contained deletion (ACD).

4.1   Only right-most constituents in NP can undergo (adjunct) extraposition

Only rightmost elements in an NP can extrapose. Compare (32)a with (32)b. This contrast would be explained by the Fox/Nissenbaum theory of extraposition as post-QR insertion of the adjunct, together with the *linear edge* condition proposed in section 3.2. As shown in (33), the unacceptable (32)b requires a post-cyclic merge operation that linearly disrupts the vP built on a previous phase. In Hebrew—where such adjectives appear to the right of the NPs they modify—the equivalent of (32)b is possible.[28]

(32) a.   I saw a man yesterday taller than John / who was very tall

   b.   * I saw a man yesterday very tall

(33) a.   [vP [vP I   saw a man yesterday] [DP a [NP [NP man] _____]]]
   [*who was very tall*]

   b.   * [vP [vP I   saw a man yesterday] [DP a [NP _____ [NP man]]]
   [*very tall*]

We can see further instances of this generalization in the following examples. Johnson (2000) notes that the universal quantifier in examples like (34) cannot take scope outside the DP in which it is embedded, as diagnosed by its inability to bind the variable.

(34) a.   * I showed [a report [about a book by almost every author]] to his students

---

[28] Thanks to Danny Fox (p.c.) for this observation.

    b. *I showed [a report [about a book that almost every author wrote]] to his students

On the other hand, adjunct extraposition from a similarly buried DP seems to yield the opposite result:

(35)   I bought [a report [about the investigation of almost every president]] yesterday that ever taped his private conversations

What this shows us, in light of Johnson's argument that QR is impossible out of the embedding DP, is that the embedding DP may *itself* undergo QR, with late adjunction permitted indefinitely low. But one factor that does appear to constrain how low countercyclic adjunction can go: it needs to be rightmost. This predicts the following to contrast with the above:

(36) a. *I bought [reports [about the investigations of almost every president and many chief justices]] that ever taped his conversations

     b. I bought [reports [about the investigations of many chief justices and almost every president]] yesterday that ever taped his conversations

Needless to say, the relatively acceptable member of the pair above can hardly be argued to QR out of both the Coordinate NP island and the higher embedding DPs:

    cf. *I showed [reports [about the investigations of many chief justices and almost every president]] to his biographer.

4.2   No extraposition to the left of (certain) vP-adjuncts

Extraposition is predicted to be impossible to the left of a vP-adjunct. What we actually find, interestingly, is a contrast that depends on whether the adjunct contains a PG:

(37) What book did you put _ on the table yesterday that John fixed

                                   * without even reading _ first
                                   without even reading it first

This asymmetry follows from the *linear edge* condition. The condition allows a vP-adjunct to be merged on a later cycle—subsequent to late-merge of the extraposed relative. Thus both instances of late merge satisfy the linear edge condition. This derivation is not available to adjuncts that contain PGs, since they must be inserted to the left of the licensing *wh*-phrase, hence during the phase in which *wh*-movement targets the vP.

This asymmetry appears as well when an ACD is followed by a vP-adjunct containing a PG, providing further support for the linear edge condition as well as for Fox (2000b)'s theory of ACD.

(38) What book did you put a copy of _ on the table that John fixed

                                  *without reading _ first / without reading it first

(39) What book did you put a copy of _ on the table that John told you to__

                                   *without reading _ first
   without reading it first

4.3 No extraposition from (certain) in-situ *wh*-phrases

For example, in a multiple-*wh*-construction, the *wh*-phrase that raises overtly targets the right edge of vP, but the *wh*-in-situ does not (it tucks in). Hence an extraposed relative associated with the *wh*-in-situ should not be possible. This is illustrated in (40).

(40)

No extraposition
(No late merge to this position)
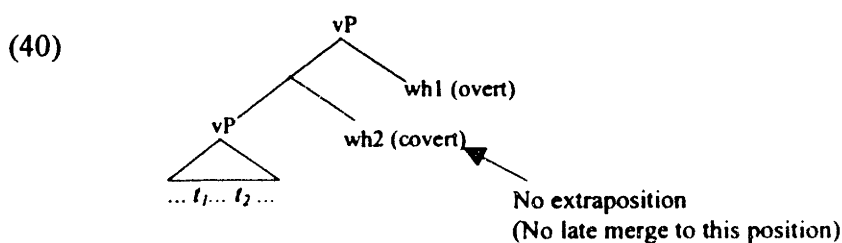
On the other hand, if the overtly raised *wh*-phrase originates from an A-position higher than vP, extraposition from the *wh*-in-situ should be acceptable—since in this case (illustrated in (41)) the covert movement will extend the vP, targeting its right edge.

(41)
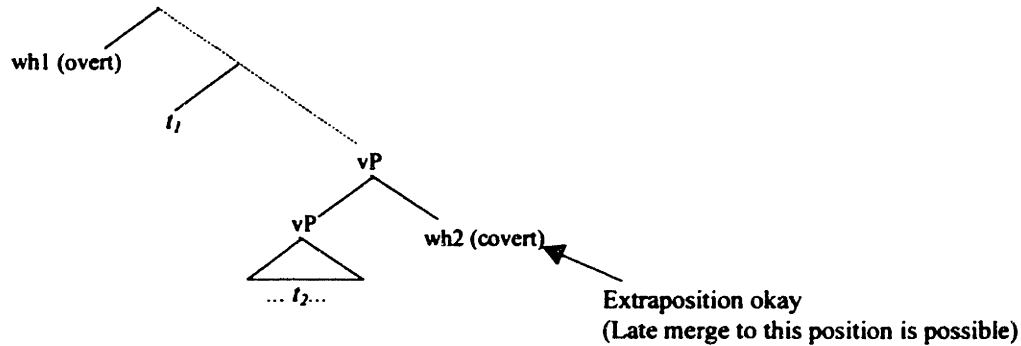
This prediction appears to be correct, although the judgements are subtle. Structure (40) is instantiated by the starred (a) examples in (42)-(44), in which both *wh*-phrases originate in the complement domain of vP. The (b) examples instantiate structure (41), in which the first *wh*-phrase originates outside the vP, allowing the *wh*-in-situ to raise to the linear edge, licensing extraposition. The (c) examples are controls designed to show that extraposition from the embedded clause is okay in principle, hence not the source of deviance in the (a) sentences.

(42) a. *?which girl did you [vp persuade _ to read *which book* by charming her] that John recommended / that John wanted you to

　　b. Which girl did you persuade _ to [vp read *which book* for her term paper] that John recommended / that John wanted her to

　　c. Who _ [persuaded Mary to read *which book* by charming her] that John recommended / that John wanted *you* to ⟨persuade her to read by charming her⟩

(43) a. *?Which patient do you [believe _ to have contracted *which disease* these days] that there's no cure for?

　　b. Which patient do you believe _ to have [contracted *which disease* last month] that there's no cure for?

c.   Which hypochondriac _ [believes himself to have contracted *which disease* these days] that there's no cure for?

(44) a. *?Which woman did you [say _ will invite which man yesterday] that she wants to dance with?

b.   Which woman did you say _ will [invite which man tomorrow] that she wants to dance with?

c.   Which woman _ [said she'll invite *which man* yesterday] that she wants to dance with?

(45)   **LF of (42)a, an instantiation of (40):**
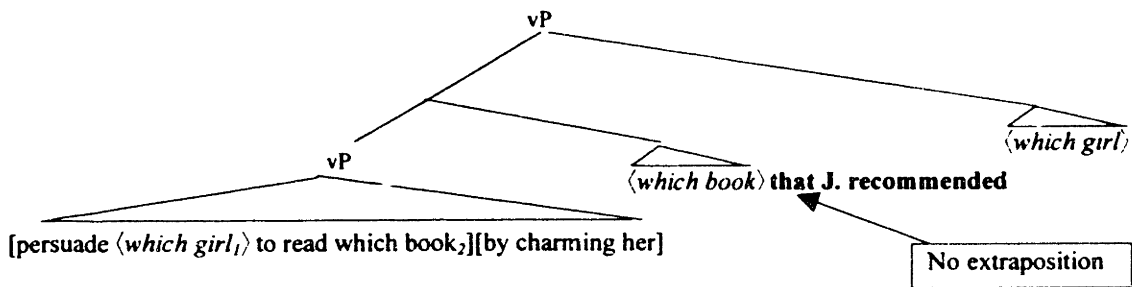
(46)   **LF of (42)b, an instantiation of (41):**
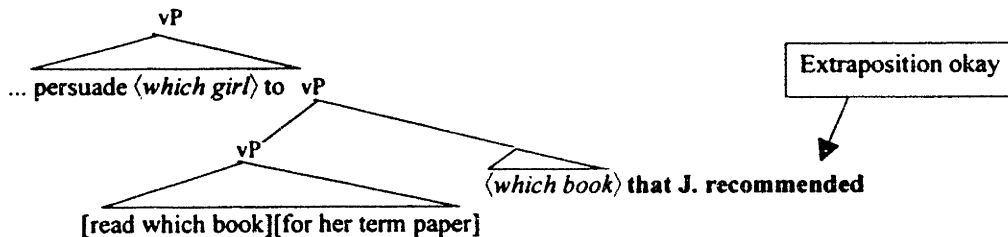
These contrasts sharpen if we use antecedent-contained deletion (ACD) in the extraposed relative. This serves as an extra guarantee that the scope of extraposition is (at least) as high as the vP in question, due to the need for parallelism between the vP in the elided sentence and the vP in the antecedent (which the *wh*-phrase targets covertly). Again, (47)a instantiates (40) and the (47)b instantiates (41):

(47) a. ?*Who did Mary introduce to which delegate (yesterday) that *you* were supposed to?

      cf. Who did Mary introduce to which delegate that you *were supposed to introduce him to?*

    b. Who introduced which delegate to Mary that *you* were supposed to?

(47)a is difficult to make sense of, even though it has a perfectly coherent paraphrase without the ellipsis, and contrasts sharply with (47)b. This fact is unexpected given standard generalizations about ACD and *wh*-in-situ but is explained on the assumption that the *wh*-phrase targets vP, tucking in below *who*, and that the relative clause must be added on a later cycle at the right edge.

Notice too that (47)a has the same status with or without 'yesterday' directly signaling that extraposition has taken place. This fact lends further support to Fox's (2000) revival of Baltin's (1987) hypothesis that ACD always involves extraposition (=late adjunction) even when there is no direct cue from the word order. With this added assumption we can test a further prediction.

4.3.1 No wide-scope ACD in certain cases of *wh*-in-situ

It is a well-known property of ACD constructions that when there is more than one potential antecedent for the ellipsis, the choice of antecedent is constrained by the scope position of the DP containing the ellipsis (Sag 1976, Larson and May 1990). The DP must take scope at least as high as the antecedent for the ellipsis. Thus interpretation (a) of the DP in (48) requires that *every city Mary did* take scope over the matrix verb *refuse*, while interpretation (b) for the DP will allow the DP to take either wide or narrow scope with respect to *refuse*.

(48)   I refused to visit [every city Mary did]
    a.   every city Mary did $\langle$refuse to visit$\rangle$
    b.   every city Mary did $\langle$visit$\rangle$

Given the fact that wide scope for the DP will allow either the higher or lower choice of antecedent, it of course follows that an overtly moved *wh*-phrase such as the one in (49) will be ambiguously interpreted as taking either the high or low antecedent for the ACD.

(49)   Which company that you weren't willing to did Mary persuade the D.A. to sue?
    a.   which company that you weren't willing to $\langle$persuade the D.A to sue$\rangle$
    b.   which company that you weren't willing to $\langle$sue$\rangle$

Fox (2000b) proposed an explanation for why the QR involved in ACD doesn't leave a full copy (i.e. why ACD constructions are immune from certain reconstruction effects). Namely, the relative clause containing the ellipsis is "extraposed" in the sense of Fox and Nissenbaum (1999)—i.e. added after QR has covertly raised the associate DP to its scope position. If this theory is correct then (48) has the following structural possibilities (and the ensuing interpretive possibilities):

(50) a.   *wide scope only*

    I [$_{vP}$ [$_{vP}$ refused to visit every city] $\langle$every city$\rangle$ Mary did $\langle$vP refuse to visit _$\rangle$]

    b.   *wide or narrow scope*

    wide: I [$_{vP}$ [$_{vP}$ refused to [$_{vP}$ visit every city]] $\langle$every city$\rangle$ Mary did $\langle$vP visit _$\rangle$]

    narrow: I refused to [$_{vP}$ [$_{vP}$ visit every city] $\langle$every city$\rangle$ Mary did $\langle$vP visit _$\rangle$]

In the derivations for (50)b, the relative clause had to be added to the structure sometime after QR to the lower vP—whether that was the final step of QR (as in the narrow scope version) or just the first step (as in the wide scope version). But in (50)a, the relative clause can only have been added after the second step of QR targeted the matrix vP.

As noted, the *wh*-phrase in (49) takes wide scope—evidenced by its surface position—and can therefore take either the higher or lower antecedent for the ACD. So again if Fox (2000b) is correct then the relative clause of (49)b had to be added to the *wh*-phrase sometime after the first step of movement targeted the lower vP. And the relative clause of (49)a had to be added after a subsequent step of movement targeted the matrix vP.

(51)   ... [vP [vP persuade the D.A. to sue $\langle$which company$\rangle$] $\langle$which company$\rangle$ that you weren't willing to $\langle$persuade the D A. to sue$\rangle$]

An ACD contained in a *wh*-in-situ should similarly be ambiguous in its choice of antecedent, given our conclusion (from chapter three) that *wh*-phrases in situ undergo covert movement to the projection in which their overt counterparts take scope.

However, on the assumption that Fox (2000b) is correct, we predict a distinction...wh-in-situ that targets a vP which was also the target of an overt wh-movement should not permit extraposition (=late insertion of a relative clause). We saw evidence of this in the last subsection. We now predict, given Fox (2000b), that ACD should not be allowed in *wh*-in-situ in exactly this configuration. On the other hand, there should be nothing wrong with an instance of ACD whose antecedent is a vP *not* targeted by the overt *wh*-movement. ... ...

(52) a.   Who did Mary persuade _ to sue which company that you weren't willing to?
     *$\langle$persuade him/her to sue$\rangle$ / $\langle$sue$\rangle$

     b.   Who persuaded Mary to sue which company that you weren't willing to?
     $\langle$persuade her to sue$\rangle$ / $\langle$sue$\rangle$

(53) a. ??Which book did you give to which girl that John asked you to?

     b.   Who gave a book to which girl that John asked him to?

Example (53)b shows that, if the covert movement of *which girl* doesn't tuck in (since it raises past the internal subject in such cases), the oddness of (53)a disappears.

(54) a. Who gave which book to which girl yesterday that (*who) John wanted YOU to?
(=which book that John wanted YOU to give to her)
(≠which girl that John wanted YOU to give it to)

(55)   Which book did you refuse to give to which girl that John hoped you would?
$\langle$refuse to give it to$\rangle$
* $\langle$give it to$\rangle$

## 4.4   No multiple extraposition to the same vP-domain

Another prediction based on the general schema (40)-(41) is that multiple extraposition should be prohibited to the same vP. The contrasts in (56) and (57) bear this out. The (a) examples are more or less acceptable on the parse indicated, with the adverbial modifying the lower vP. In these examples, the two extraposed relatives are able to appear at the right edges of different vPs (as shown in (58)), so each one satisfies the *linear edge* condition. In the (b) examples, on the other hand, the adverbial unambiguously modifies the matrix vP. Consequently both extraposed relatives must be merged at the matrix vP (59), in violation of the *linear edge* condition.

(56) a.   I persuaded someone to [present a paper tomorrow] that's over 50 pages long who isn't even registered for the class.

b. * I [persuaded someone to present a paper by charming him] that's over 50 pages long who isn't even registered for the class.

c. ?I [persuaded someone to present a paper by charming him] that's over 50 pages long

(57) a.   I persuaded no one to [present any paper next semester] that's over 50 pages who isn't registered for the class

b. *?I [persuaded no one to present any paper] yesterday (/even by pleading) that's over 50 pages who isn't registered for the class

c. I [persuaded no one who isn't registered for the class to present any paper] even by pleading that's over 50 pages.

d. I [persuaded no one to present any paper] even by pleading who isn't registered for the class

The (c) examples show that there is nothing wrong in principle with long extraposition to the matrix vP; these examples are relatively acceptable (in comparison to the (b) examples).

(58)   **LF of (56)a, an instantiation of (41):**
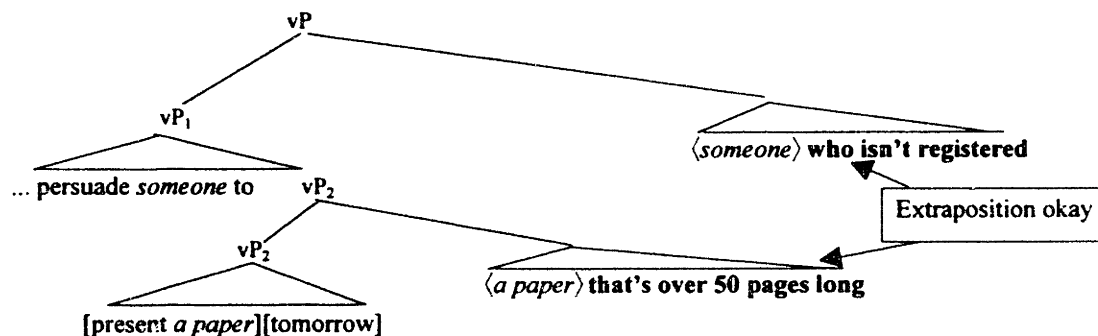
(59)   **LF of (56)b, an instantiation of (40):**
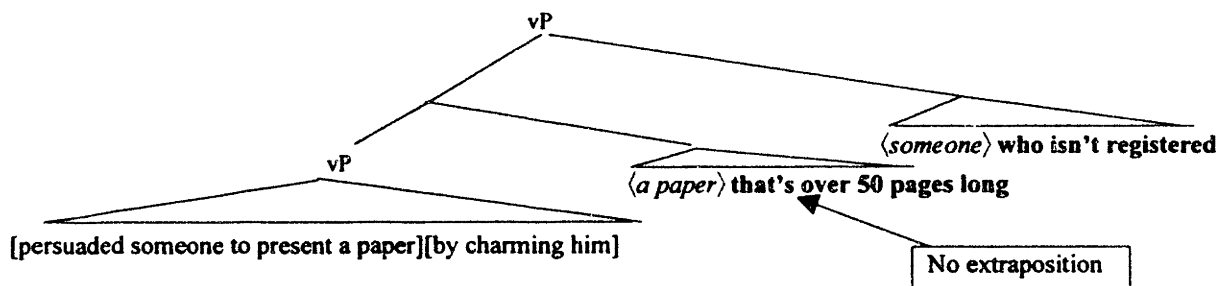
4.5   When 'tucking in' is violable

In chapter 3, we saw a case where QR was able to obviate the TIC, raising past the intermediate trace of wh-movement (indicated by the PG)—thereby licensing a (second) PG. The example,

repeated as (60), has the interesting property that the pair-list reading is forced (cf. Chapter 3 section 4.1).

(60)   ? Which car did you persuade *each republican senator* to borrow _ (after getting an opponent of _ to put a bomb in _)

Given that extraposition is the reflex of QR (chapter 4/Fox and Nissenbaum 1999), it follows that extraposition will be possible from the quantifier in a sentence like (60), yielding (61). This is, in effect, a case of extraposition (=QR) licensing a PG:

(61)   ? Which car did you persuade *each republican senator* to borrow _ (after getting an opponent of _ to put a bomb in _) (who serves on the appropriations committee)?

I speculated in Chapter 3 that QR is able to violate the TIC in such cases by the *scope economy condition* of Fox 2000a; the universal quantifier is able to scope past the *wh*-phrase (an existential) because the resulting interpretation (pair-list) is distinct from the one that would result from a shorter QR that obeys the TIC. If we can find other environments in which QR is able to obviate the TIC, we should likewise find that—just as in (61)—extraposition apparently licenses a PG.

4.5.1   Extraposition will mark widest scope among VP-internal quantifiers

The following paradigm also illustrates the pattern discussed in the previous subsection, namely that only one extraposition is possible to the same vP. (As a control, (62)c shows that two extraposed elements may appear on the right edge, as long as only one of them is extraposed from an NP internal to the vP.) These examples illustrate an additional property. Namely, when two quantifiers internal to vP create a scope ambiguity, extraposition from one of them will disambiguate the sentence in favor of wide scope (for the DP source of the extraposition).

(62) a.   I sent a student to each classroom after you told me to that had any room
       (* who had recently transferred from another department)
           *each classroom> a student, *a student > each classroom*

    b.   I sent a student to each classroom after you told me to who had recently transferred
from another department
       (* that had any room)
           **each classroom> a student, a student> each classroom*

    c.   Someone sent a student to each classroom after you told him to that had any room
       (? who was helping the principal)

(63) a.   I put a book on every table after you told me to that John wrote
          *a book > every table, *every table > a book*

    b.   I put a book on every table after you told me to that John built
          **a book > every table, every table > a book*

4.5.2   Extraordinary cases of PG licensing by extraposition

While extraposition is ordinarily unable to appear to the right of a vP-adjunct (illustrated in

(64)a), ellipsis inside the adjunct evidently allows a QP to get past it ((64)b):[29]

(64) a.   * I read a book before you read it that was on the bestseller list for six months

    b.   I read a book before you did ___ that was on the bestseller list for six months

The assumption that extraposition marks the site of QR in (64)b is corroborated by that fact that

the indefinite receives unambiguous wide scope; the sentence lacks the interpretation *"I read a*

book that was on the bestseller list for six months before you read a book that was on the

bestseller list..."

---

[29] We might conjecture here that what allows TIC to be obviated is the need to satisfy the parallelism condition on ellipsis. This conjecture assumes, contra Fox 2000a , that *scope economy* is not blind to the parallelism condition; it could turn out that the need to satisfy parallelism licenses a longer instance of QR, with the effects obscured in the cases Fox tested because the interaction was not local. In cases involving QR past a *vP-adjunct* that contains ellipsis, however, the interaction is local.

Given that QR in (64)b raises past a vP-adjunct, we might expect that the latter is forced to contain a PG, by the Obligatory PG generalization (cf. chapter 2, section 3.1). Thus, although the ellipsis would obscure the existence of the PG, we expect that (64)b has the LF shown in (65).

(65)   *LF of (64)b:*
        I read a book [O_i before you did $\langle_{vP}$ read _$_i\rangle$] [$\langle$a book$\rangle$ that was on the bestseller list...]

We can demonstrate that it is at least possible for an ellipsis inside a vP adjunct to contain a PG, with the following test. A gap in a subject can only be a PG (**John's the guy that friends of_ are crazy / John's the guy that friends of_ are crazy about_ *). A parasitic gap in a subject can only be licensed by another gap that it c-commands. We therefore conclude that if the subject within the adjunct contains a gap, there must be another gap (hence a PG) in the elided vP.

The contrast between (66) and (67)a serves as a baseline for the experiment: the unmoved NP in (66) doesn't license parasitic gaps, while the Heavy-NP-Shift in (66)a does. Moreover, HNPS past an adjunct that contains ellipsis, as in (67)b, also seems to license parasitic gaps with little or no degradation. This example indicates that it is possible for an elided vP to contain a PG (whose existence can be inferred by the presence of the subject PG, on the reasoning indicated above).

(66) a. * Yesterday I visited MILTON, and TODAY I visited MARY after learning that friends of often end up hating.

(67) a. Yesterday I visited someone who's staying at the HILTON, and TODAY I visited — after learning that friends of often end up hating — someone who's staying at the MARIOTT.

     b. Yesterday I visited someone who's staying at the HILTON, and TODAY I visited — after learning that friends of also would — someone who's staying at the MARIOTT.

The next step is to see whether extraposition will license PGs just in case the adjunct contains ellipsis—as we have postulated for (64)b. The crucial examples are those in (68).[30]

(68) a.   ?Yesterday I visited someone who's staying at the HILTON, and TODAY I visited someone — after learning that friends of also would — who's staying at the MARIOTT.

    b.   *Yesterday I visited someone who's staying at the HILTON, and TODAY I visited someone — after learning that friends of often end up hating — who's staying at the MARIOTT.

---

[30] Note that we expect an example like (i)—which is nearly identical to (68)a except that it has phonological reduction *(de-accenting)* in place of the ellipsis—to be equally acceptable. If the parallelism condition is able to motivate a QR past the adjunct, as I speculated was the case for (68)a, then parallelism should likewise motivate a TIC-violating movement past the adjunct in (i). In my own judgment, (i) has a status somewhere in between that of (68)a and (68)b. Some—but not all—of the speakers who found (68)a acceptable found (i) equally acceptable or at least not nearly as bad as (68)b.

(i)   Yesterday I visited someone who's staying at the HILTON, and TODAY I visited someone — after learning that friends of would ALSO visit — who's staying at the MARIOTT.

In a similar vein, if all this is right (and the contrast in (68) suggests that it is), then we ought to expect that QR without extraposition would be sufficient under the right circumstances to license a PG. For some reason, QR that fails to tuck in appears to be pretty difficult in the absence of major assistance from the pronunciation (intonational contour). It is possible that extraposition serves as an extra aid to the parse in which QR has extended past an adjunct. But I believe the contrast in (ii)-(iii) is not imagined.

(ii) ??Yesterday I visited MANY of the patients on the seventh floor, and TODAY I visited EVERYONE after learning that friends of ALSO would.

(iii) *Yesterday I visited MANY of the patients on the 7th floor, and TODAY I visited EVERYONE after learning that friends of would introduce me to.

At this point there are two major questions left open. First, why do the contrasts sharpen when QR is indicated by extraposition (i.e. why is (68)a better than (ii), at least for many of the speakers whose judgments were solicited)? Second, why do the contrasts sharpen (again, for many speakers) when parallelism is indicated by ellipsis (as in (68)a) rather than downstressing (as in (i))? The results from a small sample of speakers indicate a gradient pattern of judgments, for both of these questions: *ellipsis* > *deaccenting* > **no parallelism at all*; similarly *extraposition* > *plain QR* > **no QR.*.

5. Some possible extensions: On the absence of *wh*-islands for covert movement

One of the arguments given in section 2 for the claim that spellout applies cyclically to the internal domain of each phase was that it allows a simple characterization of the parametric property that determined the distribution of overt and covert *wh*-movements. Here I will try to strengthen that argument by showing that the claimed *wh*-spellout parameter explains why only overt movement is subject to *wh*-islands. The basis for the explanation is the claim that there is no such thing as a *wh*-island; the deviance of so-called *wh*-island violations results from violating the sequence of operations imposed by the spellout parameter setting. The movement itself is not blocked. Since this view of the distribution of overt and covert movement allows us to dispense with the claim that *wh*-islands exist, it follows that covert movement should not violate them. I will show in section 5.4 that this characterization of *wh*-islands predicts their absence in languages with different settings for the *wh*-spellout parameter—languages of the Bulgarian type that pronounce all *wh*-phrases in spec-CP, and languages like Chinese which have been argued to raise *wh*-phrases to spec-CP but pronounce none of them there.

5.1 Raising out of an embedded question: the 'Baker-ambiguity'

Recall the 'Baker-ambiguity' that was discussed in the appendix to chapter three: Example (69) can be interpreted as either a singular *wh*-question with an embedded multiple question (with felicitous answers like (70)a), or as a multiple question with an embedded singular *wh*-question (with felicitous answers like (70)b).

(69)   Who asked who bought what?

(70) a.   JOHN asked who bought what.

    b.   JOHN asked who bought the SOUP, and MARY asked who bought the WINE.

The LF representations of the two readings of (69) corresponding to the answers (70)a and (70)b both involve extra spec-CP positions formed by covert movement of *what*. In the embedded multiple-*wh* question interpretation, represented as (71)a, only the embedded CP has two *wh*-specifiers. In the matrix multiple-interrogative interpretation, however, *what* has raised successive-cyclically out of the embedded question on its way to the matrix spec-CP, leaving a copy in the embedded second spec-CP position (shown in (71)b).

(71) a.   [$_{CP}$ who$_i$ ... asked [$_{CP}$ who$_j$ $_{(what)k}$ ... bought what$_k$]]

    b.   [$_{CP}$ who$_i$ $_{(what)k}$ ... asked [$_{CP}$ who$_j$ $_{(what)k}$ ...bought what$_k$]]

We can be fairly certain that the matrix multiple-interrogative interpretation ((70)b, (71)b) involves covert raising of *what* out of the embedded CP. Evidence of the kind that we saw in chapter three supports the conclusion: the posited covert movement feeds Condition A (72) and licenses a parasitic gap in a matrix-vP-adjunct (73).

(72) Covert movement out of the embedded CP feeds Condition A

    a.   ?*WHICH BOY*$_i$ asked who bought *WHICH PICTURE OF HIMSELF*$_i$?
        cf. **Which boy*$_i$ *asked who/which girl bought pictures of himself*$_i$

    b.   [$_{CP}$ which boy $_{(which picture of himself)}$ ... asked [$_{CP}$ ...]]

(73) Covert movement out of the embedded CP licenses a PG in a matrix adjunct

    a.   ?*WHICH BOY* did you ask who bought *WHICH PICTURE* after asking friends of *WHO PAINTED*?

    b.   ...[$_{matrix vP}$ <u>(which boy)</u> <u>(which picture)</u>[$_{vP}$...[$_{embedded CP}$...]][$_{adjunct}$$O_1$$O_2$...PG$_1$...PG$_2$]$_{matrix vP}$]

But while these facts provide evidence that the embedded *wh*-phrase has raised to the matrix clause, how can we know whether the raising was successive-cyclic? It could turn out that the covert movement skips the embedded spec-CP position in these derivations. This is exactly what Richards (1997) proposed: raising out of the *wh*-island is licensed in these cases by the fact that the matrix *wh*-phrase has not crossed any *wh*-islands (satisfying Richards' PMC; cf. chapter 3).

Richards' proposal is incompatible with our assumption that successive-cyclic movement is forced by the PIC (15), which in turn is a reflex of cyclic spellout (9). It would not do to stipulate that Richards' PMC can license a violation of PIC; this would amount to a claim that spellout fails to apply on the embedded CP-cycle.

Moreover, facts like (74) cast some doubt on Richards' claim that movement out of embedded questions is not successive cyclic. If the matrix C^0, which hosts an island-satisfying *wh*-specifier, is thereby free to attract a second *wh*-phrase past an island (by PMC), we would not expect to detect evidence of another intermediate landing site between the island and the final landing site. But this evidence is present in (74): an intermediate trace of *which picture of himself* must be present somewhere in between the embedded subject *John* and the island in order for the reflexive to be locally bound to *John*. As (75) illustrates, the assumption of successive-cyclic movement creates the required intermediate copy, in the spec-vP$_2$ position just below *John*.

(74)   ?*WHICH* of his sisters thinks John$_i$ asked who bought *WHICH* picture of himself$_i$?

   cf. **Which of his sisters thinks John asked how many girls bought pictures of himself*

(75)   [$_{CP}$*Which* of his sisters ... thinks [$_{CP}$ J. asked [$_{CP}$who ... bought *WHICH* picture of himself]]]?

While this does not give us direct evidence for a trace in the most embedded spec-CP, it does argue that the *wh*-in-situ crosses the island to target a projection (namely spec-vP$_2$) that is not also targeted by a non-island-crossing movement—in apparent violation of the PMC.

Recalling the discussion in chapter three, we have yet another reason to be skeptical that the PMC is involved in extraction from *wh*-islands. One potential argument in favor of viewing the PMC as the principle that allows movement past the embedded interrogative CP is its generality: Richards argued that the same principle licensed apparent violations of superiority (giving rise to free ordering of second and third *wh*-phrases in Bulgarian), as well as PGs. However, in chapter three we proposed an alternative explanation for both the Bulgarian ordering puzzle, and in chapter 2 an alternative theory of PG licensing. The alternative to the Bulgarian puzzle capitalized on Richards' generalization that secondary movements tuck in; I argued that if tucking in is stated simply as a condition that blocks tree-extending movements (in the relevant environments), there is no need to assume that superiority is relaxed for second and third instances of *wh*-attraction. PMC is thus superfluous in the account, and the 'generality' argument for PMC loses some force.

Given these grounds for suspicion about the claim that PMC is what licenses movement past the island, let us take the facts as showing that successive cyclicity (PIC) is obeyed in all of these derivations. We are still faced with the question of how extraction from the embedded interrogative works. But now we are in a position to isolate the potentially problematic step: covert movement from the embedded interrogative CP (labeled CP$_2$ in (76)) to spec-vP$_1$. And we can isolate the potential problem with this step: Given that *what* must have tucked in below *who* in spec-CP$_2$, how can v_1 attract it (apparently violating locality)?

(76) $[_{CP_1}$ who$_i$ $\langle$what$\rangle_k$ ...$[_{vP_1}$ $\langle$what$\rangle_k$...asked $[_{CP_2}$ who$_j$ $\langle$what$\rangle_k$ ...$[_{vP_2}$ $\langle$what$\rangle_k$...bought what$_k$]]]]

      ↑spec-CP$_1$ / ↑spec-vP$_1$ / ↑spec-CP$_2$ / ↑spec-vP$_2$ /

5.2   How is attraction of a tucked in *wh*-phrase possible?

Given the preceding discussion, Baker-sentences like (69) lead us to a curious conclusion about embedded interrogative complementizers. In a configuration like (77), which zooms in on the potentially problematic step of (76), we have indirect empirical evidence that v can 'bypass' wh_1 to attract (covertly) the tucked in wh_2. On the surface this looks like a violation of locality ('attract closest'), a condition whose attested effects include enforcement of 'superiority' when the embedded complementizer is *non*-interrogative (78).

(77)   ... v ... $[_{CP}$ **wh$_1$** **wh$_2$** $C_{[interrogative]}$ ...   attraction of wh_2 possible (Baker-sentences)

(78)   ... v ... $[_{CP}$ **wh$_1$** **wh$_2$** $C_{[NON\text{-}interrogative]}$ ... wh_1 must be attracted first (superiority)

At the relevant point in the derivation—when v first chooses a wh-phrase to attract—the difference between the two cases lies in whether the embedded C^0 is interrogative. This suggests that interrogative C^0 has some property that renders wh_1 invisible to attraction by v.

In section 1.1, a general schema was proposed for implementing successive-cyclic wh-movement using similar mechanisms to those proposed by Chomsky (1998/99) for successive-cyclic A-movement. *Wh*-phrases have some uninterpretable feature F that makes them eligible for attraction by heads (C and v) that also have F. The final landing site (e.g. interrogative C) deletes F on the wh-phrase; a succession of non-final attractors (non-interrogative C, v) attracts phrases with F but does not delete F.

The difference between (77) and (78) falls out from this characterization of successive-cyclicity. The interrogative C in (77) first attracts wh_1 and deletes its uninterpretable feature. This

renders wh_1 ineligible as a candidate for attraction by the next successive-cyclic head (v_2). In

(78), by contrast, C_1 is non-interrogative and so does not delete the uninterpretable feature of

wh_1. Locality of *wh*-attraction is obeyed in both cases.[31]

5.3   *Wh*-islands reduce to violations of the *wh-spellout* parameter

As the discussion of Baker-sentences in the last sub-section showed, the theory of successive-

cyclic *wh*-movement assumed here predicts that raising of a *wh*-phrase out of an embedded

interrogative CP should be possible. Given the interpretive possibilities for Baker sentences, this

result is empirically desirable. However, as is well-known, raising out of embedded

interrogatives is only possible in English when the movement is covert. These environments are

islands for overt movement:[32]

---

[31] An additional assumption that seems empirically motivated by the Baker-ambiguity is that interrogative C *can*, but need not, delete the uninterpretable features of second (and subsequent) *wh*-specifiers. For multiple *wh*-interrogatives, C deletes the uninterpretable features of all *wh*-specifiers that are attracted, blocking further movement.

A prediction if this picture is correct in its most general form is that the embedded question in a Baker-sentence should *either* be a single *wh*-interrogative, with all but the overt *wh*-specifier construed with the matrix question (as in (i) below), or a multiple-interrogative, allowing *none* of the embedded *wh*-phrases to escape to the higher CP (ii). There should be no 'mixed' interpretations as a matrix multiple-*wh*-question that embeds a multiple-question, as in (iii)—unless the matrix *wh*-phrases all originate outside of the embedded question (iv). This seems to me to be correct:

(i)   Who wondered WHICH UNCLE gave WHICH PRESENT to WHICH KID?   (*I did.*)

(ii)   WHICH UNCLE wondered who gave WHICH PRESENT to WHICH KID?
         (Uncle Paul wondered who gave the toy gun to Kyle, and Uncle Jeff wondered who gave the horrible dress to Phoebe)

(iii) a. WHICH UNCLE wondered who gave what to WHICH KID?
         (≠ Uncle Paul wondered who (all) gave what to Kyle, Uncle Dan wondered who (all) gave what to Phoebe)

    b.   WHO wondered which uncle gave which present to WHOM? *(same—no 'mixed' reading with embedded pair-list)*

    c.   WHO wondered which uncle gave WHAT to which kid?   *(same)*

(iv)   WHICH UNCLE asked WHICH KID who gave what to whom?
         (Uncle Paul asked Kyle who (all) gave what to whom, and Uncle Jeff asked Phoebe who (all) gave what to whom)

[32] Chomsky 1998/99 proposed that *wh*-islands exemplify a type of locality effect resulting from intervention, claimed to hold when the closest term with the required feature is 'inactive' due to its uninterpretable features being deleted ('defective intervention effects'). But there are several reasons to suppose that *wh*-islands are not defective intervention effects. (i) The assumptions stated above—that the uninterpretable feature that makes *wh*-phrases attractable is what is deleted in the final landing site—is, in my view, simpler. (ii) Covert *wh*-movement doesn't

(79)   *? What did you ask who bought?

It turns out, though, that this is exactly what is expected given the theory of spellout developed in section 2, and in particular claim that the distribution cf overt and covert *wh*-movement is a parameterized property of interrogative C. The parameter setting for languages like English is repeated below:

(80)   **English *wh*-movement:** Apply spellout after exactly one *wh*-phrase raises to the periphery of an interrogative clause.

If (80) is inferred by language-learners as the basis for the generalization that exactly one *wh*-phrase is raised overtly for each interrogative clause, the deviance of sentences like (79) will follow. Recall that (80) imposed the ordering of the three steps shown in (81) (repeated below) for the sentence that is embedded in (79). Since *what* was assigned a pronunciation *in situ* (step 2), there is no way for the chain to be re-assigned a pronunciation at the head.

(81)   Three steps in the derivation of *"Who bought what"*:

Step one: *raise* who $[_{CP}$ who$_i$ C^0 [who$_i$ bought what]]

Step two: *spellout the internal domain* $[_{CP}$ who$_i$ C^0 ["$_{\langle who \rangle i}$ **bought what**"]]

Step three: *raise* what $[_{CP}$ who$_i$ $_{\langle what \rangle j}$ C^0 ["$_{\langle whoi \rangle i}$ **bought what$_j$**"]]

---

obey *wh*-islands, and we have seen evidence that it is indeed covert phrasal movement; yet we wouldn't want to say that defective intervention constrains only overt movement. (iii) *Wh*-islands don't exist in all languages, but we would not want to assume 'defective intervention' is open to parametric variation. The assumption that there is no intervention eliminates these problems.

Consequently, (79)—which embeds this sentence—cannot be derived without violating the spellout parameter (80),[33] postponing spellout on the embedded CP-cycle until both *wh*-phrases have raised.[34]

5.4   Predicting a cross-linguistic typology of *wh*-island effects

Richards (1997), building on observations of Reinhart (1979), Comorovski (1986) and Rudin (1988) developed a typological distribution of *wh*-island effects. W*h*-islands are found in 'mixed' languages like English that raise some but not all *wh*-phrases. However, languages that raise all *wh*-phrases to spec-CP uniformly in the overt (Bulgarian) or in the covert (Chinese) syntax, on the other hand, lack *wh*-islands:

(82) **Mixed (overt & covert *wh*-movement):** *English*

a.   *Who* do you think _ stole *what*

***wh*-island effects exist:**

b.   * *What* do you wonder *who* _ stole _

(83) **Uniformly overt (a) or covert (b) *wh*-movement to spec-CP:**

a.   *Koj kakvo* vidjahte _ da otkradva _   (Bulgarian)
      Who what see-2pl _   steal-3sg. _
      'Who did you see steal what?'

---

[33] Or, alternatively, by violating 'superiority' (attract the closest *wh*-phrase) by raising *what* first and assigning *who* a pronunciation *in situ*. Given that superiority is apparently violable in some environments (although Pesetsky (in press) argues that those violations are only apparent), we might ask what goes wrong with a derivation that raises *what* overtly over D-linked *which woman*, yielding **What did you ask I persuaded which woman to buy*. Here again (see footnote 16), we apparently have reason to invoke a proposal like Cheng's (1991) 'clause-typing' hypothesis, requiring that a *wh*-interrogative clause in an English-type language (with no overt C^0) be overtly marked with a *wh*-specifier.

[34] A potential question is what goes wrong with the derivation that *obeys* (80), raising *what* covertly to the matrix interrogative C to form **You asked who bought what* with the intended interpretation as a root question (a covert version of (79)). An answer to this question is similarly provided by adopting Cheng's (1991) clause-typing hypothesis (see previous footnote).

b. Ni   ren-wei *shei* tou-le   *shen-me*   (Chinese)
   You think   *who* steal-PERF *what*

   'Who do you think stole what?'

No *wh*-island effects:

c. Kakvo se čudiš   koj _ otkradna _   (Bulgarian)
   What wonder-2sg *who* _ stole   _

   'What do you wonder who stole?'

d. Ni   xiang-zhidao *shei* tou-le   *shen-me*   (Chinese)
   You wonder   *who* steal-PERF *what*

   'What do you wonder who stole?'

In the previous sub-section I argued that the existence of *wh*-islands in English is misleading and that interrogatives are not actually islands for extraction. The possibility of covert *wh*-movement out of these supposed islands supports this position. The lack of *wh*-islands in Bulgarian and Chinese similarly supports this claim.

Recall the parameter settings that were suggested in section 2.2 for *in situ* languages like Chinese and multiple-*wh*-fronting languages like Bulgarian (corresponding to the English parameter setting in (80)):

(84) a. **Chinese:** Don't apply spellout after any *wh*-phrase has raised to the periphery of an interrogative clause.

   b. **Bulgarian:** Don't attract any *wh*-phrase to the periphery of an interrogative clause after spellout has applied.

From these local restrictions on the sequencing of spellout the distribution of *wh*-islands illustrated in (82)-(83) follows. In Bulgarian the embedded CP can—and must—delay spellout until both *wh*-phrases have raised to the periphery (85). Interrogative C is allowed to spell out its internal domain once multiple specifiers have been attracted, so island effects are will be absent.

206

(85)   Three steps in the derivation of *"What do you wonder who stole"* (Bulgarian)[35]

Step one: *raise* who   ...[$_{CP}$ koj$_i$ C^0 [koj$_i$ otkradna kakvo]]

Step two: *raise* what   ...[$_{CP}$ koj$_i$ kakvo$_j$ C^0 [koj$_i$ otkradna kakvo$_j$]]

Step three: *spellout the internal domain*   ...[$_{CP}$ koj$_i$ kakvo$_j$ C^0 ["⟨koj⟩$_i$ **otkradna** ⟨kakvo⟩$_j$"]]

The Chinese parameter setting will similarly yield a derivation with no island effects. Spellout will apply as the first step a, prior to any *wh*-movement to spec-CP.

Step one: *spellout the internal domain*   ...[$_{CP}$ C^0 [*"shei* tou-le *shen-me"*]]
                                                    who   stole   what

Step two: *raise* who   ...[$_{CP}$ ⟨shei⟩$_i$ C^0 [*"shei$_i$* tou-le *shen-me"*]]
                                                     who

Step three: *raise* what   ...[$_{CP}$ ⟨shei⟩$_i$ ⟨shen-me⟩$_j$ C^0[*"shei$_i$* tou-le *shen-me$_j$"*]]
                                                     who   what

## 5.5   Path Containment Condition effects

An additional argument in support of this picture comes from Path Containment effects. In chapter 3 I discussed Pesetsky's observation that cases of multiple overt *wh*-movement in which one of the *wh*-phrases moves over the other create nested rather than crossing dependencies. Although such cases involve *wh*-island violations and are therefore somewhat marginal, there is a clear and strong constraint on their interpretation. The lower gap can only be construed as bound by the higher of the two operators:

[35] I am abstracting away from V raising in this derivation.

(86) a. ??This Volvo is one car [*O* that I know [who we can convince _ to buy _]]

b. *Trent Lott is one man [*O* that I know [what car we can convince _ to buy _]]
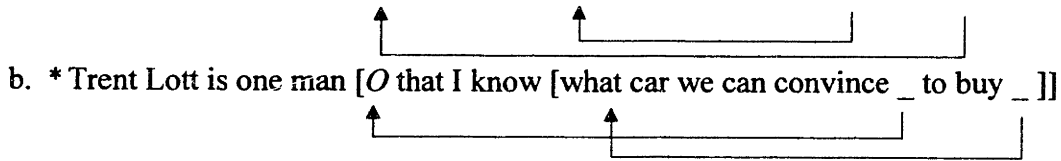
Pesetsky's *Path Containment* generalization raises an interesting question in light of the evidence both for tucking in and for successive-cyclic movement. Despite the superficial appearance of nested dependencies in the interpretable (a) example, we expect the actual chains to form an intricate pattern of nested paths all the way up to the *wh*-island, as in (87).

(87)   ... [who _ we can _ _convince _ to [_ buy _]]
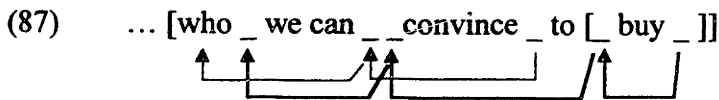
Both of these chains have to consist entirely of overt movements, given the surface form (86)a. As we saw in section 5.3, this is what creates the island violation—specifically, the fact that spellout applies to the interrogative CP in (87) after both *wh*-phrase target its periphery. The interesting fact, however, is what Pesetsky's contrast in (86) tells us about the *next* step. Namely, the only possible continuation of (87) is (88), with the *lower* of the two *wh*-phrases attracted by the embedding vP.

(88)   ... [who _ we can _ _convince _ to [_ buy _]]
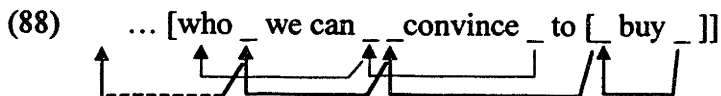
This pattern is familiar from the discussion of Baker sentences in sections 5.1 and 5.2. All of the steps in (87) obeyed superiority; on each cycle the higher of the *wh*-phrases had to be attracted first, with the second one tucking in. But the embedded interrogative C provides the final landing site for the first *wh*-phrase that it attracts—deleting its uninterpretable feature and rendering it

unattractable. Consequently, the *wh*-phrase that is tucked in is the only one eligible for attraction. out of the interrogative clause (88). The island violation is only incidental; if movement of the lower *wh*-phrase had been covert (as in Baker sentences) the pattern would have been exactly the same (minus the violation).

References

Baker, C.L. (1970). "Notes on the description of English questions: the role of an abstract question morpheme." *Foundations of Language* 6: 197-217.

Baltin, M. (1987). "Do antecedent-contained deletions exist?" *Linguistic Inquiry* 18:579-595.

Barss, A. (1986). *Chains and anaphoric dependence: On reconstruction and its implications.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Barss, A. and H. Lasnik (1986). "A Note on Anaphora and Double Objects." *Linguistic Inquiry* 17: 347-354.

Belletti, A. (1999). "'Inversion' as focalization and related questions." In *Catalan Working Papers in Linguistics* 7: 9-45. Bellaterra: Universitat Autònoma de Barcelona Servei de Publicacions.

Bobaljik, J. (1995). *Morphosyntax: the syntax of verbal inflection.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Borer, H. (1982) *Parametric variations in clitic constructions.* Doctoral dissertation, MIT.

Bošković, Ž. (1995). "On certain violations of the Superiority Condition, AgrO, and economy of derivation," ms. University of Connecticut, Storrs.

Bošković, Ž. (1997). "Multiple WH-fronting and economy of derivation." In *Proceedings of WCCFL 16,* ed. E. Curtis, J. Lyle and G. Webster. Stanford: CSLI.

Bresnan, J. (1971). "Sentence stress and syntactic transformations." *Language* 47: 257-281.

Bresnan, J. (1972). "Stress and syntax: a reply." *Language* 48:326-342.

Bresnan, J. (1976). On the form and function of transformations." *Linguistic Inquiry* 7: 3-40.

Brody, M. (1995). *Lexico-Logical Form.* Cambridge: MIT Press.

Browning, M. (1987). *Null operator constructions.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Cheng, L. (1991). *On the typology of wh-questions.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Chomsky, N. (1973). "Conditions on transformations." In *A Festschrift for Morris Halle,* ed. S. Anderson and P. Kiparsky. New York: Holt, Rinehart and Winston.

Chomsky, N. (1977). "On *wh*-movement." In *Formal Syntax,* ed. P. Culicover, T. Wasow and A. Akmajian. New York: Academic Press.

Chomsky, N. (1982). *Some concepts and consequences of the theory of Government and Binding*. Cambridge: MIT Press.

Chomsky, N. (1986). *Barriers*. Cambridge: MIT Press.

Chomsky, N. (1993). "A minimalist program for linguistic theory." In *The View From Building 20: Essays in Honor of Sylvain Bromberger*, ed. Hale, K. and S. J. Keyser . Cambridge: MIT Press. [Republished as chapter three of Chomsky 1995].

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

Chomsky, N. (1998). "Minimalist inquiries: the framework." MIT Occasional Papers in Linguistics. Cambridge: MITWPL. [Republished in *Step by Step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. R. Martin et al. (2000). 89-155. Cambridge: MIT Press]

Chomsky, N. (1999). "Derivation by phase." MIT Occasional Papers in Linguistics. Cambridge: MITWPL. [Republished in *Ken Hale: A life in language*, ed. M. Kenstowicz. Cambridge: MIT Press (in press).]

Comorovski, I. (1986). "Multiple-*wh*-movement in Romanian." *Linguistic Inquiry* 17: 171-177.

Culicover, P. and M. Rochemont (1990). "Extraposition and the complement principle," *Linguistic Inquiry* 21: 23-48.

Culicover, P. and M. Rochemont (1992). "Adjunct extraposition from NP and the ECP," *Linguistic Inquiry* 23: 496-501.

Engdahl, E. (1983). "Parasitic Gaps." *Linguistics and Philosophy* 6: 5-34.

Fiengo, R. and J. Higginbotham (1980). "Opacity in NP," *Linguistic Analysis* 7: 395-421.

Fox, D. (1995). "Economy and scope." *Natural Language Semantics* 3: 283-341.

Fox, D. (1999), "Reconstruction, binding theory and the interpretation of chains," *Linguistic Inquiry* 30: 157-196.

Fox, D. (2000)a. *Economy and semantic interpretation*. Cambridge: MIT Press / MITWPL.

Fox, D. (2000)b. "Antecedent-contained deletion and the copy theory of movement," ms. Harvard University.

Fox, D. and J. Nissenbaum (1999). "Extraposition and scope: a case for overt QR." In *Proceedings of WCCFL 18*, ed. S. Bird, A. Carnie, J. D. Haugen, and P. Norquest. Somerville: Cascadilla Press.

Fox, D. and J. Nissenbaum (in progress). "Extraposition and the nature of covert movement," ms., Harvard University, MIT.

Friedin, R. (1986). "Fundamental issues in the theory of binding." In *Studies in the Acquistion of Anaphora,* ed. B. Lust. Dordrecht: Reidel.

Gazdar, G., G. Klein, G. Pullum and I. Sag (1985). *Generalized Phrase Structure Grammar.* Oxford: Blackwell.

Groat, Erich and J. O'Neil (1994). "Spellout at the LF interface," ms., Harvard University.

Guéron, J. and R. May (1984). "Extraposition and Logical Form," *Linguistic Inquiry* 15: 1-32.

Halle, M. and J.R. Vergnaud (1987). *An essay on stress.* Cambridge: MIT Press.

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

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

Hornstein, N. (1995). *Logical Form: from GB to Minimalism.* Oxford: Blackwell.

Hornstein, N. (1998). "Adjunct control and parasitic gaps." In *University of Maryland Working Papers in Linguistics* vol. 6, ed. E. Murguia, A. Pires and L. Quintana. College Park: UMDWPL.

Johnson, K. (1997). "A review of *The Antisymmetry of syntax." Lingua* 102: 21-53.

Johnson, K. (2000). "How far will quantifiers go?" In *Step by Step: Essays on minimalist syntax in honor of Howard Lasnik,* ed. R. Martin et al. (2000). 187-210.Cambridge: MIT Press

Johnson, K. and S. Tomioka (1997). "Lowering and mid-size clauses." In *Reconstruction: Proceedings of the 1997 Tübingen Workshop,* ed. G. Katz, S.-S. Kim and H. Winhart. 185-206. Universität Stuttgart and Universität Tübingen.

Kayne, R. (1983). "Connectedness." *Linguistic Inquiry* 14: 223-249.

Kayne, R. (1995). *The Anti-symmetry of syntax.* Cambridge: MIT Press.

Kayne, R. (1998). "Overt vs. covert movement." *Syntax* 1: 128-191.

Karttunen, L. (1977). "Syntax and semantics of questions." *Linguistics and Philosophy* 1: 3-44.

Kearny, K. (1983). "Governing categories." Ms. University of Connecticut, Storrs.

Kennedy, C. (1997). "Antecedent-contained deletion and the syntax of quantification." *Linguistic Inquiry* 28: 662-688.

Kim, S. and J. Lyle (1995). "Parasitic gaps, multiple questions, and VP ellipsis." In *Proceedings of WCCFL 14,* ed. J. Camacho, L. Choueiri and M. Watanabe. Stanford: CSLI.

Kitagawa, Y. (1986). *Subjects in English and Japanese.* Doctoral dissertation, University of Massachusetts. Amherst: GLSA.

Koopman, H. and D. Sportiche (1991). "The position of subjects." In *The syntax of verb-initial languages,* ed. J. McCloskey. Elsevier.

Kuroda, Y. (1988). "Whether we agree or not." *Lingvisticae Investigationes* 12: 1-47.

Laka, I. (1992). *Negation in syntax: on the nature of functional categories and projections.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Lakoff, G. (1970). "Repartee," *Foundations of Language* 6: 389-422.

Larson, R. (1988a). "Light predicate raising," MIT Center for Cognitive Science: Lexicon Working Papers vol. 27. Cambridge, Mass.: MITWPL.

Larson, R. (1988b). "On the double object construction." *Linguistic Inquiry* 19: 335-391.

Larson, R. and R. May (1990). "Antecedent containment or vacuous movement: Reply to Baltin." *Linguistic Inquiry* 21: 103-122.

Lebeaux, D. (1988). *Language acquisition and the form of the grammar.* Doctoral dissertation, University of Massachusetts. Amherst: GLSA.

Lebeaux, D. (1990). "Relative clauses, licensing, and the nature of the derivation." In *Proceedings of NELS 20,* ed. J. Carter, R.-M. Déchaine, B. Phillip, and T. Sherer. 318-332. Amherst: GLSA.

Legate, J. (1999). "Verb phrase types and the notion of phase," ms. MIT.

Longobardi, G. (1984). "Connectedness, scope and c-command." *Linguistic Inquiry* 16: 163-192.

McCloskey, J. (forthcoming). "The morphosyntax of *wh*-extraction in Irish." To appear in *Journal of Linguistics.*

Maxfield, T. and D. McDaniel (1990). "What do children know without learning?" In *Papers in the acquisition of WH: Proceedings of the University of Massachusetts Roundtable,* ed. T. Maxfield and B. Plunkett. 101-123. Amherst: GLSA.

Merchant, J. (2000). "Economy, the copy theory and antecedent-contained deletion." *Linguistic Inquiry* 31: 566-575.

Müller, G. (1999). "Imperfect checking." *The Linguistic Review* 16: 359-404.

Munn, A. (1992). "A null operator analysis of ATB gaps." *The Linguistic Review* 9: 1-26.

Munn, A. (forthcoming). "Explaining parasitic gap restrictions." In *Parasitic gaps,* ed. P. Culicover and P. Postal. Cambridge: MIT Press.

Nissenbaum, J. (1998a). "Derived predicates and the interpretation of parasitic gaps." In *The Proceedings of WCCFL 17,* ed. S. Blake, E.S. Kim and K. Shahin. Stanford: CSLI.

Nissenbaum, J. (1998b). "Movement and derived predication: evidence from parasitic gaps." In *The Interpretive Tract:* MITWPL 25, ed. U. Sauerland & O. Percus. Cambridge, Mass: MITWPL.

Nissenbaum, J. (2000). "Covert movement and parasitic gaps." In *Proceedings of NELS 30,* ed. M. Hirotani. Amherst: GLSA.

Nunes, J. (1995). *The copy theory of movement and linearization of chains in the minimalist program.* Doctoral dissertation, University of Maryland at College Park.

Pesetsky, D. (1982). *Paths and categories.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Pesetsky, D. (1987). "*Wh*-in-situ: movement and unselective binding." In *The representation of (in)definiteness,* ed. E. Reuland and A. ter Meulen. 98-129. Cambridge: MIT Press.

Pesetsky, D. (1995). *Zero syntax.* Cambridge: MIT Press.

Pesetsky, D. (1998). "Some optimality principles of sentence pronunciation." In *Is the best good enough? Optimality and competition in syntax,* ed. P. Barbosa et al. Cambridge: MIT Press and MITWPL.

Pesetsky, D. (in press). *Phrasal movement and its kin.* Cambridge: MIT Press.

Phillips, C. (1996). *Order and structure.* Doctoral dissertation, MIT. Cambridge: MITWPL

Postal, P. (1995). "Parasitic and pseudoparasitic gaps", *Linguistic Inquiry* 25: 63-117.

Postal, P. (1998). *Three investigations of extraction.* Cambridge: MIT Press.

Reinhart, T. (1979). "A second COMP position." In *Theory of Markedness in Generative Grammar: Proceedings of the 1979 GLOW Conference,* ed. A. Belletti, L. Brandi and L. Rizzi. Scuola Normale Superiore di Pisa. 517-557.

Reinhart, T. (1983). *Anaphora and Semantic Interpretation.* London: Croom Helm.

Reinhart, T. (1991). "Non-quantificational LF." In *The Chomskian turn,* ed. A. Kasher. Oxford: Blackwell. 360-384.

Richards, N. (1997). *What moves where when in which language.* Doctoral Dissertation, MIT. Cambridge: MITWPL.

Richards, N. (1999). "Dependency formation and directionality of tree construction." In *Papers in morphology and syntax, cycle two.* MITWPL vol. 34. Cambridge: MITWPL.

Rizzi, L. (1997). "The fine structure of the left periphery." In *Elements of Grammar,* ed. L. Haegeman. Kluwer.

Rodman, R. (1976). "Scope phenomena, 'movement transformations,' and relative clauses." In *Montague Grammar,* ed. B. H. Partee. New York: Academic Press. 165-176.

Romero, M. (1997). "The correlation between scope reconstruction and connectivity effects." In *Proceedings of the sixteenth West Coast Conference in Formal Linguistics,* ed. E. Curtis, J. Lyle and G. Webster. Stanford: CSLI Publications. 351-366.

Ross, J. R. (1967). *Constraints on variables in syntax.* Doctoral dissertation, MIT. [Published as *Infinite Syntax!* Norwood, N.J.: Ablex. (1986).]

Rudin, C. (1985). *Aspects of Bulgarian syntax: complementizers and wh-constructions.* Columbus, Ohio: Slavica Publishers.

Rullmann, H. and S. Beck (1998). "Reconstruction and the interpretation of *which* phrases." In *Reconstruction: Proceedings of the 1997 Tübingen Workshop,* ed. G. Katz, S.-S. Kim and H. Winhart. Universität Stuttgart and Universität Tübingen. 223-256.

Ruys, E. (1993). *The Scope of indefinites.* Utrecht: OTS Dissertation Series.

Sag, I. (1976). *Deletion and Logical Form.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Sag, I. (1983). "On Parasitic Gaps." *Linguistics and Philosophy* 6: 35-45.

Sauerland, U. (1998). *The meaning of chains.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Speas, M. (1986). *Adjunction and projections in syntax.* Doctoral dissertation, MIT. Cambridge: MITWPL.

Steedman, M. (1997). *Surface structure and interpretation.* Cambridge: MIT Press.

Stowell, T. (1981). *Origins of phrase structure.* Doctoral dissertation, MIT.

Szabolcsi, A. (1986). "Comparative superlatives," MITWPL 8: 245- 265. Cambridge: MITWPL.

Taraldsen, K. T. (1981). "The theoretical interpretation of a class of marked extractions." In *Theory of Markedness in Generative Grammar: Proceedings of the 1979 GLOW Conference,* ed. A. Belletti, L. Brandi and L. Rizzi. Scuola Normale Superiore di Pisa.

Thompson, E. (1995). "The structure of tense and the syntax of temporal adverbs." In *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics,* ed. R. Aranovich, W. Byrne, S. Preuss and M. Senturia. Stanford: CSLI Publications.

Urribe-Etxebarria, M. (1995), "Levels of representation and negative polarity item licensing." In *Proceedings of the Fourteenth West Coast Conference on Formal Linguistics,* ed. J. Camacho, L. Choueiri, and M. Watanabe. Stanford: CSLI Publications. 571-586.

Vahedi, M. (1994). "Scrambling and word order in Modern Farsi." *Cahiers Linguistiques d'Ottawa* 22, 77-106.

Watanabe, A. (1992). "Subjacency and s-structure movement of *wh*-in-situ." *Journal of East Asian Linguistics* 1: 255-291.

THESIS PROCESSING SLIP

FIXED FIELD: ill. _____ name _____

index _____ biblio _____

► **COPIES:** (Archives) Aero Dewey Eng (Hum)

Lindgren Music Rotch Science

TITLE VARIES: ► ☐ _____

_____

_____

NAME VARIES: ► ☐ _____

_____

IMPRINT: (COPYRIGHT) _____

_____

► **COLLATION:** _____

_____

► **ADD: DEGREE:** _____ ► **DEPT.:** _____

SUPERVISORS: _____

_____

___ ___ _____

___ ___ _____

___ ___ _____

___ ___ _____

___ ___ _____

___ ___ _____

NOTES:

_____ cat'r. _____ date: _____

	page:
► **DEPT:** _____	► 516

► **YEAR:** _____ ► **DEGREE:** _____

► **NAME:** _____