

Phrasal Movement and its Kin

David Pesetsky

Dept. of Linguistics and Philosophy

Massachusetts Institute of Technology

pesetsk@mit.edu

First draft: August 1998

Revision: October 1999

ACKNOWLEDGMENTS

The investigations reported here are the result of three lucky events. The first occurred in 1986. I had recently done the work reported in Pesetsky (1987), and received in the mail a copy of Kiss (1986). Since I had argued at length that D-linked *wh*-phrases do not display Superiority effects. I was astonished by a paradigm reported by Kiss, which appears here as example (98). These facts remained stubbornly in my mind for the next decade as an unsolved puzzle. Kiss did not publish this paper in the form that I received — and, in fact, did not even recall discovering the crucial facts when she heard this work presented as a talk in 1998. But the facts are hers nonetheless.

The second lucky event occurred in Spring 1994. I had just taught my own Pesetsky (1987) to a graduate syntax class, but had not assigned a problem set on it. The next topic on the syllabus was Antecedent Contained Deletion, which I was teaching for the first time. I needed a problem set that covered both topics, and the idea popped into my head that perhaps Pesetsky (1987) might make a prediction about Antecedent Contained Deletion in *wh* phrases that should have moved overtly but didn't — a prediction which might serve as the topic of a single, unified problem set on the two topics. The result was the observation in (60) and (61). A conversation with Chris Kennedy (who had been thinking about this issue independently) reinforced my interest in this contrast.

Finally, in 1996 (a decade after I filed Kiss's facts away in my memory), I was lucky enough to hear a presentation of Beck (1996) and Beck and Kim (1996) by Sigrid Beck at an informal meeting of MIT's student-organized "LF Reading Group". Her observations immediately reminded me of Kiss's observations. The present paper grew out of my attempt to understand the connection between these two sets of observations — an effort which led me right back to the data from my problem set on Antecedent Contained Deletion.

My proposals have undergone several transformations, with the result that some whose comments are acknowledged in this work may not recognize it as the descendent of the one that they heard. I am

grateful to those with whom I discussed the work during its various stages, and to those who commented on public presentations. The earliest versions were presented in Spring 1997 in a seminar at MIT and in a University of Pittsburgh colloquium. A version more similar to the present paper, but couched in an Optimality-Theoretic setting (see footnote 57), was presented as a summer course at the 1997 LSA Summer Institute at Cornell. For helpful questions and discussion, I am especially grateful to Jonathan Bobaljik, Barbara Citko, Peter Cole, Molly Diesing, Gabriella Hermon, Satoshi Tomioka, Colin Wilson, and Susanne Wurmbrand. This version of things was also presented at the "Colloque de Syntaxe et Sémantique" held at Paris VII in October 1997.

Especially important was the opportunity to present the work once more at MIT during Fall 1997, at which time the role of feature movement fell into place. For crucial discussion during this period, I am grateful to Elena Anagnostopoulou, Sigrid Beck, Michael Brody, Noam Chomsky, Kai von Stechow, Danny Fox, Michel Degraff, Paul Hagstrom, Irene Heim, Sabine Iatridou, Roumyana Izvorski, Shigeru Miyagawa, Jonathan Nissenbaum, Norvin Richards and Uli Sauerland. This version of things was presented as a talk during the 1997-8 academic year at the University of Massachusetts at Amherst, ESCOL (at Yale), the University of Maryland, the University of São Paulo and Sophia University, and at the conferences "Formal Approaches to Slavic Linguistics" (University of Washington) and "Perfection in Syntax" (Collegium Budapest). More extended presentations at the University of Brasilia and at Kanda University (Kaihin Makuhari, Japan) during the Summer of 1998 played an important role in clarifying the ideas and teaching me the best ways to present them. The questions and comments of the audience in Brasilia, including Lucia Lobato, Cilene Rogrigues, Rozana Naves, Heloisa Salles and Clovis Santos, among others, are gratefully acknowledged. In addition, I am grateful to Nobuko Hasegawa, Kazuko Inoue, Yasu Ishii, Noriko Kawasaki, Hisatsugu Kitahara, Mari Takahashi and Keiko Yoshida, among many others, for challenging questions and helpful suggestions during the Kanda lectures. A final presentation at the University of Vienna in June 1999 helped with the discussion of German. Thanks as well to Roumyana Izvorski for information about Bulgarian, and to the German-speaking linguists at MIT for invaluable assistance with "Beck effects" and other matters.

The first draft of this manuscript was written in July and August 1998. Michael Brody, Kai von Stechow, Günther Grewendorf, Paul Hagstrom, Sabine Iatridou and Norvin Richards provided useful

comments on this draft. Special thanks to Shigeru Miyagawa for crucial encouragement and assistance as I wrote it. Three anonymous reviewers for MIT Press also provided helpful and thought-provoking comments. Their suggestions improved many parts of the manuscript. Almost all their suggestions have been taken into account here. To the fourth reviewer, who criticized the manuscript for its large number of footnotes (and for their length) I say: I have set no records for footnotes in linguistics! The footnotes attempt to acknowledge loose ends, unexplored alternatives, and interesting corollaries without interrupting the main lines of an already intricate investigation. If they distract, ignore.

Cambridge, Massachusetts

October, 1999

1. Introduction

This study is an attempt to discover the types of movement and movement-like relations that link positions in syntactic structure. I will argue that we can identify at least two kinds of movement-like relations: phrasal movement of the traditional kind, and the relation that Chomsky (1995, chapter 4) has called "feature movement."¹ My results differ sharply from Chomsky's, however, in supporting an additional distinction between overt and covert phrasal movement. If these conclusions are correct, then "feature movement" is not the proper reanalysis of covert phrasal movement, as Chomsky suggested, but is a phenomenon in its own right — a type of syntactic relation distinct from previously studied instances of movement. All in all, then, we will see evidence for three kinds of movement relations: overt phrasal movement, covert phrasal movement, and feature movement.

The focus of our study will be *wh*-questions. In particular, our classification of movement-like relations will allow us to understand the story behind *wh*-questions in which an otherwise inviolable property of movement — "Attract Closest" — appears to be violated. We will see that Attract Closest is actually not violated at all in the cases in question. The key to this conclusion will be a demonstration that more movement takes place in these configurations than we might have suspected, so that the "closest" element is really being attracted after all. Crucially, we will be able to reach this conclusion only once we accept the expanded repertoire of movement-like relations presented in this study.

The study concludes with a corollary investigation of a semantic restriction on *wh*-questions investigated by Beck (1996). This restriction appears in English in exactly those constructions that occupy us in the first sections of this study, but has a wider distribution in German — and an even wider distribution in Japanese and Korean. I will try to show that the distribution of Beck's effect in these three language-types follows straightforwardly from our hypothesis about the movement-like relations available to languages, in tandem with a typology of specifier requirements that also figures prominently in the first part of the study. In short, I will argue that our general view of movement helps us understand how the syntax of *wh*-questions differs across languages.

I begin with an introduction to the central issues of the study.

1.1. Phrasal movement: overt and covert

The term "movement" describes a situation in which a syntactic unit — for example, a phrase — appears to occupy more than one point in syntactic structure. Movement has been most easily detected by researchers when a word or phrase is pronounced in a position where we do not expect it to be pronounced (given an independently well-motivated syntax for argument-taking and modification). Often, when a word or phrase is pronounced in an "unexpected" position, it is pronounced *only* in the unexpected position.² It is not additionally pronounced in its expected ("trace") position. As was observed in the early 1970s, the pronunciation position in these cases typically c-commands the trace position. This constellation of syntactic and phonological properties constitutes what we can call *overt phrasal movement*.³

Overt phrasal movement is *movement* in that it obeys a characteristic set of command restrictions, as well as locality conditions governing the distance between the "expected" position and the position created by movement. It is *phrasal* in that the moved unit is a word or group of words. (Reconstruction effects reveal the presence of the moved phrase in both trace position and targeted position.) It is *overt* in that it affects the phonology; the sentence sounds different with the movement operation than it would have sounded without it. Many of the best-studied instances of movement belong to the category of overt phrasal movement. Of special interest to us will be overt *wh*-phrase movement of the sort familiar from many languages, including English:

(1) Which book did Mary give ___ to Sue?

As linguists, we are lucky that overt phrasal movement exists, or else movement itself might not have been discovered. Nonetheless, from a broader perspective, the pronunciation pattern of overt phrasal movement seems rather arbitrary. Why should pronunciation target the moved position and not the trace

position? Why not the other way around, for example? In fact, a large body of research has provided evidence that other pronunciation patterns do exist. The proposals of Chomsky (1976), May (1977; 1984) and Huang (1981; 1982) provided central arguments for the existence of alternative pronunciation patterns for movement. These researchers argued that sometimes a moved element is pronounced in a trace position, rather than in its final (highest) position. This type of movement is traditionally called *covert*. We will call it *covert phrasal movement*. Covert phrasal movement can be shown to be *movement* insofar as it creates a link between positions which obeys command restrictions and islands in a manner familiar from overt phrasal movement. Covert phrasal movement is *covert* in that it does not affect the segmental phonology. The moved words affected by covert phrasal movement are pronounced just as if no movement had taken place. Covert phrasal movement is *phrasal* in that entire words and word groups are copied from the trace position to the new position.

A particularly good argument for covert phrasal movement comes from "antecedent-contained" anaphora of the sort familiar from research on Antecedent-Contained Deletion (ACD) constructions like (2) (Bouton 1970; Sag 1976; May 1985; Larson and May 1990):

(2) **ACD**

Mary [_{VP} invited [_{DP} everyone that I did [_{VP}]]]

The argument runs as follows. The example in (2) most naturally has the interpretation in (3) (where "t" is the trace of relativization within DP):

(3) **Interpretation of ellipsis in (2)**

Mary [_{VP} invited [_{DP} everyone that I [_{VP} invited t]]]

It is a fact about VP-ellipsis (Hankamer and Sag 1976) that an elided VP requires a linguistic expression as its antecedent. In contrast to a pronoun, which may take its reference from a contextually salient, but unmentioned individual, an elided VP can only take a pronounced VP as its antecedent. Thus, while I may say "Thank goodness *he's* left" as a response to the departure of an salient but unmentioned individual, I cannot say "Thank goodness *I* didn't [VP]]" as a response to someone's tripping over a wire.

Consequently, the fact that the elided VP in (3) is understood as the expression *invited t* tells us that the linguistic context must contain an antecedent of the form "*invited t*". If the analysis of (2) does not involve covert movement, then we cannot understand the availability of interpretation (3). Without covert movement, (2) contains no VP of the form *invited t* — only a VP of the form *invited everyone that I did*. If, on the other hand, the phrase *everyone that I did* undergoes covert phrasal movement to a VP-external position, then it will leave behind a VP of the form *invited t*, supplying the appropriate antecedent for the elided VP:

(4) **Example (2) after covert phrasal movement**

[_{DP} everyone that I [_{VP} invited *t*]] [Mary [_{VP} invited *t*]]

Support for this proposal comes from Larson and May's (1990) discussion of configurations in which both a higher and a lower VP can serve as antecedent for an instance of ACD (a discussion anticipated by Sag 1976, 72-74). Larson and May noted that the higher VP can be chosen only if the phrase containing the ACD is interpreted with scope wider than the higher VP. The phenomenon is demonstrated in (5):⁴

(5) **Scope/ellipsis correlations**

John refused to visit [every city Mary did [VP]]

Interpretations:

a. John refused to visit [every city Mary visited *t*]

[ambiguous: compatible with narrow or wide scope of *every city*]

b. John refused to visit [every city *Mary refused to visit t*]

[unambiguous:incompatible with narrow scope of *every city*]

The story is more complex than this, of course, if traces are simply the originals of which moved phrases are copies (Chomsky 1995, chapter 3). In the process of semantic interpretation, descriptive material in this trace position must be deleted, so that the trace not only functions as a variable, but also helps provide an appropriate antecedent VP for the ellipsis site. Fox (1995), building on a discussion by Fiengo and May (1994), shows that this process (which, he argues, is motivated by ACD interpretation) interacts with Principle C of the Binding Theory just as expected. Covert movement in normal circumstances is insufficient to bleed Principle C (presumably because nothing motivates deletion of the trace-internal material that generates the Principle C violation):

(6) **No bleeding of BT(C) without ACD**

*I [sent him_i [every letter that John_j expected I would write *t*]

Covert movement with ACD, however, requires deletion of the trace-internal material, which in turn removes the Principle C violation:

(7) **Bleeding of BT(C) with ACD**

I [sent him_i [every letter that John_j expected I would [VP]] —>

[every letter that John_j expected I would [VP send him_i *t*]] I [sent him_i *t*].

Important complications in this picture are discussed by Fox (1995; 1998) and by Sauerland (1998a), and will be glossed over here. Some of these complications concern the landing site of covert movement in cases like (2) and (7) — that is, whether the relevant DP moves to an IP-initial position, as shown in (4) and (7), or to some other (perhaps lower) VP-external position, as argued by Fox (1995; 1998). The only point important to us is the fundamental conclusion that the interpretation of ACD requires phonologically vacuous (i.e. covert) movement of the phrase that contains the deletion site.

Covert phrasal movement has often been thought of as a language-specific or construction-specific variant of overt phrasal movement. It is in this light, for example, that many researchers proposed covert phrasal movement of *wh*-phrases to C as a covert version of the overt phrasal *wh*-movement displayed in (1) (for example Huang 1981, 1982; and Aoun, Hornstein and Sportiche 1981). Covert *wh*-movement has frequently been proposed for languages like Chinese and Japanese (where most *wh*-phrases are pronounced in situ) as well for *wh*-in-situ in multiple questions in languages like English. One argument for covert *wh*-movement in English multiple questions can be constructed from the fact that both overt and (putative) covert *wh*-movement obey the same command condition. The trace position must be c-commanded by the C to which the *wh* moves:

(8) **Command condition on overt and covert *wh*-movement**

a. [Give a book to John] I can guess [who C will ___]!

[cf. *I can guess who will give a book to John!*]

b. *[Give a book to t_i] I can guess [who $_i$ C Mary will ___]!

[cf. *I can guess who Mary will give a book to!*]

c. *[Give a book to whom] I can guess [who C will ___]!

[cf. *I can guess who will give a book to whom!*]

The ungrammaticality of both (8b) and (8c) probably reflects a command requirement on movement, (rather than, for example, a semantic condition on variable binding), since no similar condition affects pronouns when they function as bound variables — as (9) demonstrates. The semantics, at least, can apparently inspect a fronted VP in its (reconstructed) base position:⁵

(9) **The command condition is not semantic.**

[Give a book to his $_i$ mother] I can guess who will ___!

If our conclusions are correct, the LF representation of *Who gave what to whom?* may look something like (10), in which *wh*-movement has affected all three *wh*-elements. Overstriking indicates the pronunciation pattern:

(10) **English multiple questions: LF and pronunciation**

Who gave what to whom?

Analysis: [*Who ~~what~~ ~~whom~~ [~~who~~ gave what to whom]*].⁶

As was often noted in the 1980s (Lasnik and Saito 1984; Rudin 1985; Pesetsky 1987), this hypothesis about English multiple questions is supported by the behavior of multiple questions in Slavic languages. (I focus here on Bulgarian.) In these languages, the covert instances of *wh*-movement posited for English are overt. That is, if (10) represents a correct analysis for English multiple questions, it differs from its Bulgarian counterpart only in how its *wh* chains are pronounced:

(11) **Bulgarian multiple questions: LF and pronunciation**

Koj kakvo na kogo dade?

Who what to whom gave?

'Who gave what to whom?'

Analysis: [*Koj kakvo na kogo [~~koj~~ dade ~~kakvo na kogo~~]*]

Of course, if covert *wh*-movement in English truly mirrors overt *wh*-movement, it should be demonstrably *phrasal*. That is, we should have evidence that covert *wh*-movement copies word groups similar to those copied in overt movement. ACD provides a test for this prediction. English *wh*-phrases pronounced in situ support ACD. This fact supports the hypothesis that *wh*-phrases may undergo covert phrasal movement, though it leaves open for now the possibility that the covert phrasal movement which resolves ACD is not *wh*-movement (as argued by Hornstein 1994, 1995, and by Lasnik 1993, among others). I return to this issue in section 0 below, where I argue that the movement that what resolves ACD is indeed *wh*-movement:

(12) **ACD licensed in *wh*-in-situ**

- a. Which girl invited [which student that John did [VP]]?
- b. I need to know who can speak [which languages that Ken Hale can [VP]]?
- c. Which spymaster suspected which spy that Angleton did [VP])?

(Fiengo and May 1994, 242)

Let us now consider the mechanics of covert phrasal movement in more detail. Traditionally, the pronunciation difference between overt and covert phrasal movement has been viewed as a consequence of the timing of movement within a derivational model. In the model assumed in many studies — the so-called "(inverted) Y-model" of Chomsky and Lasnik (1977) — the pronunciation of the chains of positions linked by movement is governed by a simple principle of Phonological Spellout:

(13) **Phonological Spellout in the "Y-model":**

Pronounce only the highest position in a movement chain.

On this view, covert movement is simply movement that takes place after Spellout . Movement after Spellout thus creates a situation in which the highest position at LF is higher than the highest position at the time of Spellout. Phonological processing and covert movement take place on separate derivational tracks (the arms of the inverted "Y"). Though this view is common, few (if any) sustainable arguments seem to support the Y-model's segregation of overt and covert movement within the syntax. Indeed, the segregation of overt and covert movement into separate blocks of structure-building operations creates complications for simple views of structure-building like that advanced by Chomsky (1995, chapter 3-4), in which the phrase structure of an expression simply *is* its derivational history. Covert movement as analyzed within the Y-model would be fairly unique in "altering the historical record" — by tinkering with the internal structure of the derivation.

In Pesetsky (1997; 1998), I suggested an alternative view, which places the burden of accounting for the covert/overt distinction on the phonology (Similar ideas have also been developed by Bobaljik 1995, who calls this view "Single Output Syntax", and by Groat and O'Neil 1996. A precursor was Brody 1995, circulated in 1992.) In this view, the syntax is simpler than in the Y-model; there is just one movement component. The trade-off comes in the phonological component, where the simple principle of Phonological Spellout in (13) is replaced by phonological principles of chain pronunciation that regulate the pronunciation of moved elements in a more complex manner. These principles determine, for example, whether the head of a chain is the only position pronounced (overt phrasal movement) or whether a trace position will be pronounced instead (covert phrasal movement). An indirect but telling argument for this point of view was the observation that the dichotomy "overt vs. covert" inherent in the "Y-model" is too crude, since there are other pronunciation patterns for chains. Sometimes, no position within a chain is pronounced, and sometimes more than one chain position is pronounced (e.g. in resumptive pronoun constructions).

In this study, very little will depend on the choice between the Y-model and Single-Output Syntax.⁷ Nonetheless, the overall picture will be significantly easier to discuss within the architecturally simpler Single-Output model. I will therefore assume that there is only one cycle of syntactic movement, and will describe pronunciation distinctions like the distinction between overt and covert phrasal movement as essentially phonological. Against this backdrop, I will argue that (14) and (15) give the proper characterization of the difference between English and Bulgarian pronunciation patterns in multiple questions. Justification of the reference to the "first instance" and "secondary instances" of *wh*-movement will be offered shortly.

(14) **Pronunciation rule (English):**

- a. The first instance of *wh*-phrase movement to C is *overt* in that *wh* is pronounced in its new position, and unpronounced in its trace positions..
- b. Secondary instances of *wh*-phrase movement to C are *covert*, in that *wh* is pronounced in its trace position, and is unpronounced in its new position.

(15) **Pronunciation rule (Bulgarian):**

All *wh*-phrase movement to C is *overt*, in that *wh* is pronounced in its new position, and unpronounced in its trace positions.

1.2.Feature movement

In his first paper developing the "Minimalist Program", Chomsky (1995, chapter 3) suggested that covert phrasal movement is the default style of movement. This idea was implemented by means of a principle called *Procrastinate*, which required movement to be covert. This principle, in turn, was argued to be overruled whenever movement is motivated by a "strong" feature. A strong feature, in this system, was a feature that needed to be "checked off" by overt movement, in order to avoid violating the principle of Full Interpretation at PF. Though this proposal provoked much interesting research and discussion, it had a certain arbitrariness about it, in that it was not obvious (at least at the conceptual level) why language should incorporate a timing principle like *Procrastinate*. Why not the opposite timing principle (e.g. the Earliness Principle of Pesetsky 1989, as incorporated, for example, in Brody's 1995 "radically minimalist" model), or no timing principle at all?

In later work, Chomsky (1995, chapter 4) offered an alternative view that promised to eliminate the issue. He suggested that it had been a mistake to assume that covert movement is ever "phrasal". Chomsky's 1995 proposal starts with the idea that movement is a "repair strategy" by which an uninterpretable feature F on a head K is deleted in response to movement to K of another instance of F

(typically an interpretable instance of F.. Failure to repair a structure that contains an uninterpretable feature renders the derivation non-convergent. Movement for any other purpose is banned. Thus, for example, an uninterpretable *wh* feature on C might require movement of the corresponding *wh* feature from a *wh*-phrase elsewhere in the structure, but there could be no "gratuitous" movement of this feature in other circumstances. Likewise, an uninterpretable person-feature on T requires movement of a person-feature from some DP internal to the structure containing T (with gratuitous instances of this movement strategy prohibited). On this view, movement at its simplest should copy just the features necessary to ensure convergence (Chomsky 1995, 262). Copying of anything more than features is unexpected — especially the phonological features and dependent constituents copied in overt phrasal movement.

In this system, phrasal movement is, in a sense, the surprise. Phrasal movement is a situation in which a grammatical feature that must be moved cannot be separated from the syntactic expression in which it occurs. In the case of overt DP movement to subject position (SPEC of TP), for example, it is supposed that the D feature cannot be copied apart from the remainder of the expression that it labels. Likewise, overt *wh*-phrase movement is attributed to the inability of the phonological system to pronounce the *wh*-feature and the remainder of its phrase in separate places. Chomsky suggests that when F moves, "F carries along just enough material for convergence". Movement of more than just the relevant feature occurs only if "generalized pied piping" is necessary in order for the derivation to converge at LF or PF. Chomsky further speculates that only PF considerations force pied piping of this sort.⁸

Chomsky's 1995 proposal leaves us with two types of movement. Movement that involves copying of a phrase is motivated by PF considerations, and must perforce be "overt" movement. Movement that copies only the grammatical features that motivate the movement must perforce be "covert" movement. On this view, then, feature movement is simply the proper analysis of covert movement — "covert" in the sense that it has no effect on the phonology. The postulation of covert *phrasal* movement, from this perspective, was simply an error. Notice that this view, if correct, offers a third slant on the overt/covert contrast — distinct both from the "Y-model" tradition and from the "single-output" syntax literature. The proposal that covert movement is feature movement removes any need for either syntactic timing or phonological principles to distinguish the two movement types.⁹ Covert and overt movement may

participate in a single syntactic cycle, with the phonological distinctions arising simply from the differences between the material copied by the movement operation. Chomsky (1995, chapter 4) did not develop this consequence of his new view of movement, but it follows rather straightforwardly, nonetheless. In more recent work, Chomsky (1998) takes exactly this step.

Chomsky's (1995, chapter 4) proposals concerning movement are consonant with some of the arguments for covert movement, but not with all of them. It is, of course, consonant with any test that cares only about the relation between a source and a target — for example, the command condition discussed in connection with (8). The new view runs afoul, however, of evidence for covert movement that identifies the moving element as a word or phrase. As we saw, evidence from ACD has just this property. ACD resolution *requires* movement of an entire phrase out of categories that contain it (a conclusion supported by the Binding Theory evidence of Fiengo and May (1994) as explained by Fox 1995¹⁰).

Consequently, it does not appear correct to simply reanalyze covert phrasal movement as feature movement. Covert phrasal movement exists. I suspect that this is not a step backwards. Although Chomsky's 1995 proposal provides a rationale for his earlier assumption that covert movement is the default type of movement, it is not clear that the earlier assumption was correct. I know of no empirical evidence bearing on the matter, nor is it obvious how the conceptual chips fall. Indeed, alongside the plausible-sounding idea that the most natural style of movement would copy just the features that are necessary in order to ensure convergence, one might lay a quite different, equally plausible idea: that movement copies the *largest constituent* that bears the relevant feature. This idea would make phrasal movement (rather than feature movement) the default. Each of these ideas is natural. Let us view movement as a process by which a head H in search of a feature F scans down the tree in order to identify a constituent that bears F and copies it. According to Chomsky's 1999 idea, only the feature itself is copied; according to the alternative, the identified constituent is copied. I am hard-pressed to find even an Occam's Razor argument that favors one idea over the other.¹¹

On the other hand, whatever one's views of the actual proposals, one can distill an important question from the discussion. Granted that movement, both overt and covert, does sometimes copy phrases,

are there movement operations that simply establish a relationship between different expressions bearing a particular *feature*? In this study, I will suggest that the answer is yes. If so, then Chomsky's (1995, chapter 4) discussion of feature movement uncovered something real — but misidentified the discovery. The phenomenon called "feature movement" by Chomsky is not an alternative analysis of covert phrasal movement, as he thought, but is a distinct syntactic operation in its own right.¹²

To see in a nutshell what I have in mind, compare the behavior of *wh*-in-situ with the behavior of the "associate" in the English *there*-construction. As is well-known, the verb in the English existential *there*-construction generally agrees with a post-verbal "associate" DP, which is usually required to be indefinite:

- (16) a. There is a book on the table.
b. There are some books on the table.
- (17) a. There is likely to be a book on the table.
b. There are likely to be some books on the table.

Apparently, the Merger of *there* as a specifier of TP satisfies the "Extended Projection Principle" requirements of T (its need for a phrasal specifier), but does not satisfy the requirement that T "check off" (i.e. delete) its number features (and possibly others). The word *there*, Chomsky suggests (p.273), does not bear these features. Consequently, the features must move to T from somewhere inside TP. The associate DP furnishes the necessary features:

- (18) a. There [F_i -is] [F_i -a book] on the table.
b. There [F_i -are] [F_i -some books] on the table.
- (19) a. There [F_i -is] likely to be [F_i -a book] on the table.
b. There [F_i -are] likely to be [F_i -some books] on the table.

There is thus a movement-like relation between the associate and T. This relation is "covert", in that neither the associate nor T seems to show any phonological effect of the relationship (besides the agreement relation itself). Nonetheless, the relation is real. For one thing, c-command must be maintained between T and the associate:

(20) **C-command condition on feature-movement**

a. He said there [F_i -were] likely to be [F_i -several books] on the table.

b. *...and likely to be [F_i -several books] on the table there are.

Furthermore, the associate must be the closest DP to T:

(21) **The ban on superraising with overt phrasal movement**

*Several books_{*j*} are desirable [for it to be *t_j* on the table]

The relation between the associate and T in the *there*-construction is blocked in exactly the same circumstances:

(22) **The ban on superraising with feature movement**

*There are desirable for it to be several semanticists at the party.

So far, the relation between the associate and T is revealed by evidence familiar from arguments for covert movement (movement without a phonological effect). But is the relation covert *phrasal* movement of the associate to T? ACD provides us with a means of detecting phrasal movement, and strongly suggests that the answer to our question is "no".

To see this, we must attach a relative clause containing VP ellipsis to an associate of *there*, and test for the possibility of using a higher VP as an antecedent for the elided VP. For reasons that are unclear to me, a relative clause attached to an associate in the *there* construction favors an associate containing a strong

quantifier like *every*, rather than the typical indefinite — but some weak associates like *no one* are also natural:

(23) **Relative clauses compatible with an associate**

- a. There will be [everyone that there should [_{VP} be *t*]] at the party.
- b. There will be [almost no one that there should [_{VP} be *t*]] at the party.

If we now elide the VP in (23), we have our test for ACD. First, however, we should ensure that VP-ellipsis is compatible in principle with the *there*-construction. Example (24) shows that it is:

(24) **VP ellipsis acceptable in the *there*-associate construction**

Will there be phonologists at the party? Well, there should [_{VP}].

The examples in (25) display the crucial contrast with (24):

(25) **ACD impossible in a relative clause modifying an associate.**¹³

- a. *There will be [everyone that there should [_{VP}]] at the party.
- b. *There will be [almost no one that there should [_{VP}]] at the party.

Since (24) shows that there is nothing wrong with the ellipsis *per se*, the unacceptability of ACD in (25) shows us that — whatever is transpiring between the associate and T in a *there*-construction — it is not phrasal movement.¹⁴ Instead, it appears to be movement of something (obligatorily) smaller than the associate phrase. Feature movement fits the bill.¹⁵ This conclusion accords with Chomsky's (pp. 272ff.) hypothesis that long-distance agreement in the *there*-construction is a consequence of feature movement.

In the case of the *there*-construction, the feature movement hypothesis revises an earlier set of proposals, also by Chomsky, in which the movement properties of the associate-T relation were attributed to covert phrasal movement. Though ACD did not form one of Chomsky's arguments for this revision, it provides a strong argument in its favor. If we accept this conclusion, however, we need an explanation for the contrast between the "associate in situ" in the *there*-construction and *wh*-in-situ in multiple questions. In each case, we found evidence of a relation between a phrase pronounced in situ and a higher position. But the two constructions differ in their behavior under a test for specifically *phrasal* covert movement. While covert movement of the associate in the *there*-construction looks like feature movement, covert movement of *wh*-in-situ looks like phrasal movement. It is this pattern of evidence that suggests the existence of more than one type of movement-like relationship between syntactic positions.

I will take these conclusions as my point of departure, devoting considerable attention to distinctions between phrasal movement and "feature movement" in the sense of Chomsky (1995, chapter 4). Later in this study, however, we will also examine some possible variants this picture — though I will leave the choice among alternatives open. In particular, I will entertain the possibility that "feature movement" is actually a subcomponent of phrasal movement. I will attribute the similarities between the two operations to that fact, rather than to the existence of two substantially different varieties of movement. Note, for example, that there is no clear evidence that features of the associate in the *there*-construction actually "move" to a position near T. Instead, the evidence points to some sort of communication between the features of the associate and the features of T. For now, however, I will stick to more familiar concepts like "movement", and will present my arguments as evidence for the coexistence of phrasal and feature movement in grammar.

We are now in a position to sketch the goals of the study in some detail. If our discussion so far is on the right track, we might expect to find "minimal triplets" in which a particular head is seen to establish a relationship with a remote feature in each of the three ways we have seen—the choice depending on other properties of the grammar. That is, we might expect to find movement to a particular head showing up sometimes as overt phrasal movement, sometimes as covert phrasal movement, and sometimes as feature movement. The existence of such a "minimal triplet" is not logically entailed by our hypotheses.

Nonetheless, if we can identify such a triplet, we will significantly improve our ability to investigate the coexistence of phrasal movement with feature movement in the grammar, since we will have the opportunity to examine the differences among these syntactic operations under well-controlled conditions.

The remainder of this study presents and investigates a triplet of exactly this sort. In particular, I will argue that interrogative *wh*-constructions involve relations of all three types. We have already seen examples of overt and covert phrasal *wh*-movement to C in English. I will suggest that under certain circumstances, we can detect another kind of operation that relates *wh*-phrases to an interrogative C: feature movement (or something very much like it).

I introduced Bulgarian into our discussion of English multiple questions with malice aforethought, since Bulgarian provides some of the foundations of our argument. After presenting some evidence in favor of covert phrasal movement for *wh*-in-situ, I noted (in the footsteps of Lasnik and Saito 1984) that Bulgarian, on this view, simply shows overt phrasal movement where English shows its covert counterpart. The difference between the two languages was summarized in (14)-(15). In the next section, I present certain additional facts about Bulgarian multiple questions, for which partial explanations will be offered (borrowed from Richards 1997). In the following sections, I will show that a seemingly distinct set of puzzles concerning English multiple questions actually represents exactly the same phenomena we have just observed in Bulgarian. If this is true, we can claim that the pronunciation distinctions in (14) and (15) constitute the *only* relevant difference between multiple questions in the two languages. We only reach this conclusion, however, if we assume that multiple questions in English (and Bulgarian) in certain cases involves pure feature movement of *wh* — even though in other cases it involves overt and covert phrasal movement triggered by the same feature. In the final sections of the study, we consider some syntactic and semantic consequences of this hypothesis, which helps establish a cross-linguistic typology of *wh*-movement.

1.3. Superiority effects

The English puzzles involve the observation known as the "Superiority Effect". The Superiority Effect arises in multiple questions when more than one *wh*-phrase is relevant to the answering patterns of a given question. In such cases the syntax needs to decide on a pattern of *wh*-movement within that question. The Superiority Effect is a restriction on this decision. In English, where only one *wh*-phrase moves overtly, the Superiority Effect is the contrast observed in (26)-(27):

(26) **Superiority Effect**

- a. Who __ bought what?
- b. *What did who buy __ ?

(27) a. Who did you persuade __ to read what?

- b. ??What did you persuade whom to read __?

If we assume that the first application of *wh*-movement in these examples is the overt phrasal instance of *wh*-movement, Kuno and Robinson's (1972) description of the phenomenon describes it accurately:

(28) A *wh* word cannot be preposed crossing over another *wh*.

Chomsky (1973) suggested that (28) is a special case of a more general phenomenon, which he termed the "Superiority Condition:

(29) **Superiority Condition (Chomsky 1973)**

No rule can involve X, Y in the structure

...X... [... Z... —WYV ...]

where the rule applies ambiguously to Z and Y and Z is superior to [m-commands] Y.

The "Attract Closest" condition of Chomsky (1995, chapter 4) is, in essence, a restatement of the Superiority Condition for movement, where movement is viewed as triggered by particular features of a "target" head K. I will abbreviate "Attract Closest" as AC:

(30) **Attract Closest ("AC"; Chomsky 1995, chapter 4, pp. 280, 296)**

_____ can raise to target K only if there is no legitimate operation Move _____ targeting K, where _____ is closer to K.

When one presents the examples in (26)-(27) to speakers in a manner that is fair to the data (maintaining focal stress on *wh*-in-situ in both cases, and presenting the examples without special context), the contrasts are clear and stable. Nonetheless, there are minimal changes that one can make in these examples that produce apparent exceptions to the Superiority Effect. These facts will be crucial to our investigation.

1.3.1 D-linking. One minimal change of this sort involves what Pesetsky (1987) called "D-linking". When a *wh*-question asks for answers in which the individuals that replace the *wh*-phrases are drawn from a set that is presumed to be salient both to speaker and hearer, the multiple question can appear to violate AC. This possibility typically arises when the answers to the question are supposed to be drawn from a set of individuals previously introduced into the discourse, or when the set forms part of the "common ground" shared by speaker and hearer. *Wh*-phrases with *which* favor this type of interpretation:

(31) **Superiority effect disappears with D-linking.**

- a. Which person ___ bought which book?
- b. Which book did which person buy ___?

- (32) a. Which person did John talk to ___ about which topic?
b. Which topic did John talk to which person about ___?

The semantics of D-linked *wh* phrases closely tracks the semantics of the definite article *the*. Context sets previously mentioned in the discourse qualify a phrase as D-linked, but so do sets that are merely salient

(e.g. *which book*, in a context where speaker and hearer both know that reference is being made to a reading list for a course), and sets whose salience is culturally determined (e.g. *what day of the week*, *which sign of the zodiac*). A reliable rule of thumb is the following: if a *wh*-word in a multiple question can be felicitously paraphrased with an expression of the form *which of the X*, it can cause the Superiority Effect to disappear. The reason for this link between semantics and syntax is obscure, and will remain obscure even at the end of this study. On the other hand, the syntax of the problem is also obscure: what is the structure of D-linked questions¹⁶ in which the Superiority Effect appears to be obviated? Are they exceptions to Attract Closest, or is something else going on? By the end of this study, I hope to have answered *these* questions quite fully.

1.3.2 More than two *wh*-phrases. This factor is less studied,¹⁷ but equally striking. The Superiority Effect is limited to questions with two *wh*-phrases. It disappears in multiple questions with more than two *wh*-phrases.¹⁸

(33) **Superiority effect disappears with more than 2 *wh*-phrases.**

- a. *What did who give __ to Mary? [Sensation of Superiority effect]
- b. What did who give __ to whom? [No sensation of Superiority effect]

We will discuss this effect in greater detail later. For now, we simply wish to note the important syntactic question that this contrast raises. They are the same as the ones posed in the previous section: What is the structure of questions with more than two *wh*-phrases, in which the Superiority Effect appears to be obviated? Are they exceptions to Attract Closest, or is something else going on?

1.3.3 German. Another minimal change one can make in a multiple question that obviates the Superiority Effect is translation into German. In at least the simplest cases, German appears to lack the Superiority contrast entirely:

(34) **Superiority effect disappears in German.**

- a. Ich weiß nicht, wer was gesehen hat. 'I don't know who has seen what.'
- b. Ich weiß nicht, was wer gesehen hat. 'I don't know what who has seen.'

The correct interpretation of this fact is a matter of some current controversy, which we will turn to towards the end of this study. For now, it is sufficient to observe the problem, and to ask the same questions we have asked about the previous two sets of apparent counterexamples.

These are the problems that will occupy us for the remainder of this study. In the next section, we present a fuller picture of Bulgarian multiple questions, as necessary background to our discussion of apparent exceptions to the Superiority Effect.

2. Two Observations about Bulgarian

2.1. A Complementizer That Requires Multiple Specifiers

I have already noted that Bulgarian (unlike English) shows multiple instances of overt phrasal *wh*-movement to the left periphery of CP. I have not said anything so far about the obligatoriness or optionality of multiple overt movement. The obligatoriness of multiple overt movement partly depends on D-linking, in ways to which we return later. For now, let us consider only multiple questions without a D-linked interpretation. In these cases, the facts are somewhat subtle.

Consider a question containing three *wh*-phrases. According to consultants' judgments, the most natural multiple question using three *wh*-phrases shows all three phrases overtly fronted, as in (35a) below. Nonetheless, a construction in which only two of the *wh*-phrases are fronted is fairly acceptable. By contrast, a multiple question in which only one *wh*-phrase is overtly fronted is strongly unacceptable (in the non-D-linked usage being considered). There is thus a three-way contrast. The diacritics ("**" vs. "??" vs. "ok") are those supplied by one of my consultants, Roumyana Izvorski, but the general description of the judgments is one that contrasts perfection with unnaturalness with total unacceptability: The break between (35b-c) and (35d) is claimed to be sharper than that between (35a) and (35b-c).¹⁹

(35) **Bulgarian: all vs. 2 vs. 1 *wh*-phrase moves.**

- a. Koj kakvo na kogo dade? [all *wh*-phrases move]
 who what to whom gave
 'Who gave what to whom?'
- b. ??Koj na kogo dade kakvo? [2 *wh*-phrases move]
- c. ??Koj kakvo dade na kogo? [2 *wh*-phrases move]
- d. **Koj dade kakvo na kogo? [only 1 *wh*-phrase moves]

(R. Izvorski, I. Derzhanski p.c.)

Since there seems to be no reason to assume that the Bulgarian *wh*-phrases in a multiple question are related to distinct heads (Rudin 1988), we may safely assume that these phrases all bear the specifier relation to the same interrogative head, in the manner posited by Koizumi (1995) and Ura (1996). Let us call this interrogative head the "*multi-specifier complementizer*", which I will abbreviate as C_{m-spec} — to indicate that its "specifier potential" requires it to take more than one specifier. The ability of C_{m-spec} to attract more than one *wh*-phrase indicates that the feature of C_{m-spec} responsible for *wh*-movement (uninterpretable though it may be) does not need to delete and does not exhaust its ability to motivate *wh*-movement until its maximal projection has been fully built. We are thus free to propose that (36) is true, and that (36) and (37) are jointly responsible for the observed three-way split in judgments on Bulgarian multiple questions:²⁰

(36) **Specifier potential of C_{m-spec}**

C_{m-spec} requires more than one *wh*-specifier.

(37) **General preference**

All *wh*-phrases in a multiple question move.

I will assume that C_{m-spec} differs both syntactically and semantically from the standard interrogative complementizer C_{I-spec} , which takes a single *wh*-specifier. A multiple question in English or Bulgarian *must* be introduced by C_{m-spec} (and a non-multiple question must be introduced by C_{I-spec}). Consequently, example (35d) cannot slip through the net of constraints discussed in this section by using the complementizer reserved for non-multiple questions. I discuss the typology of complementizers in more detail in the final sections of this study (section §5.3).²¹

Besides the Multiple Specifier requirement of C_{m-spec} , there must be another condition with the effect of requiring at least one instance of phrasal *wh*-movement in an embedded question, so as to rule out (39a) with *wh*-feature movement from *what*. The specifier properties of C_{I-spec} ensure the contrast in (39) and its Bulgarian counterparts. The pronunciation principles guarantee that the one specifier required by C_{I-spec} will be overt:

(38) **Specifier potential of C_{I-spec}**

C_{I-spec} requires one *wh*-specifier.

(39) **C_{I-spec} in English**

- a. *Mary needs to know [Bill saw what].
- b. Mary needs to know [what Bill saw].

There is an alternative view of the Bulgarian data that we should consider. On this view, the only factor influencing judgments is (37), with sentences judged gradiently in accordance with how many *wh*-phrases violate (37). To decide between these two proposals on the basis of Bulgarian data, we must examine the behavior of questions with four (or more) *wh*-phrases. The two statements in (36) predict a three-way contrast in such a case. Moving all four *wh*-phrases would be perfect; moving two or three of the four *wh*-phrases would be indistinguishably unnatural to a mild degree; and moving only one *wh*-phrase would be totally unacceptable. To the extent that such a paradigm can be constructed without running afoul of other complications, the facts seem to support (36). For example, in (40), a ditransitive verb and an

instrumental phrase furnish the four *wh*-phrases. According to the judgments of Roumyana Izvorski (p.c.), there is a major acceptability break between (40d) and (40a-c).²²

(40) **4 *wh*-phrases: monoclausal examples**

- a. Koj na kogo kakvo s kakvo napisa? [all *wh*-phrases move]
who to whom what with what wrote
'Who wrote what to whom with what?'
- b. ?Koj na kogo kakvo napisa s kakvo? [3 out of 4 *wh*-phrases move]
- c. ???Koj na kogo napisa kakvo s kakvo? [2 out of 4 *wh*-phrases move]
- d. **Koj napisa kakvo na kogo s kakvo? [1 out of 4 *wh*-phrases move]

Similar tests can be performed with bi-clausal structures. Once again, there is a major acceptability break between (41d) and (41a-c).

(41) **4 *wh*-phrases: biclausal examples**

a. Koj kogo na kogo kakvo ubedi da dade?

who whom to whom what convinced to give

'Who convinced whom to give what to whom?'

b. ?Koj kogo na kogo ubedi da dade kakvo?

c. ???Koj kogo ubedi na kogo da dade kakvo?

d. **Koj ubedi kogo da dade kakvo na kogo?

2.2. Superiority Effects and the Principle of Minimal Compliance²³

We have not so far considered the question of how *wh*-phrases in a Bulgarian multiple question are ordered. As observed by Rudin (1988), the order of *wh*-phrases in a Bulgarian multiple question conforms to the following generalization:

(42) **Superiority Effect (Bulgarian)**

The leftmost *wh*-phrase in a Bulgarian multiple question is the *wh*-phrase that moves overtly in the corresponding English multiple question.

In other words, the contrast between Bulgarian (43a) and (43b) corresponds to an English Superiority Effect:²⁴

(43) a. Koj kakvo viĭda?

who what sees

cf. *Who sees what?*

b.*Kakvo koj viĭda?

what who sees

cf. **What does who see?*

(44) a. Koj kude udari Ivan

who where hit Ivan

cf. *Who hit Ivan where?*

b.*Kude koj udari Ivan

cf. **Where did who hit Ivan?*

The Superiority Effect in English follows naturally from AC, on the assumption that the overt instance of phrasal *wh*-movement in English is also the first instance of phrasal *wh*-movement in the derivation. By parity of reasoning, the leftmost *wh*-phrase in a Bulgarian question is also the first to move. This means that the second *wh*-phrase must have "tucked in" underneath the first phrase, forming a lower specifier than the *wh*-phrase that moved first:

(45) "Tucking in"

a. Koj kude C ___ udari Ivan ___ ?



who where hit Ivan

'Who hit Ivan where?'

b. *Kude koj C ___ udari Ivan ___ ?



where who hit Ivan

This is, in fact, the hypothesis of Richards (1997), who conducted an extensive study of multiple attraction constructions in Bulgarian and other languages. As Richards pointed out, second instances of movement to a given a tractor *should* "tuck in" if, in addition to the "attractor-oriented" condition *Attract Closest*, movement also obeys a "mover-oriented" *Shortest Move* condition. The first *wh*-phrase to be attracted by C in (44a) is, because of AC, the highest C in the structure. (Throughout this study, we shall label the *wh*-phrase of a question which occupies the highest structural position *before wh*-movement *wh*₁. Other *wh*-phrases will be labeled similarly.) *Wh*₁ moves as short a distance as possible, which in this case creates the first specifier of C. The second *wh* to be attracted in constructions like (43)-(44) is the other *wh*-phrase: *wh*₂. This *wh*-phrase could in principle create a specifier higher than *wh*₁, or it could create a specifier lower than *wh*₁. Shortest Move favors creation of a specifier lower than *wh*₁, i.e. "tucking in" as in (45a). In this way, the interaction of AC and Shortest Move causes the command relations between two *wh*-phrases after movement to mirror their order before movement.²⁵

As Richards notes, a tucking-in derivation is "countercyclic" in that it does not extend the tree. This would violate a "Strict Cycle" condition if each category were to count as a cycle — as suggested by

Chomsky (1995, chapter 1) under the rubric of the "Extension Condition", which requires each overt operation to extend the tree at its root. Richards suggests that it is not the tucking-in hypothesis, but the Extension Condition that is incorrect. The cases that actually motivated the Extension Condition and its predecessors involved derivations in which the overt-movement requirements of a lower head are satisfied later than the overt-movement requirements of a higher head. In later work, Chomsky (1995, chapter 4) suggested an alternative approach that zeros in more tightly on the actual cases that need to be excluded. He suggested that convergent derivations must satisfy the overt-movement requirements of a head H while still building the portion of the tree headed by H. (In his terminology, failure to satisfy the "strong feature" of a head H in a timely fashion "cancels the derivation".) This proposal, which Richards calls *featural cyclicity*, blocks most traditional Strict Cycle violations, but allows tucking in as in (45a), precisely because tucking in involves countercyclic movements to a single head.²⁶

Now let us consider AC more closely — in particular, its effects in multiple questions with three or more *wh*-phrases. All things being equal, one might expect AC to be obeyed in some fashion in each instance of *wh*-attraction to C_{m-spec} . For example, in a multiple question with three overtly moved *wh*-phrases, one might expect their order after movement to mirror their order before movement — just as in a multiple question with two overtly moved *wh*-phrases. As Bošković (1995; to appear) noted, this is not the case. Though the leftmost *wh* after movement must be the highest *wh* before movement, the second *wh* after movement is *not* required to be the second highest *wh* before movement. The orders ***wh*₁ *wh*₂ *wh*₃** and ***wh*₁ *wh*₃ *wh*₂** are equal (or nearly equal) in acceptability:

(46) ***Wh*₁ *wh*₂ *wh*₃ and *wh*₁ *wh*₃ *wh*₂ both acceptable**

a. Koj na kogo kakvo dade?

who to whom what gave

b. (?)Koj kakvo na kogo dade?

(47) a. Koj kogo kak udari?

who whom how hit

b. Koj kak kogo udari?

(48) a. Koj kogo kakvo e pital?

who whom what AUX asked

b. Koj kakvo kogo e pital?

The AC requirement seems to be "turned off" for movement to a given C_{m-spec} once it has been satisfied by a prior operation of *wh*-movement to that C_{m-spec} . As Richards puts it, it is as if the first operation of *wh*-movement pays an AC "tax" which allows subsequent instances of *wh*-movement to the same target to ignore this constraint, as in the following sketch of a derivation for (46b):

(49) **Paying the AC tax**

before wh-mvt:

C_{m-spec} [*koj dade na kogo kakvo*]

step one: C_{m-spec} attracts the *wh*-feature of *koj* (wh_1), pay "AC tax".

koj C [___ *dade na kogo kakvo*]

step two: C_{m-spec} attracts either of the remaining *wh*-phrases, since AC tax has been paid. Tuck in under *koj*.

koj kakvo C [___ *dade na kogo* ___]

step three: C_{m-spec} attracts the other *wh*-phrase, which tucks in under *kakvo*

koj kakvo na kogo C [___ *dade* ___ ___]

Richards (1997) argues that this view of the matter is exactly correct. He suggests further that the logic of the "AC tax" is a special case of a much more general meta-constraint that he calls the **Principle of Minimal Compliance**. I will not try to develop a precise characterization of this more general principle. (That task is undertaken by Richards 1997). The phenomenon relevant to our concerns is the simple observation that "attractor-oriented" constraints on movement apply only to the first movement operation that targets a given attractor.

(50) ***Principle of Minimal Compliance (PMC; Richards 1997)***

For any dependency D that obeys constraint C, any elements that are relevant for determining whether D obeys C can be ignored for the rest of the derivation for purposes of determining whether any other dependency D' obeys C.

An element X is *relevant* to determining whether a dependency D with head A and tail B obeys constraint C iff

- a. X is along the path of D (that is, X=A, X=B, or A c-commands X and X c-commands B).
- AND
- b. X is a member of the class of elements to which C makes reference.

(51) ***PMC for our purposes***

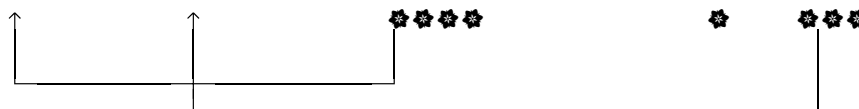
Once an instance of movement to α has obeyed a constraint on the distance between source and target, other instances of movement to α need not obey these constraints.

The PMC applies to a number of syntactic conditions. Certain island effects, for example, such as Ross's (1967) Complex NP Constraint, hold in their strongest form only of the first instance of *wh*-movement to a given C. Example (52a), for instance, displays the expected judgment for *wh*-movement out of a clause contained inside an NP. Example (52b) contains exactly the same instance of *wh* movement as (52a), except that here it is (by hypothesis) the second instance of movement to its target, rather than the first. Strikingly, the effects on acceptability of the Complex NP Constraint violation fade²⁷ to near the vanishing point.

(52) **PMC improves subjacency violations.**

- a. *Koja kniga **otreãe** senatorãt [mãlvata ãe iska da zabrani ___]?
 which book denied the-senator the-rumor that wanted to ban
 ‘Which book did the senator deny the rumor that he wanted to ban?’

- b. ?Koj senator koja kniga **otreãe** ___ [mãlvata ãe iska da zabrani ___]?



which senator which book denied the-rumor that wanted to ban

‘Which senator denied the rumor that he wanted to ban which book?’

(Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c. to Richards 1977)

These facts recall properties of English multiple questions. As is well-known, *wh*-in-situ in English multiple questions do not display island effects — or at least display them very weakly (Hankamer 1974). This fact has always been a puzzle, particularly if, as we have argued, *wh*-in-situ standardly undergo covert *wh*-movement. Traditional accounts (e.g. Huang 1982) attribute the absence of island effects in these environments to a special fact about covert *wh*-movement. As Richards notes, the phenomenon can instead be seen as the English counterpart of the effect in (52). On this view, the absence of subjacency effects with covert *wh*-movement in multiple questions is not due to the fact that the movement is covert, but to the fact that it follows an earlier application of *wh*-movement in its clause.²⁸

In this respect, Richard's Bulgarian discoveries are a perfect mirror of data from Italian discovered by Longobardi (1991), as analyzed by Brody (1995). In Bulgarian we can see multiple instances of overt movement, with the first instance paying the "subjacency tax" that allows later instances to by-pass the subjacency effect. In Italian, we can observe multiple instances of *covert* movement with the same properties. The Italian case does not involve *wh*-movement, but scope assignment to negative phrases. I follow Brody's presentation here. The relation between a negative phrase like *nessuno* 'no one' and its scope position (marked by *non* with post-verbal negation) can be separated by embedded clauses:

(53) **Italian: long distance scope**

Non approvarei che tu gli consentissi di vedere nessuno.
neg I-would-approve that you him allow to see no one
'I would not approve that you allow him to see anybody.'

Nonetheless, the relationship obeys adjunct islands (as seen in (54a)) and subject islands (as seen in (54b)).

This fact leads one to suspect that covert movement of the negative element takes place:

(54) **Italian: long distance scope obeys islands.**

a. *Non fa il suo dovere per aiutare nessuno.
neg does his duty for help no one
'He does not do his duty in order to help anyone.'

b. ?*Chiamare nessuno sarà possibile.
To call no one will-be possible.
'It will not be possible to call anyone.'

Crucially, when two negative phrases share the same scope, the lower phrase does not have to obey the islands seen in (54):

(55) **Italian: PMC improves island violations.**

a. ?Chiamare nessuno servirà a niente, ormai.
to call nobody will serve for nothing now
'To call nobody will do any good now.'

b. Non fa niente per aiutare nessuno.
neg does nothing to help nobody
'He does not do anything in order to help anyone.'

English, Bulgarian and Italian thus all display the same phenomenon: the first occurrence of movement to a given position obeys islands, but not the second — whether the two movements differ in overtness (English), are both overt (Bulgarian), or both covert (Italian).²⁹

A cautionary note is in order. Although AC is subject to the PMC, Shortest Move is another matter. Once a given instance of movement to a given C has satisfied Shortest Move, it must not be the case that subsequent instances of movement to that C may violate Shortest Move, or else we could not observe a "tucking in" requirement. This follows if Shortest Move is not a constraint governing the distance between an Attractor and Attractee, but rather a constraint that chooses the exact position near a given Attractor to which copying takes place.

2.3 Summary of Bulgarian *wh*-movement

Let us collect and organize our observations concerning Bulgarian:

- **Multiple-Specifier requirement:** The complementizer of a multiple question in Bulgarian appears to strongly require at least two *wh*-phrases as specifiers. This is a separate requirement from the general preference for moving all *wh*-phrases in a multiple question.
- **Superiority:** The left-most *wh*-phrase in a multiple question is the *wh*-phrase that was highest before movement — *wh*₁ in our terminology. If this is the same fact as the Superiority Effect in English, it suggests that the left-most *wh*-phrase is also the first to move. This entails that multiple movement to a given head proceeds via "tucking in".
- **PMC:** In Bulgarian, Attract Closest is only required to be satisfied by the first instance of phrasal *wh*-movement to C. Consequently, the second *wh*-phrase to move is not necessarily the second highest *wh*-phrase before movement (*wh*₂).

Finally, we are now able to observe that the pronunciation principles in (14)-(15), repeated below, were appropriately formulated — assuming, of course, that we were correct to identify the left-most *wh*-phrase in Bulgarian with the single overtly moved *wh*-phrase in English:

(56) **Pronunciation rule** (English) [= (14)]

- a. The first instance of *wh*-phrase movement to C is overt, in that *wh* is pronounced in its new position, and unpronounced in its trace positions.
- b. Secondary instances of *wh*-phrase movement to C are covert, in that *wh* is pronounced in its trace position, and is unpronounced in its new position.

(57) **Pronunciation rule (Bulgarian)** [= (15)]

All *wh*-phrase movement to C is overt, in that *wh* is pronounced in its new position, and unpronounced in its trace positions.

As promised earlier, I will argue that (56) and (57) are the *only* relevant differences between English and Bulgarian *wh*-questions. In particular, once *wh*-feature movement is introduced into the discussion, we will see that several peculiarities of English multiple questions are simply the reflection, in English, of the Bulgarian properties summarized in this section.³⁰ Our preliminary discussion of phrasal movement vs. feature movement in section 1, combined with our sketch of Bulgarian *wh*-syntax in this section, now permits us to develop the argument.

3. Does *wh*₁-in-situ undergo covert phrasal movement?

3.1.No.

As noted earlier, *wh*-questions show a number of apparent exceptions to the AC condition, over and above those attributable to the PMC. These are cases in which *wh*₁ is pronounced in situ, while some other *wh*-phrase is pronounced in the specifier of CP position. (I describe the situation in terms of pronunciation only so as not to prejudge the actual movement patterns.) One exception of this type that we have already mentioned involves D-linking. Relevant examples were seen in (31)-(32), of which (31) is repeated here:

- (58) **D-linking apparent exception to Superiority** [= (31)]
- a. Which person ___ bought which book?
 - b. Which book did which person buy ___?

Example (58)a shows *wh*₁ (presumably) moved³¹ and pronounced in specifier of CP. Example (58b), on the other hand, shows *wh*₂ moved and *wh*₁ in situ. Obviously, the acceptability of (58b) reveals an inadequacy somewhere in our discussion — but where? In order to answer this question, we need to know what the syntax of (58b) is.

- For example, does *wh*₁, the in-situ *wh*-phrase in (58b), undergo phrasal *wh*-movement of some sort? If the in-situ *wh*-phrase *does* undergo phrasal movement, the movement is obviously covert. If so, we must ask why?
 - Did *wh*₁, the higher *wh*-phrase, undergo phrasal movement first — but the output of movement was pronounced as covert movement for some reason, in violation of the Pronunciation rule (56)?
 - Or was AC violated, so that the first *wh* attracted was *wh*₂, the lower *wh* (with the Pronunciation rule (56) operating exactly as expected)?
- If, however, the in-situ *wh*-phrase does *not* undergo phrasal movement, we will also have to ask why? Is *wh*₁ in (58b) for some reason ineligible for *wh*-movement? Or is AC violable in situations other than those given by the PMC? Or is something else going on?

The first issue to settle, clearly, concerns whether or not *wh*₁ undergoes phrasal movement. We have in our repertoire exactly one test for specifically phrasal movement: Antecedent-Contained Deletion. We have already used this test as part of an argument that *wh*-in-situ may in principle undergo covert phrasal movement. The relevant cases were examples like (12), repeated below, in which a relative clause modifying a *wh*-phrase in situ contains an instance of ACD:

(59) **Wh-in-situ licenses ACD**

- a. Which girl invited [which student that John did [VP]]?
- b. I need to know who can speak [which languages that Ken Hale can [VP]]

These examples, of course, do not involve any apparent exception to the Superiority Effect. Neither example shows *wh*₁ in situ, so neither example tests whether the covert phrasal movement detected in (59) is also available to *wh*₁ in situ. By attaching a modifier that contains ACD to an otherwise legal instance of *wh*₁-in-situ, we should be able to determine (all things being equal) whether this type of *wh*-in-situ also undergoes covert phrasal movement. Although the relevant examples are (inevitably) complex, the experiment is straightforward, and its results are clear.

Let us begin by constructing, as an experimental control, a slightly more complex counterpart to the examples in (59), containing three relevant DP positions. Example (60) is the necessary control. The three DP positions that we will play with are the subject of *order*, the object of *order*, and the object of *congratulate*. I have glossed (60) with its intended interpretation.

(60) ***wh*₁ not in situ, ACD within *wh*₂**

I need to know which girl ___ordered [which boy that Mary (also) did] to congratulate Sarah.

I need to know for which girl x and for which boy y such that Mary ordered y to congratulate Sarah, x also ordered y to congratulate Sarah.

[i.e. I need to know the girl-boy pairs such that both the girl and Mary ordered the boy to congratulate Sarah]

The first crucial comparison is (61). Both (60) and (61) contain the same instance of ACD inside an in situ *wh*-phrase in the middle DP-position. In (60), the *wh*-phrase that moved overtly came from a higher DP position. In (61), the *wh*-phrase that moves overtly will come from the lower DP position. Thus, in

(60), ACD was not contained within an instance of *wh* in situ, but in (61), it is. In my experience, speakers uniformly find a sharp difference in the acceptability and interpretability of the two examples. Example (60) is difficult to interpret, perhaps, but (61) is absolutely impossible:

(61) ***wh*₁ in situ, ACD within *wh*₁**

*I need to know which girl Sue ordered [which boy that Mary (also) did] to congratulate __.

*I need to know for which girl x and [which boy y such that Mary ordered y to congratulate x],
Sue also ordered y to congratulate x .*

[i.e. I need to know the girl-boy pairs such that both Sue and Mary ordered the boy to congratulate the girl]

I conclude from the contrast between (60) and (61) that in apparent exceptions to the Superiority Effect like (58b), *wh*₁ does not undergo covert phrasal movement. Still, to assure ourselves that this explanation is both plausible and correct, we must construct other experiments that test other explanations of the contrast between (60) and (61). For example, one might ask if the meaning of (61) is in some way especially difficult to construct. This can be tested by examining constructions whose meaning is similar or identical to (61), but whose syntax is different. Such experiments can be performed. Example (61) involves a *wh*-phrase overtly moved from the third DP position. This *wh*-phrase binds an elided bound variable inside the second DP position. An equivalent interpretation can be obtained by switching the initial positions of the bound variable and *wh*-phrase, and overtly moving *wh*₁ (which now contains the second *wh*-phrase in situ). The relevant construction is (62), which expresses the meaning of (61), but is quite acceptable. If there is any deviance to), it may be related to the failure of surface c-command between *which girl* and *her*, but the contrast with (61) is sharp.

(62) **It's not the meaning of (61) that renders it unacceptable...**

?[Which boy that Mary ordered to congratulate which girl] did Sue (also) order ___ to congratulate her?

As just noted, the interpretation of the elided VP in (61) includes a variable bound by the overtly moved phrase *which girl*. This raises another possible account of the deviance of (61). One might wonder if this variable within an elided VP obeys the Complex NP Constraint, as a trace inside an overt VP would. That is, could the contrast between (60) and (61) stem from a violation of the Complex NP Constraint in (61)? The answer seems to be no. Consider, for example, an instance of ACD which minimally differs from (61) in that the position occupied by *wh* 1-in-situ in (61) is occupied by a non-*wh*-phrase. An example is, which replaces *wh* 1-in-situ with a quantified phrase headed by *every*. Since *every boy that Mary did* is not a *wh*-phrase, it does not compete with *which girl* for movement to C. Consequently, no special problem is expected with ACD — and none is found. Crucially, however, the ellipsis in (63) contains a variable bound by *which girl* across a Complex NP boundary. Consequently, it is unlikely that this factor accounts for the unacceptability of (61):

(63) **...nor the Complex NP Constraint violation within the ellipsis site...**

I need to know which girl Sue ordered [every boy that Mary (also) did] to congratulate ___.

I need to know for which girl x, if Mary ordered a boy to congratulate x, then Sue also ordered that boy to congratulate x.

It is also worth verifying that if the *wh*-in-situ and the overtly moved *wh*-phrase are reversed (that is, if the apparent violation of the Superiority Effect is eliminated), ACD becomes acceptable. Though the example might (for some speakers) involve a Weak Crossover effect, the conclusion seems to be correct. A useful prime for (64) are other sentences in which an in situ *which*-phrase binds a variable to its left, e.g. *Which man ordered a boy that liked her to congratulate which girl.*

(64) **...nor the apparent Superiority violation**

(*wh*₂ in situ, ACD within overtly moved *wh*₁)...

(?) I need to know [which boy that Mary (also) did] Sue ordered ___ to congratulate which girl.

*I need to know for which girl x and [which boy y such that Mary ordered y to congratulate x],
Sue also ordered y to congratulate x .*

[i.e. I need to know the girl-boy pairs such that both Sue and Mary ordered the boy to congratulate the girl]

A few other tests can be performed. Hiroshi Aoyagi (personal communication), for example, notes that it is important to make sure that the variable in *wh*₁ (inside the ellipsis site) in (61) can in fact be bound by the overtly moved *wh*, as it needs to be. For example, one might ask whether *wh*₁ undergoes covert movement to a position *higher* than the overtly moved *wh*-phrase, making binding of a variable within *wh*₁ impossible for simple reasons of scope and c-command. Examples like (65) seem to suggest that this is not the case. Here the parasitic gap within the in-situ *wh*₁ is successfully bound by the moved *wh*-phrase:

(65) **...nor binding into *wh* in situ...**

Which girl did you persuade [which friend of ___] to congratulate ___.

Another relevant example is due to James McCawley (personal communication). Example (66) below is like (59), except that it has in common with (61) a variable — here a pronoun — following *wh*₂-in-situ.

While it was actually McCawley's judgment that the example is deviant, I disagree:

(66) **...nor the variable bound by the moved *wh* to the right of the *wh* in situ.**

I need to know which girl __ ordered [which boy that Bill (also) did] to congratulate her.

In my view, the tests we have performed point to one clear conclusion. Whatever the syntactic fate of *wh*₁ in situ in apparent Superiority violations like (58b), covert phrasal *wh*-movement is *not* one of the things that can happen to it. The next question to consider, of course, is what *does* happen to *wh*₁ in situ in such examples.

3.2. Covert phrasal movement and case

Before turning to this question, however, I want to point out an interesting corollary result. The comparison between (60) and (61) has revealed a link between the availability of ACD inside a *wh*-phrase and the overall pattern of *wh*-movement: *Wh*₁ in situ in a multiple question does not undergo covert phrasal movement. Since it does not undergo covert phrasal movement, we can conclude a fortiori that it does not undergo covert phrasal *wh*-movement, which is our main concern in this study. But it is worth noting some implications of the stronger conclusion, because it bears directly on a current debate concerning the syntactic configurations that resolve ACD. If we consider the various explanations for ACD resolution that have been offered in the literature, and focus our attention on those explanations that posit covert phrasal movement, we can distinguish two families of subproposals.

One school of thought — represented, for example, by May (1985, 11-14) — has argued that ACD resolution not only provides evidence of covert phrasal movement, but also reveals severe limitations on its distribution. We can call this school of thought the *limited movement* family of theories. In particular, May suggested that constraints on ACD teach us that covert phrasal movement is limited to special types of constituents — constituents whose special syntactic or interpretive needs motivate covert phrasal movement. The contrast between (67a) and (67b) was taken to show that quantified expressions, but not DPs headed by proper names, undergo the covert phrasal movement that can resolve ACD:

(67) **Quantifiers vs. proper names in ACD**

- a. Dulles suspected [everyone who Angleton did].
- b. *Dulles suspected [Philby, who Angleton did].

May's original explanation for (67) does not seem quite correct, however. The contrast between (67a) and (67b) does not seem to correlate directly with the quantificational nature of the relevant phrase, since focusing the VP inside the relative clause renders (67b) completely acceptable:

(68) **Focus eliminates the quantifier/proper name contrast.**

- a. Dulles suspected [Philby, who Angleton also did].
- b. Dulles suspected [Philby, who Angleton did not].
- c. Dulles did not suspect [Philby, who Angleton did].

The contrast between (67b) and (68) is familiar from other cases of VP-ellipsis (or even simple VP-repetition, in some cases) and seems to be unrelated to ACD:³²

(69) **Contrast not limited to ACD**

- a. *Dulles suspected Philby. Angleton did .
- b. Dulles suspected Philby. Angleton also did .
- c. Dulles suspected Philby. Angleton did not .
- d. Dulles did not suspect Philby. Angleton did .

The real puzzle, then is the absence of this focus requirement in restrictive relatives such as the relative clause in (67a). I will not attempt to solve this puzzle here. In any case, Fiengo and May (1994, 242) note that a wide variety of NP-types support ACD.

(70) **ACD with other NP-types**

- a. Dulles suspected some spy that Angleton did .
- b. Dulles suspected the spy that Angleton did .
- c. Dulles suspected many spies that Angleton did .

These facts raise the question of whether, in Fiengo and May's words, "the construction can [still] be taken as a diagnostic for quantification". They suggest that it can. For example, they take (70b) to argue in favor of the Russellian analysis of definite descriptions as quantificational. And they note, with Diesing (1992) (though disagreeing over details), that the determiners in (70c) and similar examples receive a strong interpretation, which suggests some sort of QR-like operation on the object NP.

The other school of thought (which we can call the *case movement* family of theories) rejects this conclusion. In particular, it abandons the view that the phrasal movement which resolves ACD in a direct object is limited to particular subtypes of DPs. According to this proposal, developed independently by Lasnik (1993) and by Hornstein (1994; 1995), the covert phrasal movement that licenses ACD is not limited to "special DPs", but is actually a much more "garden-variety" type of movement — namely, movement for reasons of Case. Lasnik and Hornstein develop their proposals in the context of Chomsky's (1995, chapters 2-3) proposal that object Case is checked — not within VP — but in a VP-external projection (AGRoP in Chomsky 1995, chapter 1; *v*P in Chomsky 1995, chapter 4) to which the object moves. They suggest that it is this kind of Case-driven movement, not QR or covert *wh*-movement, that resolves ACD in the standard cases.³³ They develop this proposal as part of a broader proposal that QR does not exist, which I will not report here.³⁴

The central difference between the "limited movement" and the "case movement" approaches lies in whether ACD resolution within direct objects is brought about by special processes that are limited to certain DP-types (quantified phrases and *wh*-phrases), or whether ACD resolution is available to all direct object DPs that bear object Case. Some of the arguments developed by Kennedy (1997) help answer this question, by arguing for the weak point that covert operations other than Case movement must at least be *allowed* to resolve ACD. As Kennedy notes, there are instances of ACD that require a type of movement that Case

considerations do not motivate. For example, "NP-contained ACD" of the sort seen in (71) (Kennedy's (68)-(70)) has an interpretation entirely unexpected under theories in which ACD is resolved through Case movement:

(71) **NP-contained ACD (Kennedy's examples)**

- a. Beck read [a report on [every suspect Kollberg did]].
- b. Melander requested [copies of [most of the tapes Larsson did]].
- c. Kollberg took [pictures of [the same people Beck did]].

If any DP in (71) is assigned Case outside its VP, it is the external bracketed DP. The interpretation of ellipsis in these examples, however, is unexpected if ACD is resolved by Case movement. Instead, the interpretation is that derived if the *internal* bracketed DP undergoes covert movement (e.g. QR) to a VP-external position:

(72) **LF representations for (71)**

- a. [every suspect Kollberg read a report on]_x Beck read a report on *x*
- b. [the tapes Larsson requested copies of]_x Melander requested copies of most of *x*.
- c. [Kollberg took pictures of the same people]_x Beck took pictures of *x*.

Thus there must be some type of covert phrasal movement distinct from Case movement that resolves ACD.³⁵

The contrast between (60) and (61) provides a good argument in favor of an even stronger position. If object Case really is checked outside VP, through movement into AGRoP or vP, this kind of movement cannot involve the entire object phrase, since it does not resolve ACD. The point is straightforward. Examples (60) and (61), if we have interpreted them correctly, display an interaction between *wh*-movement and ACD resolution. More specifically, they show that when a lower *wh*-phrase undergoes overt *wh*-

movement past a higher direct-object *wh*-phrase, the higher phrase cannot contain an instances of ACD. If direct object phrases uniformly undergo a type of movement for Case purposes which is capable of resolving ACD, then the interaction with *wh*-movement seen here could only entail (all things being equal) that movement of a lower *wh*-phrase over a higher prevents the higher *wh*-phrase from undergoing movement for Case purposes. But this conclusion cannot be correct. If object Case cannot be checked on the higher *wh*-phrase, *wh*₁, in these configurations, then — independent of ACD — such configurations should have the status of Case filter violations. But, of course, they do not. Remove the ACD from (61) and the structure, though complex, is not a Case filter violation, as (73) shows. Whatever might be wrong with (61), Case is not the culprit:³⁶

(73) ***Wh*₁ in situ does not violate the Case filter.**

I need to know which girl Sue ordered which boy to congratulate ___.

A "limited movement" hypothesis offers a more plausible perspective on the problem of (61). Suppose only covert *phrasal wh*-movement can resolve ACD in a relative clause that modifies an in-situ interrogative *wh*-phrase. Though we have not yet explained why covert *phrasal wh*-movement is not available in these cases, the unavailability of covert *phrasal wh*-movement should not be totally surprising — even at this point in our story. After all, there *is* something special about *wh*₁-in-situ from the perspective of *wh*-movement. Furthermore, the hypothesis that *wh*₁-in-situ does not undergo covert *phrasal wh*-movement helps us understand the contrast between (61) with ACD and (73) without ACD, since the presence of *phrasal* movement is crucial only to ACD.

Consequently, if Case on direct objects is checked external to VP,³⁷ the kind of direct object movement that accomplishes this task cannot be *phrasal* movement. Otherwise, whether or not *wh*₁-in-situ also undergoes *phrasal wh*-movement, Case movement alone would license ACD in examples like (61). To put the matter differently: Hornstein and Lasnik were correct in noting that Case-driven *phrasal* movement to a VP-external position should license ACD, but they were wrong about the existence of this sort of *phrasal* movement.

Indeed, the conclusion that object Case movement (if it exists) is not phrasal excludes not only proposals that posit that English direct objects undergo *covert* phrasal movement, but also proposals like Johnson's (1992) and Koizumi's (1995) that English direct objects undergo *overt* phrasal movement outside VP for Case reasons. The moment we posit *phrasal* object movement for Case reasons, we lose almost all reasonable hope of accounting for the ACD contrasts that concern us. Either there is no movement for case purposes to a VP-external position, or else there is only feature movement. The latter is, in fact, Chomsky's (1995, chapter 4) proposal.^{38,39}

These conclusions from (60)- (61) fit quite nicely with arguments by Kennedy (1997) that also make the stronger case — i.e. not merely that operations like QR and covert *wh*-movement serve to resolve ACD, but also that covert object Case movement does not. For example, Kennedy notes (p.675), that the *de dicto* reading normally available for the subjects of infinitives embedded under *want* disappears when that subject contains an elided VP whose antecedent is the VP headed by *want*.

(74) **De dicto reading disappears (Kennedy)**

- a. Kollberg [_{VP} wants [everyone Beck wants to interrogate] to answer these questions].
- b. Kollberg [_{VP} wants [everyone Beck does] to answer these questions].

As Kennedy points out, (74a) "is ambiguous between a [*de re*] reading...in which Kollberg's desire concerns particular individuals, and [*de dicto*] one ... in which his desire concerns any individual who meets the requirements imposed by the restriction (that Beck wants to interrogate him)". By contrast, (74b) allows only the *de re*, "particular individual" reading. It cannot mean that Kollberg wants the set of people who answer the questions to coincide with the set of people that Beck wants to answer the questions, whoever that set may happen to consist of. In fact, simpler examples can be constructed (as Danny Fox, p.c., has pointed out) with bare DP objects.⁴⁰ These also display a *de re/de dicto* ambiguity eliminated by ACD:

(75) **De dicto reading disappears (Fox)**

- a. Kollberg [_{VP} wanted [every book that Mary wrote]].
- b. Kollberg [_{VP} wanted [every book Mary did]].

In (75a), Kollberg might desire a particular set of books (without even knowing that they were written by the same person), or his desire may concern any book that Mary happens to have written. In (75b), Kollberg's desire must concern a particular set of books.

These contrasts follow straightforwardly if ACD resolution requires covert phrasal movement of the bracketed phrase to a position that also happens to be outside the semantic scope of *want*. But this position cannot be the case position for the bracketed DP, or else the *de dicto* reading would also be unavailable in (74a) and (75a), contrary to the facts. Therefore, the case position for the embedded subject, wherever it is, must have two properties: (1) it lies within the scope of *want*, and (2) movement to it does not license ACD resolution.⁴¹

4. What *does* happen to *wh*₁-in-situ?

4.1. Does *wh*₁-in-situ undergo feature movement?

So far, I have argued that in D-linked exceptions to the Superiority Effect, *wh*₁-in-situ does not undergo covert phrasal movement. This argument was possible only because the ACD behavior of *wh*₁-in-situ could be contrasted with the behavior of instances of *wh*-in-situ that do undergo covert phrasal movement. So far, this conclusion accords with Pesetsky (1987), where I also argued that *wh*₁-in-situ fails to undergo covert phrasal movement.⁴² At this point, however, the work of the last few years raises new questions. In Pesetsky (1987), I assumed that if an in-situ *wh*-phrase fails to undergo covert phrasal movement, it undergoes no movement at all. For that reason, I suggested that a special mechanism ("unselective binding") assigns such *wh*-in-situ their scope, but that no instance of *wh*-movement applies to them. Now, however, we can envision the possibility of a less ad hoc proposal. In particular, we can ask

whether an instance of *wh*-in-situ that fails to undergo *wh*-phrase movement might nonetheless undergo *wh*-feature movement. In (58b), for example (*Which book did which person buy ___?*), even though *wh*₁, *which person*, fails to undergo covert phrasal *wh*-movement, could it be that its *wh*-feature has moved on its own to the interrogative *C_{m-spec}*, leaving the rest of the phrase behind?

This possibility suggests a new perspective on apparent exceptions to the Superiority effect. (I call them "apparent exceptions" from this point on.) Perhaps *wh*₁-in-situ in these examples actually undergoes *wh*-feature movement. If so we might entertain the possibility that this instance of feature movement is actually the *first* instance of *wh*-movement in its clause — with the first instance of overt phrasal movement actually constituting the *second* instance of *wh*-movement overall. On this view, our pronunciation principle for English *wh*-movement remains accurate in exactly the form given earlier. The first instance of *phrasal* movement is overt, even if the first instance of phrasal *wh*-movement is not the first instance of *wh*-movement:

(76) **Feature movement analysis of apparent exceptions to Superiority Effect**

structure before movement:

C_{m-spec} [*which person bought which book*]

step one: *C_{m-spec}* attracts the *wh*-feature of *which person*

F_i-C [*F_i-which person bought which book*]

step two: *C_{m-spec}* attracts the *wh*-phrase *which book*

which book F_i-C [*F_i-which person bought ___*]

pronounced result:

Which book did which person buy?

If this hypothesis is correct, then — although such examples sound as if they violate AC — in reality, AC is not violated at all. The first *wh*-feature attracted by C_{m-spec} in the course of the derivation is the closest instance of this feature. The only relevant difference between *Which book did which person buy ___?* and *Which person ___ bought which book?* lies in whether the first instance of *wh*-movement is feature movement or phrasal movement:

This hypothesis, in turn, raises the possibility that AC for the *wh*-feature is inviolable, just as suggested by Chomsky (1995, chapter 4).⁴³ The appearance of counterexamples arises simply from the failure to consider the possibility that the first instance of overt movement to C_{m-spec} might not be the first instance of movement to C_{m-spec} . On this hypothesis, the failure of ACD in a relative clause attached to *wh* 1-in-situ⁴⁴ (as seen in (61)) stems from the same cause as the failure of ACD in relative clauses attached to the associate of *there*, studied earlier in connection with example (25). In each case, movement takes place from the constituent that contains ACD, but because the movement is not phrasal, ACD is not resolved.

4.2. What is special about questions with D-linking?

This proposal comes no closer than previous efforts to explaining why the semantic property of D-linking yields an exception to the Superiority Effect in (76), but it suggests a new way of looking at the syntactic side of the problem. Nothing so far has excluded the possibility of a derivation like (76) for English examples without D-linking — yet these do show a Superiority Effect. We need to know what is wrong with a derivation like (77):

(77) **Spurious Derivation of an apparent exception to Superiority**

C_{m-spec} [who bought what]

step one: C_{m-spec} attracts the *wh*-feature of *who*

F_i -C [F_i -who bought what]

step two: C_{m-spec} attracts the *wh*-phrase *what*

what F_i -C [F_i -who bought ____]

pronounced result:

*What (did) who buy?

I want to suggest that what is wrong with (77) is that it violates a constraint that we have already seen in a different form in Bulgarian: the "multiple-specifier requirement" of C_{m-spec} presented in (36), repeated below:

(78) **Multiple-Specifier requirement of C_{m-spec}**

C_{m-spec} requires more than one *wh*-specifier.

Recall that this requirement displayed itself in Bulgarian as the strong requirement that at least two *wh*-phrases undergo overt phrasal movement in a multiple question. If a multiple question involves a total of only two *wh*-phrases, and one of them undergoes feature movement to C, the Multiple-Specifier requirement cannot be satisfied. This, I suggest, is what goes wrong with the derivation sketched in (77). Feature movement involves communication between the *wh*-feature of *who* and a corresponding feature of C_{m-spec} , but does not involve copying of a phrase into a specifier position, which could help satisfy the Multiple-Specifier requirement.⁴⁵

If this idea is correct, then the Multiple-Specifier requirement must somehow be suspended in questions that involve D-linking, or else the derivation in (76) would also lead to deviance. Let us suppose that this is the case:

(79) **Multiple-Specifier requirement of C_{m-spec} (revised)**

Except in questions where a *wh*-phrase is D-linked, C_{m-spec} requires more than one *wh*-specifier.

Alternatively, D-linked multiple questions might be formed (and receive normal "pair-list" readings) with the complementizer otherwise reserved for non-multiple questions — the interrogative complementizer C_{I-spec} that takes only one *wh*-phrase as specifier. On this view, we would not need the exception clause built into (79), but we would need to understand why the semantics of D-linked multiple questions allows a complementizer that is otherwise not found in such questions. I do not know at present how to decide between the two hypotheses.⁴⁶

More generally, it remains unclear why D-linked phrases should constitute an exception to the Multiple-Specifier requirement. In this domain, the present proposal leaves an important question unanswered. This is not grounds for giving up on the proposal, however. Even when one is forced to leave some aspect of a phenomenon unexplained, one can often tell whether the discussion is on the right track by examining whether the unexplained portion of the phenomenon has internal coherence. If so, there are grounds for optimism about the overall approach — grounds for suspecting that the unanswered question is a real question about the true theory, and not merely an artifact of missteps and false notions. In the present case, if the absence of Superiority Effects with D-linked *wh* in English is due to an exception to the Multiple-Specifier requirement, we immediately expect this exception to show up in a slightly different way in Bulgarian. In English, the Multiple-Specifier requirement, combined with the pronunciation principles we stated earlier, produces one overt occurrence of *wh*-phrase movement, and one covert occurrence. In Bulgarian, the Multiple-Specifier requirement, combined with the pronunciation principles for that language, forces the *overt* fronting of at least two *wh*-phrases. Consequently, if the D-linking exception to English Superiority effects arises from an exception to the Multiple-Specifier requirement, we expect D-linking to improve Bulgarian multiple questions in which just one *wh*-phrase has fronted. As already mentioned in

footnote 24, this is the case. Examples like the double-asterisked (35d), (40d) and (41d), in which only one out of three or four *wh*-phrases has fronted, are significantly improved when one or the other *wh*-phrase is understood as D-linked (that is, as demanding answers drawn from a set which counts as old information for speaker and hearer). The same is true of questions with only two *wh*-phrases ("binary questions"):

(80) **D-linked *wh*-in-situ in Bulgarian**

a. Koj dade kakvo na Stefan? [* if non-D-linked; better if D-linked]

who gave what to Stefan

b. Koj kakvo dade na Stefan? [no D-linking preference]

(Roumyana Izvorski, p.c.)

Furthermore, as mentioned in footnote 19, a binary question may constitute an exception to Superiority, so long as the question involves D-linking — just as in English:

(81) Kakvo koj dade na Stefan? [* if non-D-linked; better if D-linked]

If we are correct in our suppositions about *wh*-feature movement, (81) receives an analysis akin to (76), with the Multiple-Specifier requirement irrelevant because of the D-linking exception.⁴⁷

4.3. Non-binary multiple questions

As it happens, Pesetsky (1987) (citing Wachowicz (1974) and Choe, personal communication) already presented the Polish counterpart of the previous section's Bulgarian facts as an argument for the hypothesis that D-linked *wh*-in-situ may fail to undergo covert phrasal *wh*-movement. The way that paper saw things, questions with D-linking were not an exception to a Multiple-Specifier requirement, but rather were an exception to a proposed requirement that each *wh*-phrase undergo some form of phrasal *wh*-movement by LF. Polish was seen as a language that "wears its LF on its sleeve", thanks to the phenomenon of multiple overt *wh*-phrase movement. The discussion had certain marked deficiencies, however. For its conclusions to have force, it was necessary to assume that Polish *wh*-movement is always overt (an assumption not made explicit in Pesetsky 1987). Furthermore (needless to say) the possibility of feature movement was not considered at all. In this context, it was reasonable to assume that D-linking is an exception to a movement requirement for *wh*-phrases, rather than an exception to a requirement on the specifiers of C_{m-spec} .

In fact, however, evidence now available to us favors the latter view over the former. Recall that the Multiple-Specifier requirement is satisfied once two *wh*-phrases have moved to SPEC of C_{m-spec} . Superiority effects arise in binary questions without D-linking because such questions contain only two *wh*-phrases in the first place. Consequently, a derivation in which one of the two *wh*-phrases undergoes feature movement inevitably violates the Multiple-Specifier requirement. The situation should be different, however, in multiple questions that contain *more than two* *wh*-phrases. In such a question, nothing said so far should block a derivation in which one or more of the *wh*-phrases undergoes *wh*-feature movement, so long as at least two *wh*-phrases undergo the phrasal version of *wh*-movement. In particular, in a question with three or more *wh*-phrases, *wh*₁ (the highest *wh*-phrase) should be able to undergo *wh*-feature movement as the first instance of *wh*-movement (satisfying AC), with two of the lower *wh*-phrases undergoing phrasal movement. Of these instances of phrasal movement, the first will be overt in English, and the rest will be covert. In other words, independent of D-linking, questions with three or more *wh*-

phrases should show no Superiority effect. In fact, this is the case, as I already pointed out when presenting example (33), repeated below.

(82) **No Superiority effect with three or more *wh*-phrases.**

- a. *What did who give ___ to Mary? [detectable Superiority effect]
- b. What did who give ___ to whom? [No detectable Superiority effect]

Example (82a) contains only two *wh*-phrases. Consequently, both must undergo phrasal *wh*-movement, in order to satisfy the Multiple Specifier requirement of C_{m-spec} . Since AC is inviolable, *who* must move first, and *what* second. By the pronunciation principles, movement of *who* should be the overt instance of movement. Since this is not the case in (82a), the sentence is deviant. Example (82b), on the other hand, contains three *wh*-phrases. The highest *wh*-phrase, *who*, is thus free to undergo *wh*-feature movement, as the first instance of *wh*-movement to the complementizer of the question. *What* and *whom* then undergo phrasal *wh*-movement, satisfying the Multiple Specifier requirement. The first instance of *wh*-movement is overt, and the result is (82b).

The disappearance of Superiority effects with non-binary multiple questions makes it clear that the grammar of multiple questions contains a Multiple Specifier requirement, and not a requirement that all *wh*-phrases undergo phrasal *wh*-movement. This, in turn, tells us exactly which property of grammar D-linked questions are exceptions to. Despite our ignorance concerning the semantic source of this D-linking exception, the strong hidden parallels between Bulgarian and English revealed by this approach suggest that we are on the right track. The failure of D-linked questions to display the Superiority Effect in English groups with the fact that D-linking in Bulgarian obviates the requirement that two *wh*-phrases move overtly. The fact that English questions with more than two *wh*-phrases also fail to show the Superiority Effect groups with the fact that Bulgarian multiple questions with two overtly moved *wh*-phrases are significantly better than multiple questions with only one overtly moved *wh*-phrase.

An additional observation may support this perspective. From time to time, I have encountered speakers of English who report a residual Superiority Effect in those cases where other speakers report its disappearance. For these speakers, the Superiority contrast weakens in questions with D-linking and in non-binary multiple questions, but remains detectable. It is not implausible that this "lesser" Superiority contrast reflects the effect in English of Bulgarian's "general preference" reported in (37) for multiple questions in which all *wh*-phrases are attracted to C_{m-spec} . Cases of apparent violations of Superiority, if our proposals are correct, always involve at least one *wh*-phrase (*wh*₁) undergoing feature movement, rather than phrasal movement, to C_{m-spec} .⁴⁸

The same considerations may help us understand comparable cases in Bulgarian. The Superiority Effect in Bulgarian shows up as the observation that where *wh*₁ has clearly undergone overt *wh*-movement, it is leftmost among the *wh*-phrases that have fronted. We have already seen this effect in binary questions in (43)-(44). The same effect can be observed in ternary questions. Where it is clear that *wh*₁ has undergone overt *wh*-phrase movement, it must be leftmost among the various *wh*-phrases. Examples (83a-b) must be parsed either with *koj* in situ, violating the Multiple Specifier requirement, or else with *koj* having undergone overt *wh*-phrase movement later than a lower *wh*-phrase — in violation of AC. That is why these examples are completely unacceptable:

(83) **Violation of multiple specifier requirement or violation of AC**

a. **Na kogo koj dade kakvo?

to whom who gave what

b. **Kakvo koj dade na kogo?

Such examples contrast detectably with minimally different (though still unacceptable) examples in which *koj* is arguably completely in situ. While one may remain skeptical of arguments based on a contrast between strong and weaker unacceptability ("**" vs. "*"), the contrast here is apparently quite clear.⁴⁹

(84) **Multiple specifier requirement satisfied /AC satisfied**

*Na kogo kakvo dade koj?

to whom what gave who

If our hypothesis is correct, the first instance of *wh*-movement to C_{m-spec} in (84) is *wh*-feature movement from *koj*. The second instance of *wh*-movement involves the *wh*-phrase that ends up left-most, and third instance of *wh*-movement involves the second overtly fronted *wh*-phrase. The contrast between (83) and (84) follows as the Bulgarian counterpart to the English difference between binary and non-binary multiple questions that appear to violate Superiority. Of course, (84) is not fully acceptable, as the asterisk indicates. This fact may reflect the preference for multiple questions in which all *wh*-phrases undergo *wh*-phrase movement (especially if this preference is stronger for Bulgarian speakers, who can hear the difference, than it is for English speakers).⁵⁰

One final empirical point must be investigated. The proposal in this section has presupposed that *wh* 1-in-situ in non-D-linked ternary questions has essentially the same syntax as its counterpart in D-linked binary questions. That is, it fails to undergo phrasal *wh*-movement — but it does undergo *wh*-feature movement, if our hypothesis is correct. ACD evidence bearing on this point is hard to come by. In order to construct a fair test, we need to avoid D-linking, and we need at least four argument places to work with (in order to vary the status of the *wh*-phrase with ACD from *wh* 1-in-situ to *wh* 2-in-situ while holding other conditions constant). To the extent that such examples can be constructed, and to the extent that they do not pose insuperable burdens on linguistic memory, the facts appear to be as expected. It is exceedingly difficult to parse (85a-b), but insofar as one can judge the examples at all, they seem to contrast as indicated below. Example (85b) suffers from the repetition of *who*, but its instance of ACD seems to be interpretable, unlike ACD in (85a). The problem with the judgment is the excessive concentration required in order to deliver it:

(85) **ACD in *wh*₁ within three-*wh* questions**

- a. *What did Sue order [who that Mary (also) did] to give ___ to whom?
- b. What did who order [who that Mary (also) did] to give ___ to whom?⁵¹

4.4. *Wh*-feature movement vs. no *wh*-movement at all

The discussion in subsections §4.1-4.3 has provided only one real argument that directly bears on the existence of *wh*-feature movement. This was the observation that if *wh*-feature movement can be the first instance of *wh*-movement, examples that look like violations of AC really do not violate the constraint at all. If we accept the evidence from the *there*-construction in favor of the existence of non-phrasal feature movement, this consideration is rather significant, since it explains the existence of apparent Superiority violations as a consequence of independently detected properties of grammar. In this sense, it is superior to certain potential alternatives.

For example, one might suppose that the *wh*-phrases that undergo *wh*-feature movement (according to our hypotheses) are actually "invisible" to *wh*-attraction by C_{m-spec} and undergo no *wh*-movement whatsoever. On this view, so long as the Multiple Specifier requirement is satisfied in the end, an invisible *wh*₁ would be free to remain overlooked by *wh*-attraction. The Multiple Specifier Requirement would filter out non-D-linked binary questions in which one of the two *wh*-phrases had been marked invisible.⁵² The cost of this proposal, however, would be the postulation of a construction-specific "invisibility" motivated only by apparent Superiority violations. The feature-movement proposal may be the less complex of the two — but we would still like more evidence bearing on the issue.

It is clear what we want to find out. We know from ACD that in apparent violations of Superiority, *wh*₁ undergoes neither overt nor covert phrasal *wh*-movement. Can we tell whether it undergoes any sort of

wh-movement — in particular, feature-movement? More specifically, do we have tool for detecting whether *wh*₁-*in-situ* undergoes feature-movement before the first instance of overt *wh*-phrasal movement?

Surprisingly, we do have such a tool. One might think it impossible to distinguish phonologically vacuous feature movement from the absence of movement, but we can actually distinguish these possibilities by looking at the impact that feature movement should have on other, later operations that do have phonological consequences. Richards' (1997) *Principle of Minimal Compliance* (PMC), discussed in section 2.2, provides us with just the tool we need to detect this impact.⁵³

Recall how we introduced this principle. Following Richards (who in turn used evidence from Bošković 1995; to appear), we noted that the first instance of *wh*-movement in a Bulgarian multiple question has a notable effect on subsequent instances of *wh*-movement within that multiple question. In particular, once overt *wh*-phrase movement has taken place in accordance with AC, subsequent instances of *wh*-phrase movement do not need to satisfy AC.⁵⁴ As we discussed, and as illustrated in the derivation from (49) repeated below, it is as if satisfying AC with the initial instance of movement to *C_{m-spec}* pays an "AC tax" that provides immunity to AC for other instances of *wh*-movement to that *C_{m-spec}*. As a consequence, orders like *wh*₁ *wh*₂ *wh*₃ and *wh*₁ *wh*₃ *wh*₂ are nearly equal in acceptability, as example (46), repeated below, shows:

(86) **Bulgarian: *wh*₁ *wh*₂ *wh*₃ and *wh*₁ *wh*₃ *wh*₂ both acceptable**

a. Koj na kogo kakvo dade? (= (46))

who to whom what gave

b. (?)Koj kakvo na kogo dade?

(87) **Demonstration of AC tax** (=49))

before wh-mvt:

C_{m-spec} [*koj dade na kogo kakvo*]

step one: C_{m-spec} attracts the *wh*-feature of *koj* (H), pay "AC tax".

koj C [__ *dade na kogo kakvo*]

step two: C_{m-spec} attracts either of the remaining *wh*- phrases, since AC tax has been paid. Tuck in under *koj*.

koj kakvo C [__ *dade na kogo* __]

step three: C_{m-spec} attracts the other *wh*-phrase, which tucks in under *kakvo*

koj kakvo na kogo C [__ *dade* ____]

This effect, if we have understood it correctly, provides us with a way of telling whether an instance of overt *wh*-movement in an English multiple question is the first or the second instance in its clause. If overt *wh*-movement in English has been preceded by another type of *wh*-movement, it should show no effects of AC, since the earlier instance of *wh*-movement has already paid an "AC tax". On the other hand, if the overt *wh*-movement in a particular multiple question is the first instance of *wh*-movement, it should show the effects of AC.

Consider in particular a question with three *wh*-phrases: *wh*₁, *wh*₂ and *wh*₃. By the Multiple Specifier requirement, we know that at least two of them must undergo phrasal movement. As a factual matter, we know that questions with three *wh*-phrases may appear to violate Superiority. Consider such a case. By hypothesis, *wh*₁ is not undergoing phrasal movement at all, but *wh*₂ and *wh*₃ do. The question is, which one undergoes phrasal movement first?

If we suppose that apparent Superiority violations are allowed because *wh*₁ can be marked as "invisible" to movement, then what we hear as overt *wh*-movement is the first instance of *wh*-movement in its clause. This instance of movement would have to involve *wh*₂, because *wh*₂ is the closest *attractable wh*-phrase. (If *wh*₂ were marked as "invisible", the first instance of *wh*-movement would involve *wh*₃, but the structure would be filtered out by the Multiple Specifier requirement.) Consequently, apparent Superiority violations in questions with three *wh*-phrases would need to show *wh*₂ overtly moved, with *wh*₃ remaining in situ. On the other hand, if apparent Superiority violations arise simply because *wh*₁ may undergo *wh*-feature movement, then this instance of feature movement is expected to pay the "AC tax" for its multiple question. The next instance of *wh*-movement must be phrasal, or else the Multiple Specifier requirement would not be satisfied in the end. This instance of *wh*-movement will also be overt. But because the AC tax has already been paid, we expect both *wh*₂ and *wh*₃ to be possible candidates for the overt instance of *wh*-movement.

The facts support the feature movement proposal. In ternary questions with *wh*₁-in-situ, either *wh*₂ or *wh*₃ may undergo overt *wh*-movement. The only interfering factor is a slight dispreference for sequences of identical *wh*-phrases in these structures. This factor degrades sequences like *who did who...* In the pair of paradigms presented below, it can be seen that this dispreference cross-cuts the effect under investigation. In (88), it favors overt movement of *wh*₂, but in (89), it favors overt movement of *wh*₃:⁵⁵

(88) **English: *wh*₁ *wh*₂ *wh*₃ and *wh*₁ *wh*₃ *wh*₂ both acceptable**

- | | |
|-----------------------------------|---|
| a. Who ___ gave what to whom? | [<i>wh</i> ₁ moves overtly] |
| b. What did who give ___ to whom? | [<i>wh</i> ₂ moves overtly] |
| c. ?Who did who give what to ___? | [<i>wh</i> ₃ moves overtly] |

- (89) a. Who ___ persuaded whom to buy what? [wh₁ moves overtly]
b. ?Who did who persuade ___ to buy what? [wh₂ moves overtly]
c. What did who persuade whom to buy ___? [wh₃ moves overtly]

The derivation of (89c) proceeds as follows:

(90) **Derivation of (89c)**

before wh-mvt:

C_{m-spec} [who persuaded whom to buy what]

step one: C attracts the wh-feature of who (H), pay "AC tax" (PMC).

F_i-C [F_i-who persuade whom to buy what]

step two: C attracts either of the remaining wh- phrases, since PMC no longer requires obedience to AC.

what F_i-C [F_i-who persuade whom to buy ____]

step three: C attracts the other wh-phrase

what whom F_i-C [F_i-who persuade ___ to buy ____]

Pronounced result: *What did who persuade whom to buy?*

The free choice of which *wh*-phrase moves overtly in ternary questions with *wh*₁-in-situ is thus the same fact, in essence, as the free choice of which *wh*-phrase moves *second* in Bulgarian overt phrasal movement. (In addition, (84a-b) has already displayed examples in Bulgarian of the derivation now being considered in English — examples in which the first instance of *wh*-movement is feature movement in a Bulgarian multiple

question.) We thus have support for the hypothesis that in apparent Superiority violations in English, the first instance of *wh*-movement is actually feature movement — specifically, feature movement that satisfies AC.⁵⁶ Thus, apparent exceptions to the Superiority Effect, when examined more closely, are really no exceptions at all.

This evidence also bears on an issue that we have not discussed so far. We set for ourselves the problem of explaining apparent violations of AC. Our proposal has the interesting consequence that these violations are merely apparent. AC is being satisfied by an instance of *wh*-movement that had not previously been postulated. (The alternative involving "invisibility" also had this property.) We might have chosen to consider another possibility: that apparent violations of AC really do violate this constraint. One approach of this kind might use the logic of *Optimality Theory* (Prince and Smolensky 1993) in which certain constraints on linguistic form can be violated when this is the only way to satisfy other, more highly valued (highly ranked) constraints. If the apparent Superiority violations discussed here were actual violations of AC, one would not expect to discover the evidence we have uncovered that these violations are merely apparent.⁵⁷ In particular, one would be surprised to see AC violated in ternary questions, with no apparent pressure on output forms to minimize this violation by preferring overt movement of the second closest *wh*-phrase in cases where the first closest cannot (for some reason) move. Consequently, this set of apparent violations does not provide support for the view that constraints on movement are ranked and violable in an OT fashion.

This observation is of interest, since proposals that tend in the opposite direction (for different sets of data) have been advanced in a number of recent studies, including Grimshaw (1997) and many papers in Barbosa et al. (1998). In Pesetsky (1997; 1998), I suggested that those OT interactions that are visible in syntactic phenomena actually reflect the way in which phonology interprets syntax (e.g. via the pronunciation principles discussed above), and speculated that principles governing movement might not be ranked and violable in an OT fashion. The proposal about apparent Superiority violations presented here removes a case that might have seemed to support the opposing view.⁵⁸ As far as we can tell, except for the effects of the PMC, AC is inviolable. In addition, we can draw a methodological moral: exceptions and violations are not necessarily what they appear to be. More takes place in syntax than meets the ear.⁵⁹

4.5. Interim summary

The central argument for our proposal concerning feature movement comes from the fact that the peculiarities of English multiple questions turn out to faithfully reproduce the peculiarities of Bulgarian multiple questions if the feature-movement proposal is assumed. The only difference we must posit between Bulgarian and English multiple questions concerns the pronunciation of *wh*-phrase movement structures (i.e. how many instances of *wh*-phrase movement to a given C_{m-spec} are overt; see (56)-(57). Otherwise, the syntax of multiple questions is identical in the two languages. We are now in a position to summarize this result. The relevant example numbers are given after each comment below:

The Multiple Specifier requirement of C_{m-spec} :

In a non-D-linked question, at least two *wh*-phrases must be attracted by C_{m-spec} .

Bulgarian: In a non-D-linked binary question, at least two *wh*-phrases move overtly. This can be seen easily in the distribution of overt *wh*-fronting. (35), (40), (41)

English: In a non-D-linked binary question, *wh*-feature movement is not an option for either *wh*-phrase. This cannot be viewed directly (though ACD teaches us that the in situ *wh* may undergo covert *wh*-phrase movement), but is crucial to the distribution of Superiority Effects, as indicated below.

The Superiority effect:

In a non-D-linked binary question, where the Multiple Specifier requirement forces both *wh*-phrases to undergo *wh*-phrase movement, *wh*₁ moves first.

Bulgarian: In a non-D-linked binary question, both *wh*-phrases are overtly fronted, and *wh*₁ is leftmost (given the "tucking in" property enforced by Shortest Move). (43)-(44), (83)

English: In a non-D-linked binary question, one *wh*-phrase is overtly fronted; the other is covertly fronted. *Wh*₁ is the overtly fronted *wh*. (26)-(27)

The D-linked exception to the Multiple Specifier requirement:

In a D-linked question, there is no requirement that at least two *wh*-phrases must be attracted by C_{m-spec} .

This means that *wh*₁ might undergo feature movement.

Bulgarian: In a D-linked question, it is not necessary that at least two *wh*-phrases move overtly. (80)

English: In a D-linked binary question, *wh*-feature movement is an option for either *wh*-phrase. (76)

The absence of the Superiority effect in questions with D-linking:

Bulgarian: In a D-linked question, *wh*₁ may remain in situ. (81)

English: In a D-linked binary question, *wh*₁ may remain in situ. (31)-(32)

[ACD shows that it does not undergo covert phrasal movement. (60)- (61)]

The absence of Superiority effects with non-binary questions:

When a multiple question contains more than two *wh*-phrases, *wh*₁ may undergo *wh*-feature movement, so long as at least two *wh*-phrases undergo *wh*-phrase movement.

Bulgarian: In a ternary question, *wh*₁ may be pronounced in situ. (83)-(84)

English: In a ternary question, *wh*₁ may be pronounced in situ. (33)/(82)

The Principle of Minimal Compliance and AC:

Once a first instance of *wh*-movement to a given C has obeyed AC, subsequent instances need not obey AC.

Bulgarian: In a ternary question, in which *wh*₁ has moved overtly and is the leftmost moved *wh*-phrase, the order of subsequent *wh*-phrases is free. (46)-(48) [Also: when *wh*₁ undergoes feature movement and is pronounced in situ, the choice of leftmost overtly moved *wh* is free. (84a-b)]

English: When *wh*₁ undergoes feature movement and is pronounced in situ, the choice of overtly moved *wh* is free. (88)-(89).

The next two sections take up (inconclusively) some loose ends left untied. In particular, we need to examine a few possible competitors to the hypothesis of "feature movement", to note their advantages and disadvantages – but, in the end, to show that our overall story is compatible with a variety of hypothesis concerning the true nature of what we have been calling "feature movement". We then move on to a deeper discussion of the specifier potential of interrogative complementizers.

4.6. Is "feature movement" really *feature* movement?

One possible competitor to the "feature movement" hypothesis is relevant to an issue left open in the discussion so far: the choice between phrasal and feature movement in a derivation. Consider what happens when C_{m-spec} has been introduced into a structure and proceeds to attract instances of the *wh*-feature to it. I have suggested that this attraction may proceed in either of two ways: either the feature itself is attracted from an expression that bears it, or else the largest phrase that bears the feature is copied as a specifier of C_{m-spec} . We have also seen that certain properties of the attractor — i.e. properties *external* to the laws of movement — dictate which type of attraction occurs in particular cases. External properties include the fact that C_{m-spec} requires two specifiers and the fact that C_{I-spec} requires one, facts that we will explore in greater detail in the next section.

Are there other, more general preferences for one type of movement over the other, rooted in the laws of movement, in addition to "external" filters that may require phrasal movement in certain circumstances? As we noted earlier, Chomsky (1995, chapter 4) answered this question in the affirmative. He suggested that attraction of the *smallest* possible unit — the feature itself — is a more economical operation than attraction of any larger unit. The basis for Chomsky's particular claim was the idea that feature movement is the proper reanalysis of covert phrasal movement — coupled with the claim (inherited from earlier work) that covert movement is the default. Our discussion throughout this study has argued against the first of these claims. As noted earlier, it is not clear whether there is any particular reason to hold to the second of the claims, though it can certainly be modified so as to be consistent with our discussion in this study.

Still, it is worth mentioning some variant proposals might be equally appealing in other ways. For example, let us start with the observation that words and phrases bear grammatical features by virtue of their introduction as a subconstituent of these words and phrases. In the case of an English *wh*-word, it is not implausible to imagine that the *wh*-feature is introduced by means of the *wh*-morpheme /h^w/ with which these *wh*-words begin. (The fact that two English "*wh*-words" actually begin with pronounced /h/ is plausibly a reflex of the phonological processes mentioned below.) In several cases, the morpheme that follows /h^w/ also appears in a demonstrative whose initial morpheme is /D/.⁶⁰

(91) **The English *wh* morpheme**

h ^w ʌt	'what'	(many speakers lose aspiration)	cf. <i>that</i>
h ^w en	'when'	"	cf. <i>then</i>
h ^w eyr	'where'	"	cf. <i>there</i>
h ^w ay	'why'	"	
h ^w lā	'which'	"	
h ^w aw	'how'	(with dissimilative loss of rounding; cf. the absence of /k ^w aw../) ⁶¹	
h ^w	'who'	(with the rounding feature appearing on epenthetic vowel)	

The morpheme /h^W/ is cognate with /k/ in Bulgarian and other Slavic languages. Slavic *wh*-morpheme /k/ alternates rather regularly with demonstrative /s/, much as English /h^W/ alternates with demonstrative /D/.

Other, unrelated languages show similar patterns, e.g. Japanese:

(92) **The Japanese *wh*-morpheme**

dore 'which one'	cf. kore	'this one'	sore	are	'that one' ⁶²
dono 'which'	cf. kono	'this'	sono	ano	'that'
dotira 'whither'	cf. kotira	'this direction'	sotira	atira	'that direction'
doo 'how'	cf. koo	'this manner'	soo	—	'that manner'
dare 'who'	cf. kare	'he'	—	—	

If the "*wh*" morpheme in languages like English, Bulgarian and Japanese has no grammatical properties besides its role as the bearer of the *wh*-feature, *wh*-feature movement might simply be the attraction of this morpheme from inside the word that contains it (cf. also Watanabe 1992 and Tsai 1994's proposals along similar lines). One might ask whether other instances of "feature movement" might be better viewed as "morpheme movement" — though obvious questions arise where a phonologically distinct morpheme cannot be identified, e.g. nominative case in English. If this proposal is correct, then feature movement is simply the copying of a morphosyntactic constituent, just like more familiar examples of movement.⁶³ When *wh*-feature movement takes place from a phrase like *which book* or *who*, it is the *wh*-morpheme that is being copied to C (perhaps adjoining to it). The movement is *covert* in that the morpheme remains pronounced in its original position. This may simply reflect its status as a bound morpheme. If this view is correct, then "feature movement" is movement of the *smallest* unit that bears the feature in question.⁶⁴

AC raises some obvious questions if "feature movement" is morpheme movement. The *wh*-feature present in *wh*-words is also present in larger "*wh*-phrases" that contain these words, at least insofar as attraction of *wh* may attract a phrase containing the morpheme — the "phrasal movement" that we have discussed throughout.⁶⁵ Crucially, it is the distance from C_{*m-spec*} to the maximal *wh*-phrase — not the

distance from C_{m-spec} to the *wh*-morpheme, that counts for the calculation of "closeness" relevant to AC, even when the maximal phrase, rather than the minimal unit, is copied:

(93) **Closeness**

A feature α is *closer* to K than β iff K c-commands an occurrence of α , and β asymmetrically c-commands an occurrence of α .

Phrasal movement can now be seen as the copying of the closest bearer of the relevant feature. feature movement, by contrast, can now be seen as copying the smallest bearer of a feature. As suggested by Norvin Richards in unpublished work, and as mentioned earlier in this study, both conditions fit the spirit of "economy". Either is a plausible candidate for a genuine principle of grammar. Quite possibly, movement is possible if it meets either of these two criteria.⁶⁶ Putting the matter in procedural terms, the target scans down the tree for the closest occurrence of the feature it attracts. When it finds an occurrence on a constituent, it has a choice: it either copies that constituent or else copies the smallest subconstituent that contains the feature. Excluded is copying of an intermediate constituent that bears the relevant feature, since it is neither the closest nor the smallest expression of the relevant feature.

It is entirely possible, of course, that we will learn of other considerations that restrict the choice between feature (morpheme) movement and phrasal movement. For example, feature-movement for case purposes may be intrinsically preferred over phrasal movement. We were able to demonstrate the properties of feature movement successfully with the *there*-construction in part because (evidently) covert phrasal movement to T is impossible in this construction. In our discussion of ACD in multiple questions, we learned that covert phrasal movement to an VP-external object Case position is impossible. (Overt phrasal movement to subject position is the consequence of other considerations, often grouped under the rubric "Extended Projection Principle".) Conceivably, these two observations teach us that preferences for feature movement over phrasal movement exist. The question before us would be whether this preference is limited to movement for case purposes, or whether the preference is more general, in the spirit of Chomsky's (1995, chapter. 4) proposals. For now, I take these to be open questions.

4.7. Is "feature movement" really feature *movement*?

Another open question concerns island effects. When we first introduced Richards' (1997) Principle of Minimal Compliance, we mentioned Richards' observations concerning the interaction of the PMC with island constraints on *wh*-movement. If the first instance of *wh*-movement in a Bulgarian multiple question satisfies subjacency, the subjacency effect weakens or disappears for subsequent instances of *wh*-movement. The first instance of movement pays a "subjacency tax" valid for the subsequent instances of movement. We repeat the relevant example below:

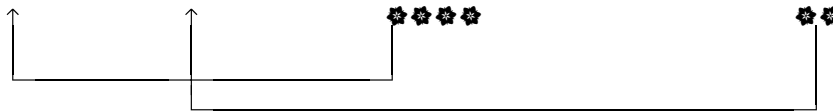
(94) PMC improves subjacency violations.

a. *Koja kniga otreãe senatorãt [mãlvata ãe iska da zabrani ___]? [(52)]

which book denied the-senator the-rumor that wanted to ban

'Which book did the senator deny the rumor that he wanted to ban?'

b. ?Koj senator koja kniga otreãe ___ [mãlvata ãe iska da zabrani ___]?



which senator which book denied the-rumor that wanted to ban

'Which senator denied the rumor that he wanted to ban which book?'

(Roumyana Izvorski, Ani Petkova, Roumyana Slabakova, p.c. to Richards 1977)

Our subsequent discussion of *wh*-feature movement raises the question of whether *wh*-feature movement, like overt phrasal movement in Bulgarian, pays a "subjacency tax", just as it pays an "AC tax". The answer seems to be that it does not:

- (95) ***Wh*-feature movement does not pay a "subjacency tax".**
- a. ??Which book did the senator deny the rumor that he wanted to ban.
 - b. ??Which book did which senator deny the rumor that he wanted to ban.
- (96)
- a. *What would you be upset if the reporter revealed __ about whom?
 - b. *What would who be upset if the reporter revealed __ about whom?
- (97)
- a. *What subject do you want to applaud the person that gave __ to whom?
 - b. *What does who want to applaud the person that gave __ to whom?

One can imagine a number of reasons why *wh*-feature movement might not pay a subjacency tax — in contrast to phrasal movement. For example: *wh*-feature movement is *covert*, in that pronunciation continues to target the *wh*-morpheme in its lower, original position. Quite generally, island violations do not cause deviance when the trace position receives a pronunciation. (This was the explanation for the ameliorating effect of resumptive pronouns offered by Perlmutter 1972; see also Pesetsky 1997, 1998.) Conceivably, movement that fails to "leave a gap" does not pay a subjacency *tax* because it does not trigger the subjacency *effect*. Likewise, if feature movement is really *morpheme* movement, its covert status may explain the absence of a subjacency effect and the absence of subjacency-tax paying.

Another possibility is suggested by the research of Ochi (1998), who develops earlier ideas by Takahashi (1994). Recall from section 1.1 that Chomsky (1995, chapter 4) decomposed phrasal movement into two distinct operations: Attract F (feature attraction) and "generalized pied piping" (which forces the copying of phrasal material). Ochi (1998) argues that it is the "generalized pied piping" component of phrasal movement that is responsible for many of the island effects often grouped under the rubric of the Subjacency Condition. He suggests that these island effects arise from a conflict between a requirement that generalized pied piping copy a phrase to the closest possible landing site and other, conflicting requirements that prevent movement from targeting these very positions. As Ochi points out, this proposal entails that pure feature movement should not show these island effects at all — to which we may now add that pure feature

movement should also not pay any "tax" that allows later instances of phrasal movement to escape these island effects.⁶⁷

As a final possibility, we might take the contrast between (94) and (95)-(97) as a sign of a deeper inaccuracy in the picture we have painted so far. Phrasal movement obeys both AC and subjacency, while what we have called "feature movement" obeys only AC. The fact that phrasal movement obeys a superset of the conditions on "feature movement" might be telling us that what we have been calling "feature movement", though it might be a subcomponent of movement, is something quite different from the operations traditionally called "movement". This is, in fact, the proposal advanced by Chomsky (1998). He suggests that phenomena that he previously identified (Chomsky 1995) as instances of feature movement are actually the result of a simpler operation called *Agree*. *Agree* merely establishes a link between an attracting head and the features that it seeks, without the copying operation characteristic of movement. Movement itself is simply *Agree* plus this copying.⁶⁸

To keep the discussion throughout this work as clear and simple as possible, I will not explore any further the consequences of choosing one or another variant of the Feature Movement proposal. What we think we know is this: there is an operation *O* that has certain manifest properties:

How *O* is like phrasal movement

1. *O* is feature-driven.
2. *O* obeys Attract Closest.
3. *O* interacts with phrasal movement for the purposes of the PMC as applied to Attract Closest.

How *O* is unlike phrasal movement

1. *O* does not copy and delete phrases.
2. *O* does not display subjacency effects.
3. *O* does not interact with phrasal movement for the purposes of the PMC as applied to subjacency effects.

Other questions about this operation cannot be answered using the facts at our disposal.

On the other hand, there is an entirely different set of facts that reinforces the conclusions we have already reached and the particular analysis of apparent Superiority violations that has been presented. By reinforcing our conclusions with new evidence, we support the proposal advanced in this monograph—incomplete though it may be in certain respects. We turn to this new source of evidence in the next section.

5. The Intervention Effect and the Typology of Interrogative Complementizers

One of the linchpins of our hypothesis has been the idea that multiple questions are introduced by a complementizer (C_{m-spec}) that requires *wh*-movement to establish a multiple-specifier configuration. This property of C_{m-spec} helped us understand why only binary questions show the Superiority Effect in English. The argument that the multiple-specifier requirement is not a *deus ex machine* came from Bulgarian, where failure to move at least two *wh*-phrases in a multiple question produces sharp unacceptability (except in a D-linked environment), which contrasts with mild deviance produced when additional *wh*-phrases are not fronted. Still, it would be useful to have converging evidence that the number of specifiers associated with C_{m-spec} is relevant to the detection of a Superiority Effect.

This section provides converging evidence of just this sort, and in the process sheds light on other aspects of our story. In particular, we will tie a loose end from the opening sections of this study. In section 1.3, we saw three situations in which an otherwise expected Superiority Effect is not found. Two of these situations have been dealt with: English multiple questions with D-linking and multiple questions with more than two *wh*-phrases. The third case has not been discussed yet. This is the case of German, where, as we saw in example (34), even binary questions fail to display a Superiority Effect. We will turn to German after discussing some additional data from English. More interesting still is the fact that English and German will turn out to fit into a simple typology of *wh*-specifier interactions — of which Japanese will exemplify the one language type undiscussed so far. This will be the topic of the final section of this study.

The English data I will discuss are partly original to this study, but are modeled in crucial ways on German paradigms investigated by Beck (1996), whose work inspired this section. After considering the crucial English cases, I will turn to Beck's German discoveries.

5.1. The Intervention Effect in English multiple questions

In the previous section, we examined several apparent exceptions to the Superiority Effect — acceptable multiple questions with *wh* 1-in-situ. The impossibility of ACD in *wh* 1 taught us at a minimum that there is something special about *wh* 1 in such cases. A phenomenon first noted in unpublished work by Kiss (1986) shows another property of *wh* 1-in-situ that teaches us a similar lesson. In apparent counterexamples to the Superiority Effect, something special happens when a scope taking element such as negation intervenes between *wh* 1 and the C_{m-spec} with which it is associated. For many speakers, at least as a first reaction, such examples are completely unacceptable. For other speakers (and perhaps for all speakers when the possibility is pointed out to them), the examples are acceptable so long as they receive a "single pair" rather than a pair-list reading.

A representative paradigm is displayed in (98). Examples (98a) and (98b) are experimental controls — normal multiple questions in which *wh* 1 has undergone overt phrasal movement. Nothing special happens when negation intervenes between C_{m-spec} and *wh* 2-in-situ in such cases. Example (98c) shows *wh* 1-in-situ (i.e. an apparent Superiority violation). Here, if we are correct, the first instance of *wh*-movement is feature movement of *which person*. In this example, negation has not moved with the auxiliary verb to C, and thus remains lower than *wh* 1-in-situ. Once again, not surprisingly, we note no special effect. Example (98d) is the example that contrasts crucially. It is identical to (98c), except that the clitic form of negation moves to C_{m-spec} with the auxiliary verb. As a consequence of this movement, negation intervenes between *wh* 1 and C_{m-spec} . Example (98d) does produce special judgments:

- (98) **Intervention effect with *not*** [Kiss, 1986; Hornstein 1995]
- a. Which person ___ did not read which book?
 - b. Which person ___ didn't read which book?
 - c. Which book did which person not read ___?
 - d. *Which book didn't which person read ___?

(cf. also *Which book did which person read ___?*)

Though, as the asterisk indicates, many speakers regard (98d) as completely unacceptable, the marginal acceptability of the single-pair reading in this and similar examples will be noted by superscripting and underscoring the judgment indication. The single-pair reading for (98d) is one in which it asks for the name of a single person and a single book, such that the person did not read the book. For example, suppose you pass a classroom and hear the teacher yelling. You know that this teacher only yells at his class when a student has failed to read his book for that day. In such a context (and only in this type of context), some speakers will accept a question like *Which book didn't which person read today?* For the time being, I will put aside the single-pair readings of examples like these, and confine my presentation of judgments on multiple questions to the pair-list reading.

I will call the type of effect seen in (98d) an "Intervention Effect", following terminology of Hagstrom (1998). The Intervention Effect is less sharp when *wh*₁ is not a subject, but much the same judgment obtains:⁶⁹

- (99) **Intervention effect with *not* — non subjects**
- a. Which issue should I not discuss ___ with which diplomat?
 - b. ??Which diplomat should I not discuss which issue with ___?

(cf. *Which diplomat should I discuss which issue with ___*)

The same judgment can be produced when other negative expressions and another downward-entailing phrases intervene between *wh*₁ and *C_{m-spec}*, showing that elements other than clausal negation function as interveners:

(100) **Intervention effect with *no one***

a. Which book did no one give __ to which student?

b. ??Which student did no one give which book to __ ?

(cf. *Which student did Mary give which book to __?*)

(101) **Intervention effect with *never***

a. Which topic did he claim which student would never talk about __?

b. ??Which topic did he never claim which student would talk about __?

(cf. *Which student did he never claim __ would talk about which topic?*)

(102) **Intervention effect with *very few***

a. Which picture did very few children want to show __ to which teacher?

b. Which teacher did very few children want to show which picture to __?

(cf. *Which teacher did she want to show which picture to __?*)

(103) **Intervention effect with *only***

a. Which girl did only Mary introduce __ to which boy.

b. ??Which boy did only Mary introduce which girl to __.

(cf. *Which girl did only Mary introduce __ to which boy?*)

The Intervention Effect is not a property of "overt syntax", but cares instead about the scope (LF position) of the intervener. For example, in (104), which should be compared to (103), the acceptability of the pair-list reading re-emerges so long as *only Mary* receives matrix scope. Some degree of focus stress on *only Mary* facilitates this reading:

(104) **Intervention effect disappears with wide scope of *only***

Sue asked which boy only Mary introduced which girl to ___.

(i.e. *Mary is the only person such that Sue asked which boy this person introduced which girl to.*)

Although downward entailing elements produce the most easily detected Intervention Effects, it seems that other quantifiers also produce a similar effect.⁷⁰ Consider, for example, a non-multiple question with a universal quantifier, such as (105):

(105) ***Wh*-quantifier scope ambiguity**

Which newspaper did everyone write to ___ about this book?

This type of question displays a well-known ambiguity (Engdahl 1980; May 1985) between a "pair-list" or "distributive" reading (for which a well-formed answer provides a list of people paired with the newspapers they wrote to) and a single-answer reading (which asks for the name of a single newspaper such that everyone wrote to it).⁷¹ While controversies abound concerning the syntax and semantics of the distributive reading (cf. Chierchia 1993, for example), one common and plausible conjecture derives the distributive reading from an LF in which the universal quantifier takes wide scope over the *wh*-expression. If we imagine that wide scope for *everyone* results from movement to a position above *which newspaper* (CP-adjunction, perhaps), the LF for the distributive reading of (105) is approximately (106):

(106) **LF for the distributive reading of (105)**

everyone_x [which newspaper_y C_{I-spec} x write to y about this book]

The ambiguity survives in multiple-*wh* questions like (107):

(107) Which newspaper did everyone write to ___ about which book? (ambiguous)

On a wide-scope reading for *everyone*, (107) asks for triplet answers which, for each person picked out by *everyone*, gives the newspaper and the book such that person wrote to the newspaper about the book. That is, a possible answering pattern might be (108):

(108) **Wide-scope answering pattern**

Bill wrote to the *New York Times* about book X, Mary wrote to the *Boston Globe* about book Y, and Tom wrote to the *Maquoketa Sentinel* about book Z.

On a narrow-scope reading for *everyone*, (107) asks for pair-list answers of the form exemplified by (109):

(109) **Narrow-scope answering pattern**

Everyone wrote to the *New York Times* about book X, everyone wrote to the *Boston Globe* about book Y, and everyone wrote to the *Maquoketa Sentinel* about book Z.

Given the ambiguity of (107), the non-ambiguity of (110) is striking. Example (110) is the same as (107), except that *wh*₂ has undergone overt phrasal movement instead of *wh*₁. Consequently, *everyone* intervenes between *wh*₁ and *C_{m-spec}*. If all things were equal, (110) might be expected to allow the same two answering patterns as (107). Though it takes some work to see, I believe that (110) lacks the reading that would yield the narrow-scope answering pattern in (109). Unless the two *wh*-phrases are accorded a single-pair reading, a question like (110) can only be asked in expectation of an answer like (108). A "single-pair" reading for the two *wh*-phrases is also available in (110) — that is, a reading that invites a simple answer like *Everyone wrote to the New York Times about book X*. This is not surprising, since the Intervention Effect is not found with single-pair readings.⁷²

(110) **Intervention effect disappears with wide scope of *every***

Which book did everyone write to which newspaper about ___? (unambiguous)

This fact seems to reinforce the observation that the Intervention Effect is a constraint on Logical Form that examines the scope position of the intervener.

This conclusion is supported by other, related tests. An expression like clausal negation whose scope is fixed at roughly its surface position produces the unacceptability already noted when it intervenes between *wh* 1-in-situ and C_{m-spec} . There are also certain types of expressions that do allow LF scope wider than surface scope that still do not allow scope wider than *wh* within the confines of a question. Such expressions, as predicted, also produce unacceptability rather than disambiguation when inserted between *wh* 1-in-situ and C_{m-spec} . These expressions include *no one* and *only NP*, which we have already examined, and also *almost every NP* (e.g. *almost everyone*), which contrasts minimally with the expression *everyone* that we have just studied.⁷³ Assigning wide scope to these expressions at the outer boundary of a question seems to be impossible. This scope assignment may run afoul of a something like (111), with the consequences documented in (112):⁷⁴

(111) **Unaskable questions**

A clause interpreted as a question may not request anything less than a full answer.

(112) **Examples of unaskable questions**

a. Which newspaper did almost everyone write to ___ about this book?

cannot be used to express the following request for information:

"Give me an almost complete list of people paired with the newspapers they wrote to about this book."

b. Which book did no one give ___ to John?

cannot be used to express the following request for information:

"Don't tell me which book anyone gave to John"⁷⁵

c. Which teacher did very few children want to visit ___?

cannot be used to express the following request for information:

"There are a number of children who wanted to visit a variety of teachers. Answer the question "what teacher did he/she want to visit?" for a very small number of these children (and don't bother providing the rest of the answers). "

d. Which girl did only Mary introduce ___ to John?

cannot be used to express the following request for information:

"I know that several people introduced girls to John. Tell me which girl Mary introduced to him, but don't tell me which girls other people introduced to him."

Since the only way around the Intervention Effect in examples like (100b) and (103b) involves wide scope for the intervener that violates (111), the examples cannot be saved. Exactly the same explanation explains the contrast between (110) with *every* and a comparable example with *almost every*

(113) **Intervention effect with *almost every***

??Which book did almost everyone write to which newspaper about ___?

As predicted, it is difficult or impossible to understand (113) as any sort of pair-list multiple question. To the extent that it is acceptable at all, it is only a request for a single-pair answer (e.g. *Almost everyone wrote to this newspaper about that book.*).

Quantifier float provides a similar demonstration. Floated quantifiers are restricted in scope to their surface position. For example, while (114a) has a reading in which *each student* has wider scope than *some teacher*, (114b) does not:

(114) **Frozen scope with quantifier float**

- a. At least one teacher made each student sing the national anthem.
- b. At least one teacher made the students each sing the national anthem.

The same observation holds for the relative scope of *each* and *wh*. While (115a) may be a request for an answer that provides adult-kid-book triplets, (115b) is at best a request for adult-book pairs, such that each kid in the group will persuade the adult to read the book.⁷⁶

(115) **Frozen scope with quantifier float in multiple questions**

- a. Tell me which adult each kid will try to persuade ___ to read which book.
- b. Tell me which adult the kids will each try to persuade ___ to read which book.

If we now alter these examples by overtly fronting *wh*₂ (*which book*) rather than *wh*₁, the impossibility of wide-scope when *each* is floated should produce an Intervention Effect from which the only (marginal) escape is the single-pair reading. This seems correct:

(116) **Intervention effect with quantifier float**

- a. Tell me which book each kid will try to persuade which adult to read ___.
- b. ??Tell me which book the kids will each try to persuade which adult to read ___.

The paradigms I have discussed so far seem to support the following (interim) characterization of the Intervention Effect as it shows up in English:

(117) **Intervention Effect in English**

A scope-bearing element (esp. quantifiers and negation) may not intervene at LF between *wh* 1-in-situ and the C_{m-spec} with which it is associated.

The data presented in support of this characterization come from instances of *wh* 1-in-situ in binary questions with D-linking. Our other case of *wh* 1-in-situ — questions with more than two *wh*-phrases — appears to show the same pattern of Intervention Effects. This is particularly evident when negation is the intervener. Taking as baseline a triple-*wh* question without negation, but with *wh* 1-in-situ, it is fairly easy to see that negation intervening between *wh* 1 and C_{m-spec} (as in (118b) and (119b)) has an effect not found when negation is lower than *wh* 1 (as in (118a) and (119a)). The effect is the familiar one: the example is either unacceptable or else acceptable with only a single-pair reading.

(118) **Intervention Effect with *not* in non-binary, non-D-linked multiple questions**

baseline: What did who give __ to whom?

a. What did who not give __ to whom?

b. *What didn't who give __ to whom?

(119) *baseline*: What did Bill persuade who to give __ to whom?

a. What did Bill persuade who not to give __ to whom.

b. ??What did Bill not persuade who to give __ to whom

The other interveners discussed above seem to behave similarly. Because questions that include three *wh*-phrases plus a quantifier are complex, judgments on these examples do not come easily. Thoughtful consideration is necessary:

(120) **Intervention effect with *no one* in non-binary, non-D-linked multiple questions**

a. What did who say no one gave ___ to whom?

b. ~~??~~What did no one say who gave ___ to whom?

(121) **Intervention effect with *never* in non-binary, non-D-linked multiple questions**

a. What did he claim who would never talk about ___ with whom?

b. ~~??~~What topic did he never claim who would talk about ___ with whom?

(122) **Intervention effect with *very few* in non-binary, non-D-linked multiple questions**

a. Who did very few people persuade ___ to give what to whom for Christmas?

a. ~~??~~What did very few people persuade who to give ___ to whom for Christmas?

(123) **Intervention effect with *only* in non-binary, non-D-linked multiple questions**

a. What did who persuade only Mary to buy ___ where.

b. ~~??~~What did only Mary persuade whom to buy ___ where.

5.2. The Intervention Effect in German separation constructions

The Intervention Effect in English is particularly interesting because it singles out *wh* 1-in-situ. As (118a) and (182a) show, no intervention effect is observed with other *wh*-phrases — even in questions that display *wh* 1-in-situ. Since we do not know for sure whether there is an overall preference for phrasal movement or for feature movement, the one English *wh*-phrase that we can be absolutely sure has undergone feature movement is *wh* 1-in-situ in apparent counterexamples to Superiority. Consequently, it is quite tempting to see the Intervention Effect as a diagnostic for some property that distinguishes *wh*-feature movement from *wh*-phrasal movement of the sort we have been examining.⁷⁷

One property that distinguishes *wh*-feature movement from *wh*-phrase movement in English is the fact that feature movement leaves the restriction on *wh*-quantification inside the clause, while phrasal

movement typically pied-pipes the restriction with the *wh*-phrase. It could be that "separation" of the sort seen in *wh*-feature movement is the cause of the Intervention Effect:⁷⁸

(124) **Intervention Effect (universal characterization)**

A semantic restriction on a quantifier (including *wh*) may not be separated from that quantifier by a scope-bearing element.

German provides interesting support for this view — support which will ultimately provide evidence for several aspects of our overall story. My discussion of German is a reworking of data and discussion from Beck (1996). I have added to her paradigms in several cases, and have truncated the presentation of data in others, but my debt to Beck's work should be clear throughout. I will not take a stand on the question of *why* a constraint like (124) should hold, but will limit my discussion to the truth of the generalization itself and its role in supporting the proposals about *wh*-movement presented here. For a view of intervention effects quite similar to (124), and a possible account, see Honcoop (1998, esp. p.19 and pp.81ff).

German displays several constructions in which phrasal *wh*-movement raises a portion of an argumental phrase overtly, leaving the remainder behind in the clause. We can call this family of constructions "separation". Among the phrases that can separate in this manner are phrases of the form *wh-word + partitive-PP*, *wh-word + adjective*, and *wh-word + all*. I will leave open the exact nature of *wh*-separation. It might involve an internal structure for the phrases in question that simply permits a subconstituent to count as a *wh*-phrase and extract accordingly. Alternatively, it might involve scrambling of expression that is to be stranded out of its *wh*-phrase, with phrasal *wh*-movement of the remnant. Either way, it is a construction that strands — inside its clause — material belonging to the restriction of a *wh*-phrase. Strikingly, the construction is subject to the Intervention Effect, as the paradigms in (125)-(126) make clear:

(125) **Who + among DP**

a. *No separation, no intervener*

[Wen von den Musikern] hat Hans __ getroffen?

whom of the musicians has Hans __ met

'Who among the musicians has Hans met?'

b. *Separation, no intervener*

Wen hat Hans [__ von den Musikern] getroffen?

whom has Hans of the musicians met

c. *No separation, intervener*

[Wen von den Musikern] hat keine Studentin getroffen?

whom of the musicians has no student met

'Who among the musicians has no student met?'

d. *Separation, intervener*

?Wen hat keine Studentin [__ von den Musikern] getroffen?

whom has no student of the musicians met

(126) **Who + all**

a. *No separation, no intervener*

[Wen alles] hat Hans ___ gesehen?

whom all has Hans seen

'Who all did Hans see?'

b. *Separation, no intervener*

Wen hat Hans [___ alles] gesehen?

whom has Hans all seen

c. *No separation, intervener*

[Wen alles] hat niemand ___ gesehen?

whom all has no one seen

'Who all did no one see?'

d. *Separation, intervener*

?Wem hat niemand [___ alles] gesehen?

whom has no one all seen

(127) **What + adjective**

[no separation not possible, for unknown reasons]

b. *Separation, no intervener*

Was hat Gretchen heute [__ schönes] gemacht?

what has Gretchen today nice done

d. *Separation, intervener*

?Was hat niemand heute [__ Schönes] gemacht?

what has no one today nice done

'What nice thing did no one do?'

Other facets of our discussion of English can be reproduced here as well. For example, (128) has only a distributive reading:

(128) **Distributive reading only**

Wen hat jeder __ alles gesehen.

whom has everyone all seen

'Who all did everyone see?'

The common behavior of German separation constructions and English *wh* 1-in-situ supports our discussion nicely.⁷⁹ If English *wh* 1-in-situ involves feature or morpheme movement (as argued above), it is actually simply an instance of "separation": a piece of the *wh*-phrase — in this case, just the *wh*-morpheme or feature — moves to the complementizer, stranding the restriction inside its clause. If we adopt the alternative that posits an agreement operation in these cases, we still see a separation of the interrogative C-system from the restriction of a *wh*-phrase. Consequently, even if this variant of our story about English

*wh*₁-in-situ is correct, these constructions should share the Intervention Effect with the more obvious instances of separation that we find in German.⁸⁰

In fact, one dialect of English actually displays a German-like separation construction that seems to show the Intervention Effect. As described by McCloskey (1997), the West Ulster dialect of Irish English shares with many colloquial registers a series of *wh*-phrases modified by *all*: *who all*, *what all*, etc. The semantic effect of the morpheme *all* is roughly to presuppose that the answer will be a plurality. The West Ulster dialect, unlike others, allows separation of *all* from its *wh*-phrase:

(129) **West Ulster *what all* separation**

- a. What all did Mary get on her birthday?
- b. What did Mary get __ all on her birthday?

When this occurs, an Intervention Effect seems to be found:⁸¹

(130) **West Ulster Intervention Effect with separation**

- a. ??What did Mary not buy __ all up the town?
- b. *What did only Mary get __ all on her birthday?
- c. What did everybody get __ all on Christmas morning? [distrib. reading only]
- d. *What did almost everybody get __ all on Christmas morning?
- e. *What did very few people get __ all for Christmas?

The observation must be treated with some caution, since even the "unseparated" variants of (130) are felt by speakers to be strange (e.g. *What all did Mary not buy up the town?.*). Nonetheless, the separated versions are said to be detectably worse.

In any case, if we are right about the unity of the Intervention Effect in English multiple questions and German separation constructions, then we must accept a characterization like (124) instead of the characterization offered by Beck (1996). Beck offers an interpretation of the Intervention Effect in German

separation constructions according to which the effect is an island condition on traces formed by covert phrasal movement — in particular, a condition on a rule which moves the restriction to its quantifier in separation constructions. (A similar formulation is given by Hasegawa 1994, who treats it as an LF-particular case of the Minimal Link Condition.) In Beck's formulation, "Quantifier-Induced Barriers" (or *QUIBs*) cannot be crossed by traces that created solely "at LF" (i.e. by covert phrasal movement):

(131) **Beck's proposals [Beck 1996, 39]**

a. **Quantification-Induced Barrier (QUIB)**

The first node that dominates a quantifier, its restriction, and its nuclear scope is a Quantifier-Induced Barrier.

b. **Minimal Quantified Structure Constraint**

If an LF trace t is dominated by a QUIB Q , then the binder of t must also be dominated by Q .

Our discussion of ACD taught us that the stranded material in English *wh*-feature movement does *not* undergo covert phrasal movement of any sort. Consequently, if the Intervention Effect in German separation constructions is the same phenomenon we have examined in English, the Minimal Quantified Structure Constraint as stated in (131b) is probably incorrect. Instead, something like (124) must be true.⁸²

5.3. The Intervention Effect in German multiple questions

If the Intervention Effect diagnoses the presence of "restriction material" stranded within an interrogative clause, it can be used as a probe to discover instances of feature movement in languages and constructions where our previous test, ACD, is unavailable. German, for example, lacks VP-ellipsis and seems to lack other instances of anaphora with the right set of properties.⁸³ If we use the Intervention Effect as a probe for feature movement in German multiple questions, we discover a notable difference between English and German. I will argue that a partial explanation for this difference is available to us, and that this explanation in turn supports the overall architecture of our story.

In German, unlike English, the Intervention Effect in multiple questions is not limited to cases of *wh* 1-in-situ. Instead, *no wh*-in-situ may be separated from C by the sorts of elements that produce the Intervention Effect.⁸⁴ (I will not prejudge the nature of the complementizer by calling it *C_{m-spec}*, for reasons that will become clear shortly.) Example (132), for example, shows a multiple question in which nominative *wh* 1-in-situ has undergone overt phrasal *wh*-movement and accusative *wh* 2 has not. When the dative NP that intervenes between *wh* 2 and its C is an ordinary NP like *dem Karl*, the result is fully acceptable. When the intervener is a negative quantifier, as seen in (133b), the result is unacceptable in a quite familiar fashion. The result is unacceptable unless (for some speakers) a single-pair answer is expected.⁸⁵ That this is an "intervention" effect is made clear by the contrast between (133a) and (133b). These examples differ only in that the accusative *wh*-phrase appears to the left of the dative in (133a), presumably as a consequence of scrambling:

(132) **Baseline: no intervener, no intervention effect**

Welche Kinder haben __ dem Karl welche Bilder zeigen wollen?
 which children-NOM have the Karl-DAT which pictures-ACC show wanted
 'Which children wanted to show Karl which pictures?'

(133) **Intervention Effect (German) with *no one***

- a. Welche Kinder haben welche Bilder niemandem zeigen wollen?
 which children-NOM have which pictures-ACC no one-DAT show wanted
 'Which children wanted to show nobody which pictures?' (*scrambling of ACC*)
- b. ^{??}Welche Kinder haben niemandem welche Bilder zeigen wollen?
 which children-NOM have no one-DAT which pictures-ACC show wanted

The set of interveners that produce the effect appears to be the set now familiar to us. Although I have not displayed the relevant examples, scrambling the accusative *wh* 2 over the dative NP eliminates the Intervention Effect. Examples (134)-(135) display downward-entailing interveners:

(134) **Intervention Effect (German) with *very few***

??Welche Kinder haben sehr wenigen Lehrern welche Bilder zeigen wollen?
which children-NOM have very few teachers-DAT which pictures-ACC show wanted
'Which children wanted to show very few teachers which pictures?'

(135) **Intervention Effect (German) with *only***

??Wer hat nur dem Karl welche Bücher gegeben?
who-NOM has only the Karl-DAT which books-ACC given
'Who gave only Karl which books?'

Intervention by *jeder* 'everyone' is possible, so long as *jeder* is assigned wide scope:

(136) **Intervention Effect (German) with *everyone***

Wen hat jeder wo gesehen?
whom has everyone where seen
'Who did everyone see where?'

(wide-scope for *every* or single-pair narrow scope; otherwise "??")

Since wide scope is impossible for *fast jeder* 'almost everyone', there is no reading (except perhaps the single-pair reading) that by-passes the Intervention Effect:

(137) **Intervention effect (German) with *almost everyone***

??Wen hat fast jeder wo gesehen?
whom has almost everyone where seen
'Who did almost everyone see where?'

The question we wish to answer concerns the stark difference in the types of *wh*-in-situ subject to the effect. Why is only *wh*₁-in-situ subject to the effect in English, while all *wh*-in-situ are subject to the effect in German? If we are right about the character of the Intervention Effect, then the answer must be (138):

(138) **Movement in German Multiple Questions**

In a German multiple question, all *wh*-in-situ undergo *wh*-feature movement.

German multiple questions behave like separation constructions because, in a sense, they *are* separation constructions.⁸⁶

What property of German is responsible for (138)? In particular, what distinguishes German from Bulgarian and English, which allow multiple *wh*-phrase movement? Though one might try to answer this question in a number of ways,⁸⁷ I will explore the possibility that the difference arises from differences in the inventory of complementizers available to the two languages.

If our proposals are correct, multiple questions in English and Bulgarian boast a special complementizer which we have called C_{m-spec} . This complementizer, as discussed, has a *wh*-feature which can and must be deleted after attracting more than one instance of a corresponding *wh*-feature to it, either by means of phrasal movement or by means of feature movement. C_{m-spec} also had a specific syntactic requirement: it must be associated with more than one specifier. This fact was visible in Bulgarian in the patterns of overt *wh*-phrase movement, and formed an important component of our explanation for the Superiority Effect in English.

C_{m-spec} obviously differs from the complementizer found in single-*wh* questions, since these questions cannot satisfy a multiple-specifier requirement — at least not through *wh*-movement. For this reason, I suggested that single-*wh* questions are introduced by a complementizer that requires a single specifier: C_{1-spec} . The requirements of the two complementizers, first stated in (36) and (38), are repeated below. I omit, for now, the special case of D-linked questions:

(139) **Single-Specifier requirement of C_{I-spec}**

C_{I-spec} requires one *wh*-specifier.

(140) **Multiple-Specifier requirement of C_{m-spec}**

C_{m-spec} requires more than one *wh*-specifier.

Recall that both complementizers are taken to bear an uninterpretable feature, in something of the sense of Chomsky (1995, chapter 4). It is this feature which triggers the rule *Attract*, which leads to *wh*-movement.

In English and Bulgarian, it appears to be the case that a question is always introduced by the interrogative complementizer with the greatest "specifier potential" consistent with the instances of the *wh*-morpheme or feature that it can attract. Consequently, a question containing only one instance of the *wh*-feature within it will be introduced by C_{I-spec} , but a question containing two or more instances of the *wh*-feature must be introduced by C_{m-spec} . Otherwise, we would not detect the Superiority Effect in cases of *wh* 1-in-situ within binary sentences. The D-linking exception, as discussed earlier, is either an exception to (140) or else an exception to the rule of complementizer choice.

Suppose the rule of complementizer choice (informally stated as (141)) is universal:⁸⁸

(141) **Complementizer choice rule**

Use the complementizer with the maximum satisfiable specifier potential.⁸⁹

Let us now advance the hypothesis that the availability of a particular complementizer depends on the resources of the lexicon in a particular language. If this hypothesis is correct, the special facts of German *wh*-questions must be teaching us that the German lexicon has C_{I-spec} , but lacks C_{m-spec} . As a consequence of the presence in the lexicon of C_{I-spec} , German *wh*-questions show one instance of phrasal

wh-movement, which is overt, as in English. As a consequence of the absence in the lexicon of C_{m-spec} , any other instance of *wh*-movement in a multiple question must involve feature movement.

If the Intervention Effect found with pair-list multiple questions in German is due to the separation of the *wh*-feature from the restriction, it must be the case that the pair-list reading depends on *wh*-movement. That is, something like (142) must be true, presumably as a consequence of mechanisms of semantic interpretation which I will not explore here:

(142) **Syntax of the Pair-list reading for multiple questions**

A *wh*-expression participates in a pair-list reading within an interrogative clause only if its *wh*-feature has been attracted to the complementizer of the clause.

In English, (142) can be satisfied by multiple phrasal *wh*-movement, which does not involve separation of the *wh*-feature from its restriction. Consequently, English multiple questions do not have to show an Intervention Effect when a *wh*-in-situ is separated from the complementizer by a scope-bearing element (except when feature movement has in fact taken place, as in apparent Superiority violations). Since German *wh*-in-situ has only feature movement at its disposal as a means of satisfying (132), pair-list readings for multiple questions always invoke an Intervention Effect when a scope-bearing element intervenes between *wh*-in-situ and the complementizer. A single-pair reading, on this view, results from the failure of a *wh*-feature to undergo *wh*-movement, which in turn accounts for the lack of Intervention Effects with this reading.

The notion of "specifier potential", though implicit in the work on multiple specifiers initiated by Ura (1996), is novel, as is the idea that questions pick the complementizer with the greatest potential. It is therefore useful (just as it was when C_{m-spec} was first introduced) to show that these ideas are not *dei ex machinis*, by showing that they play a role in the explanation of other phenomena. In the present case, the Superiority Effect furnishes such a demonstration.

The fact that English binary questions show a Superiority Effect follows from the fact (now a consequence of (141)) that such questions involve C_{m-spec} — the interrogative complementizer that requires at least two specifiers. It is this fact that ruled out the spurious derivation of *What did who buy?* in (77), repeated below:

(143) **Spurious Derivation of an apparent exception to Superiority (English) (= (77))**

C_{m-spec} [*who bought what*]

step one: C_{m-spec} attracts the *wh*-feature of *who*

F_i -C [F_i -*who bought what*]

step two: C_{m-spec} attracts the *wh*-phrase *what*

what F_i -C [F_i -*who bought ____*]

pronounced result:

**What did who buy?*

If C_{m-spec} could be replaced with C_{I-spec} in English, the derivation would go through — but that would run afoul of the rules of complementizer choice which require the multiple question to be introduced by C_{m-spec} . On other hand, C_{I-spec} is exactly what German must use, if our proposals are correct. Consequently, German should show no Superiority Effect in such examples. The derivation in (144) should be just as possible as (145):

(144) **An apparent exception to Superiority: not spurious** (German)

C_{I-spec} [wer sah was]

who saw what

step one: C_{I-spec} attracts the *wh*-feature of *wer*

F_i-C_{I-spec} [F_i -wer sah was]

step two: C_{I-spec} attracts the *wh*-phrase *was*

pronounced result:

Was sah wer?

(145) **No apparent Superiority violation** (German)

C_{I-spec} [wer sah was]

who saw what

step one: C_{I-spec} attracts the *wh*-phrase of *wer*

wer [__ sah was]

step two: C_{I-spec} attracts the *wh*-feature of *was*

wer F_j -C [__ sah __]

pronounced result:

Wer sah was?

In fact, as is well-known since at least Haider (1986, 114), both outputs are acceptable. German speakers sometimes feel that (144) is "more natural" than (145), but I have never encountered a German speaker who felt that the output of (144) was deviant, nor do these examples require D-linking (but see Wiltschko 1997 and Grohmann 1998 for alternative views).⁹⁰ Both outputs have a pair-list interpretation.

As Grohmann (1998) notes, the absence of the Superiority Effect in German is not limited to cases in which wh_1 is a subject. Cases in which it is an object also fail to show the effect:

(146) **Apparent Superiority violation: wh_1 as object**

a. Wen hast du überredet was zu kaufen?

whom-ACC have you persuaded what to buy

'Who did you persuade to buy what?

b. Was hast du wen überredet zu kaufen?

(Grohmann 1998, ex. (30))

I take these examples to show that German generally lacks the Superiority Effect.⁹¹ This is an interesting result, since it is not otherwise obvious (independent of our proposals) why German should show greater freedom than English with respect to Superiority, but less freedom than English with respect to the Intervention Effect. Our proposals link these two facts.⁹²

Before proceeding further, I need to note that the status of Superiority Effects in German is rather controversial. A number of investigators, including Büring (1994) and Fanselow (1991; 1997) have pointed out that long-distance wh -movement shows contrasts that look like a Superiority effect. The examples in (147) are from Fanselow (1997):

(147) **A superiority effect with long-distance *wh*-movement?**

a. Wer __ glaubte, dass der Peter ihr wen vorstellte?
who-ACC believed that the Peter-NOM her-DAT whom-ACC introduced
'Who __ believed that Peter introduced her to whom?'

b. ?*Wen glaubte wer, dass der Peter ihr __ vorstellte?
whom -ACC believed who that the Peter-NOM her-DAT introduced
'Whom did who believe that Peter introduced __ to her?'

One way of explaining the contrast in (147), considered by Fanselow, relies on the idea that the Superiority Effect is active in German in just the same fashion as it is in English — i.e. the opposite of what we have concluded here. What allows apparent violations of Superiority in German, in this proposal, is the possibility of scrambling. In a multiple question in which the object undergoes overt *wh*-movement, while the subject remains in-situ, the object (on this view) first scrambles to a position higher than the subject. The object is now structurally higher than the subject, and consequently not only can but must be the first *wh*-element attracted by the interrogative complementizer:

(148) **The scrambling proposal (not adopted here)**

C [*wer sah was*]

who saw what

step one: *was* scrambles to a position higher than *were*

C [*was* [*wer sah t_{scrambling}*]]

step two: C attracts the *wh*-phrase *was*, since it is closer than *wer*

wer C [*__* [*wer sah t_{scrambling}*]]

pronounced result:

Was sah wer?

In order to overcome the Superiority Effect in (147) in the same fashion, the embedded object would have to scramble out of the embedded finite clause in order to land in a position higher than the matrix subject. As an independent fact about German, scrambling from finite clauses is forbidden:

(149) **No scrambling from finite clauses**

*...daß der Fritz den Josef_i glaubte, daß der Peter ihr —_i vorstellte.

that the Fritz-NOM the Josef-ACC believed that the Peter-NOM her-DAT introduced

'...that Fritz believed that Peter introduced Josef <scrambled> to her.'

(Fanselow 1997)

Consequently, (147b) could only show *wh*-movement directly from the embedded clause. Since the matrix subject is closer to the interrogative complementizer, *wh*-movement from the embedded clause, at least as the first instance of *wh*-movement, is forbidden by AC.

This is an attractive competitor to our proposal, since (like our proposal) it reduces the difference in Superiority Effects between German and English to an independent difference: here, the availability of scrambling in German and its unavailability in English. There is an alternative explanation for the contrast in (147), however, which does not rely on facts about scrambling. Example (147b) shows an instance of overt *wh*-movement from an embedded finite clause, while (147a) does not. As it happens, many speakers reject *wh*-movement from embedded finite clauses altogether. Even the many speakers who do not reject *wh*-movement from embedded finite clauses find such examples more difficult to accept than more local instances of *wh*-movement. This factor alone could explain the contrast between (147a) and (147b).⁹³

Still, if a *wh*-phrase may scramble before undergoing *wh*-movement, and if this scrambling operation allows objects to count as closest to the complementizer than subjects, we do have a source for apparent Superiority violations independent of the hypotheses presented here. Indeed, it is not established that scrambling to the left of the subject *can* feed overt *wh*-movement in the manner assumed in the scrambling proposal. For one thing, it seems to be impossible to scramble a *wh*-phrase that when that phrase does not undergo subsequent overt phrasal *wh*-movement, as noted by several researchers. Fanselow (1990), for example, observed that while scrambling of a non-*wh*-phrase is acceptable inside a multiple question (as seen in (150)), scrambling of the *wh*-phrase that does not undergo overt phrasal *wh*-movement is impossible (as seen in (150)). (This generalization is due to Engel 1972, according to Sauerland 1998b; see also Müller and Sternefeld 1996, 483.) Example (150) shows that Superiority is not what is at stake in (150), while (150) is the variant of (150) without scrambling. Example (151) provides a similar paradigm from Grewendorf and Sternefeld (1990):⁹⁴

(150) **No scrambling of *wh*-in-situ**

a. Wie hat wer gestern das auto repariert?

how has who yesterday the car fixed

'How did who fix the car yesterday?'⁹⁵

b. Wie hat das Auto gestern wer repariert?

how has the car yesterday who fixed

'How did who fix the car yesterday?'

c. Wie hat der Mann gestern was repariert?

how has the man yesterday what fixed

'How did the man yesterday fix what?'

d. *Wie hat was der Mann gestern repariert?

how has what the man-NOM yesterday fixed

'How did the man yesterday fix what?'

(Fanselow 1990, 117)

(151) a. Was hat wer dem Studenten ausgeliehen?

what has who the student-DAT given

'What did who give to the student?'

b. Was hat der Professor wem ausgeliehen?

what has the Professor-NOM whom-DAT given

'What did the professor give to whom?'

c. Wem hat der Professor was ausgeliehen?

whom-DAT has the Professor-NOM what given

'To whom did the professor give what?'

d. ??Was hat wem der Professor ausgeliehen?

what has whom-DAT the professor-NOM given

'What did the professor give to whom?'

(Grewendorf and Sternefeld 1990, 5)

Furthermore, if separation constructions provide reliable markers of positions passed through by *wh*-movement, there are other arguments that scrambling to the left of the subject cannot feed *wh*-movement. One argument was pointed out to me by Günther Grewendorf (personal communication). In (125), we examined a separation construction in which a *wh*-word like *wer/wen* 'who' is separated from a restrictor like *von den Musikern* 'of the musicians', stranding the restrictor in object position. Such separation is possible in derived positions as well. (152a), for example, shows *von den Musikern* 'of the musicians' stranded in a SPEC,CP through which a *wh*-phrase has passed. (The presence of overt material in SPEC,CP correlates with "verb -second" (movement of V to C) and with the absence of the complementizer *dass*, as discussed by Den Besten 1983.) The possibility of "separation" in SPEC,CP contrasts sharply with the impossibility of a comparable separation in a post-complementizer scrambling position, seen in (152c), which contrasts minimally with the unscrambled (152b):

(152) **No separation in scrambled position.**

a. *Separation in SPEC,CP*

?Wen glaubst du [__ von den Musikern] hat Hans gesehen.

whom believe-2p you of the musicians has Hans seen

'Who among the musicians do you believe that Hans saw?'

b. *Separation in object position*

Ich weiß nicht, wen du glaubst dass Hans [__ von den Musikern] gesehen hat

I know not whom you believe-2p that Hans of the musicians seen has

'I don't know who among the musicians you believe that Hans saw?'

c. *Separation in scrambled position*

*Ich weiß nicht, wen du glaubst dass [__ von den Musikern] Hans gesehen hat

I know not whom you believe-2p that of the musicians Hans saw

'I don't know who among the musicians you believe that Hans saw?'

Similar paradigms can be constructed with other separation constructions, except that separation across a finite clause boundary seems degraded in the first place in some of these cases, and separation in SPEC,CP is not possible.⁹⁶

(153) **No separation in scrambled position (*was alles*)**

a. *Separation in SPEC,CP*

*Wen glaubst du [__ alles] hat Hans eingeladen.
whom believe-2p you all has Hans invited

b. *Separation in object position*

% Ich weiß nicht, wen du glaubst, dass Hans [__ alles] eingeladen hat.
I know not whom you believe that Hans all invited has
'I don't know who-all you believe that Hans invited.'

c. *Separation in scrambled position*

*Ich weiß nicht, wen du glaubst, dass [__ alles] Hans eingeladen hat.
I know not whom you believe that all Hans invited has
'I don't know who-all you believe that Hans invited.'

There might, of course, be some special explanation for the failure of separation in the scrambled position. I suggest instead that (152) and (153) are teaching us the same thing as (150) and (151) — that *wh*-phrases are simply not scrambled to the left of the subject in the first place.

5.4. Appendix to section 5.3: Further discussion of Fanselow's proposal

Fanselow (1997) provides two arguments in favor of the scrambling proposal which need to be considered. Since this section does not contribute to the forward flow of argumentation, it can be viewed as an appendix to the preceding section, and skipped without losing the thread. Nonetheless, Fanselow's arguments are an important challenge to this portion of our ideas about *wh*-movement, and merit discussion.

Argument #1: Suppose certain separation constructions actually involve — not direct *wh*-movement of a portion of the bigger phrase that might have moved — but scrambling of the phrase ultimately left behind, followed by *wh*-movement of the remnant that includes the trace of scrambling:

(154) **Separation via Scrambling (Fanselow 1997)**

starting point: C ... X [wieviel Bücher über Hans] ...
how many book about Hans

step one: über Hans scrambles to the left of X
C ... über Hans_i X [wieviel Bücher t_i] ...

step two: the remnant *wh*-phrase undergoes phrasal *wh*-movement
[wieviel Bücher t_i] C ... über Hans_i X ...

Müller (1996) and Müller and Sternefeld (1993) have observed that a constituent from which scrambling has taken place may not itself undergo scrambling. If this is the case, then the *wh*-phrase cannot itself undergo scrambling as a precursor to *wh*-movement. Now consider a situation in which *über Hans* is replaced by a *wh*-phrase such as *über wen* 'about whom' in the starting point of (154). If *über wen* undergoes scrambling, then, by Müller's generalization, the remnant [wieviel Bücher t] may not itself undergo scrambling. If Superiority holds in German as it does in English, we expect the (now) higher *wh*-phrase *über wen* to be the only phrase attractable by the interrogative complementizer. This prediction is borne out:⁹⁷

(155) a. Ich wüßte gerne, über wen er [wieviel Bücher t] gleichzeitig kritisieren kann.

I knew eagerly about whom he how-many books simultaneously criticize can
'I wanted to know how many books about whom he can simultaneously criticize.'

b. ?*Ich wüßte gerne, [wieviel Bücher t] du gleichzeitig über wen kritisieren kannst.

(cf. the same example with *über Hans* replacing *über wen* is fine)

Notice, however, that (155b) differs from (155a) in other ways that we know influence acceptability. For one thing, (155b) is a weak crossover configuration: the variable associated with *wen* is an argument of [*wieviel Bücher t*], which c-commands *wen*. The weak crossover effect may be particularly strong with *wh*-feature movement (an observation also made, in slightly different terms, by Miyagawa 1998).⁹⁸ This fact alone might produce the judgment given, which would be (in our approach) akin to the effects seen in English examples like (156) and (157b):⁹⁹

(156) ??[How many books about him_i] did who_i show __ to whom?

(157) a. Which of his friends __ ordered which boy to congratulate Sue __?

b. [*]Which of his friends did Sue order which boy to congratulate __?

Argument #2: If separation constructions arise from a derivation like (154), then the "separated" piece of the *wh*-moved phrase is "an indicator of the uppermost position the *wh* phrase had prior to *wh*-attraction," since, by Müller's generalization, the *wh*-phrase could not have scrambled before undergoing *wh*-movement. Therefore, if another *wh*-phrase is found higher than the separated piece, and the Superiority Effect holds in German just as it does in English, it should be that higher *wh*-phrase that needs to undergo overt *wh*-movement.

Thus, for example, in (158a), the underscored phrase *von den Studenten* marks the highest position that the remnant [*wen t*] could have occupied before *wh*-movement. Consequently, when the subject *der Peter* in (158a) is replaced by the *wh*-phrase *wer* 'who' in (158b), the result is a Superiority violation (from Fanselow's perspective), since *wer* — as the higher *wh*-expression — should have undergone overt *wh*-movement instead of [*wen t*]. In (158c), scrambling of *wen von den Studenten* took place before *von den Studenten* scrambled further. Since the launching site for *wh*-movement of [*wen t*] is higher than the subject *wh*-phrase *wer* in this example, it is [*wen t*] that undergoes overt *wh*-movement, with no problem.

- (158) a. [Wen t_i] hat denn der Peter im Sommersemester von den Studenten_i prüfen wollen?
whom has then the Peter in-the Summer term of the students examine wanted
- b. ?*[Wen t_i] hat denn *wer* im Sommersemester von den Studenten_i prüfen wollen?
- c. [Wen t_i] hat denn von den Studenten_i *wer* im Sommersemester prüfen wollen.

Here too there is an immediate alternative explanation. Example (158b) is ruled out by the Intervention Effect, since the scope-bearing element *wer* intervenes between a part of the restriction on *wen* and *wen* itself. As Günther Grewendorf (personal communication) has pointed out, *wh*-phrases in situ count as Interveners for the Intervention Effect:

- (159) a. Was hat der Professor dem Studenten [__ alles] geraten
what has the professor the student-DAT all advised
'What-all has the professor advised the student?'
- b. *Was hat *wer* dem Studenten [__ alles] geraten
what has who the student-DAT all advised
'What-all has who advised the student?'

I now return to the main line of discussion.

5.5. Japanese and a typology of *wh*-specifiers

The earlier parts of this section provided support for a key component of our account of apparent superiority violations: the notion of "specifier potential". According to our proposal, the German lexicon contains only C1-spec, while the English (and Bulgarian) lexicon contains both C1-spec and Cm-spec. This lexical difference, I argued, is the key to gross differences in the distribution of the Intervention Effect in German and English multiple questions. Since German allows only one *wh*-specifier per question, the

restrictions of the other *wh*-phrases in a multiple question remain separated from the interrogative complementizer. This generates an Intervention Effect when a scope-bearing element occupies an intervening position. Since English allows multiple *wh*-specifiers, the Intervention Effect is found only in cases where an apparent violation of Superiority indicates that feature-movement has taken place from a *wh*-phrase in situ.

Alongside interrogative complementizers that require multiple specifiers and interrogative complementizers that require a single specifier, it is reasonable to expect to find an interrogative complementizer that tolerates no specifiers whatsoever: i.e. *C*_{0-spec}. As it happens, our hypotheses make very specific predictions about the properties questions introduced by *C*_{0-spec}. I will argue that *C*_{0-spec} does exist, and that these predictions seem to be fully confirmed. In particular, I will suggest that Japanese (and Korean) *wh*-questions are introduced by *C*_{0-spec}. Japanese/Korean, German and English/Bulgarian thus round out the paradigm of "specifier potential" for interrogative complementizers:

(160) **Maximum specifier potential for the interrogative C in three types of languages**

	<i>C</i> _{0-spec}	<i>C</i> _{1-spec}	<i>C</i> _{m-spec}
Japanese/Korean	√		
German		√	
English/Bulgarian			√

For Japanese, I rely mostly on data and insights from Miyagawa (1998). For Korean (which I shall not focus on), I rely on Beck and Kim (1996).

What do we expect from a language that must introduce *wh*-questions with *C*_{0-spec}? Obviously, such a language will show *wh*-in-situ questions. As is well-known, Japanese and Korean are languages of this sort:

(161) ***wh*-in-situ (Japanese)**

John-ga nani-o katta no?

John-NOM what-ACC bought Q

(162) ***wh*-in-situ (Korean)**

Suna-ka muo•s-u•l sass-ni?

Suna-NOM what-ACC bought Q

Except for differences in specifier potential, the interrogative complementizer used for *wh*-questions should be fundamentally identical across languages. Consequently, we expect that Japanese and Korean C_{0-spec} , like C_{1-spec} and C_{m-spec} , bears an uninterpretable *wh*-feature. If this feature is not deleted in the course of the derivation, the derivation should crash. Since C_{0-spec} does not allow *wh*-phrase movement to form a specifier, only *wh*-feature movement (or its counterpart in the variant proposals discussed above) can delete the uninterpretable feature on C_{0-spec} . Consequently, a *wh*-question in such a language — be it a single question or a multiple question — must contain at least one *wh*-phrase *not* separated from C_{0-spec} by an intervening scope-bearer. If the question contains just one *wh*-phrase, and this *wh*-phrase is separated from C_{0-spec} by a scope-bearing intervener, the question should be unacceptable. As noted by Hoji (1985), this is indeed the case in Japanese. A scope-bearing element in a single-*wh*-question produces an Intervention Effect whenever it comes between the *wh*-phrase and the interrogative complementizer. As in German, scrambling eliminates the violation:

(163) **Intervention effect with *wh* and negation/only (Japanese)**¹⁰⁰

a. *Hanako-sika nani-o yoma-nai no?

Hanako-only what-ACC read-Neg Q

b. *Scrambling eliminates the violation*

Nani-o Hanako-sika yoma-nai no?

(Tanaka 1999, *per* Hagstrom 1998)

(164) **Intervention effect with *wh* and *every* (Japanese)**¹⁰¹

a. ?*Dono hito-mo nani-o yonda no?

every person what-ACC read Q

[less deviant for some speakers, allegedly not necessarily with wide-scope *every*]

b. *Scrambling eliminates the violation:*

Nani-o dono hito-mo yonda no?

(Hoji 1985)

(165) **Intervention effect with *wh* and *almost every* (Japanese)**

a. *Hotondo dono hito-mo nani-o yonda no?

almost every person what-ACC read Q

[for all speakers]

b. *Scrambling eliminates the violation*

Nani-o hotondo dono hito-mo yonda no?

(Miyagawa 1998)

(166) **Intervention effect with *wh* and *or* (Japanese)**

a. ?*[John-ka Bill]-ga nani-o nomimasita ka?

John or Bill-NOM what-ACC drank Q

b. Nani-o [John-ka Bill]-ga nomimasita ka?

(Hoji 1985; see also Hagstrom 1998)

In essence, examples like the (a) sentences in (163)-(166) present the Japanese speaker with a Hobson's choice. If no feature movement takes place, the uninterpretable *wh*-feature in C_{0-spec} is not deleted, resulting in "crash". If feature movement does take place, the result is an Intervention Effect. Quite comparable data have been provided for Korean by Beck and Kim (1996), for example:

(167) **Intervention effect with *wh* and Negative Polarity Item (Korean)**

a. *Amuto muo•s-u•l sa-chi anhass-ni?

Anyone what-ACC bought-CHI not-did Q

'What did no one buy?'

b. Muo•s-u•l amuto sa-chi anhass-ni?

(168) **Intervention effect with *wh* and *only* (Korean)**

a. *Minsu-man nuku-lu•l poass-ni?

Minsu-only who-ACC saw Q

b. Nuku-lu•l Minsu-man poass-ni?

(Beck and Kim 1996, example (80))

(169) **Intervention effect with *wh* and *every* (Korean)**

a. ??Nukuna-ka o•nu• kyosu-lu•l chonkyo•ngha-ni?
 everyone-NOM which professor-ACC respect Q

b. O•nu• kyosu-lu•l nukuna-ka chonkyo•ngha-ni?

If the question contains multiple *wh*-phrases, the nature of the Intervention Effect is determined by the position of the intervener. When an intervener c-commands all *wh*-expressions, the same deviance results as in single *wh*-questions that have the same problem:

(170) **Intervener (*only*) c-commands multiple *wh*-phrases (Japanese)**

[compare (172) below]

?*Taroo-wa [Hanako-sika dare-ni nani-o yom-ana-i to] Tomoko-ni itta no?
 Taroo-TOP Hanako-only who-to what-ACC read-NEG-PRES C T.-to said Q
 'To whom did Taroo tell Tomoko that only Hanako read what?'
 (Miyagawa 1997 handout; quoting Tanaka 1999)

The cause of the deviance is the same as in the single-*wh* questions just examined. Either the complementizer is left with an undeleted *wh*-feature, or else an Intervention Effect is invoked.

When the intervener c-commands one, but not all the *wh*-expressions in Japanese, the result is not deviant (Watanabe 1992; Tanaka 1999). (This is often known as the "additional *wh*-effect", following Saito 1994.) This is because *wh*-feature movement (or its counterpart in variant theories) can take place from the highest *wh*-phrase, eliminating the uninterpretable *wh*-feature on the complementizer. Nonetheless, the resulting sentences are limited to the single-pair reading, just as we have seen in English and German (Miyagawa 1997). The reason is the observation stated in (142) above, which characterized the distribution of the single-pair reading in English and German:

(171) **Intervention effect in multiple questions with *or* (Japanese)**

a. ???John-ga [MIT-ka Harvard]-ni nani-o ageta no?
John-NOM [MIT-or Harvard]-Dat what-ACC gave Q
'What did John give MIT or Harvard?'

b. Dare-ga [MIT-ka Harvard]-ni nani-o ageta no?
who-NOM [MIT-or Harvard]-Dat what-ACC gave Q
'What did who give MIT or Harvard?' [single-pair only]

(172) **Intervention effect in multiple questions with *only* (Japanese)**

a. ?*Taroo-wa [Hanako-sika nani-o yom-ana-i to] Tomoko-ni itta no?
Taroo-TOP Hanako-only what-ACC read-NEG-PRES C T.-to said Q
'What did Taroo say that only Hanako read?'

b. Taroo-wa [Hanako-sika nani-o yom-ana-i to] dare-ni itta no?
Taroo-TOP Hanako-only what-ACC read-NEG-PRES C who-to said Q
'To whom did Taroo say that only Hanako read what?' [single-pair only]

As in German, scrambling eliminates the intervention effect:

(173) **Scrambling (Japanese)**

a. ?*[John-ka Bill]-ga dare-ni nani-o ageta no?

John or Bill-NOM who-DAT what-ACC gave Q

'What did John or Bill give to who?'

b. ? Dare-ni [John-ka Bill]-ga ___ nani-o ageta no?

[single-pair only]

c. ?? Nani-o [John-ka Bill]-ga dare-ni ___ ageta no?

[single-pair only]

d. Dare-ni nani-o [John-ka Bill]-ga ___ ___ ageta no?

[pair-list reading is ok]¹⁰²

(Hagstrom 1998)

In Korean, similar facts seem to obtain — except that I have not investigated whether examples in which some, but not all *wh*-phrases are c-commanded by the intervener have the single-pair reading. Beck and Kim (1996) mark such examples with a star, which I reproduce below in square brackets. If the star is real, even on a single-pair reading, the difference between Korean and Japanese will require further investigation:

(174) **Scrambling (Korean)**

a. *Amuto nuku-lu·l o·ti-eso· manna-chi anhass-ni?

Anyone who-ACC where-LOC meet-CHI not-did Q

'Where did no one meet whom?'

b. [*]Nuku-lu·l amuto o·ti-eso· manna-chi anhass-ni?

[ok if single-pair?]

c. Nuku-lu·l o·ti-eso· amuto manna-chi anhass-ni?

d. O·ti-eso· nuku-lu·l amuto manna-chi anhass-ni?

As a final piece of reassurance, it is worth observing, with Miyagawa (1998), that Japanese displays the Intervention Effect in non-*wh* environments exactly where we expect to find it: in separation constructions. For example, when existential *dareka* 'some (adj.)' scrambles in a manner that strands its restriction, as in (175a-b), a scope-bearing phrase like 'only Hanako' may not intervene between *dareka* and its restrictor:

(175) **Intervention effect with Q-NP split (Japanese)**

a. Hanako-ga gakusei-o dareka yonda (koto)
Hanako-NOM student-ACC some invited (fact)
'(the fact that) Hanako invited some student'

b. Dareka_i Hanako-ga gakusei-o *t_i* yonda (koto)

c. Hanako-sika gakusei-o dareka_i yoba-nakat-ta (koto)
Hanako-only student-ACC some invite-Neg-Past fact
'(the fact that) only Hanako invited some student'

d. *Dareka_i Hanako-sika gakusei-o *t_i* yoba-nakat-ta (koto)
some Hanako-only student-ACC invite-Neg-Past fact

Similar facts obtain when a numeral is scrambled away from its restriction:

(176) **Intervention effect with Num-NP split (Japanese)**

b. ? San-satu_i Taroo-ga gengogaku-no hon-o *t_i* yonda.
3-CL Taroo-NOM linguistics-GEN book-ACC read

d. * San-satu_i Taroo-sika gengogaku-no hon-o *t_i* yoma-nakatta.
3-CL Taroo-only linguistics-GEN book-ACC read-NEG

Japanese and Korean thus appear to meet our expectations for a language whose interrogative complementizer is identical to that of English and German, except that it does not take a specifier.¹⁰³ Since the notion of "specifier potential" played a crucial role in our original discussion of apparent AC violations, the results in this section help support conclusions we have already reached. The notion of "specifier potential" has been shown to play a key role in explaining variation in the syntax of *wh*-questions. In particular, the scale of specifier potential has been shown to correlate inversely with the number of environments in which *wh*-questions show intervention effects:

(177) **Environments in which Intervention Effects are found in *wh*-questions:**

	single- <i>wh</i> questions	multiple- <i>wh</i> questions	multiple- <i>wh</i> questions with apparent AC violations
<i>C_{0-spec}</i>	√	√	√
<i>C_{1-spec}</i>		√	√
<i>C_{m-spec}</i>			√

This makes sense, since the fewer opportunities a language has for phrasal *wh*-movement, the greater the number of environments in which a restriction on a *wh*-phrase must remain inside TP, yielding Intervention Effects.¹⁰⁴

5.6. Some unanswered questions about the typology

Some important questions remain, to which the answers are not completely clear. One question concerns the repertoire of complementizers in a given language. I have described the list of available complementizers as fixed in the lexicon of each language. This is important to the action of the Complementizer Choice Rule in (141). If Japanese, for example, had the option of using *C_{m-spec}* in a multiple question, the Complementizer Choice Rule should require its use in Japanese just as it does in English, where the forced use of *C_{m-spec}* forms a key component of the explanation for the Superiority

Effect. As it happens, on one interpretation of Takahashi (1993), we may suspect that we are faced with exactly this problem.¹⁰⁵

Though Japanese is famously a "*wh* in situ language", it allows *wh*-phrases to be fronted by a process that in certain cases has the properties of English or Bulgarian *wh*-movement, according to Takahashi. These properties show up when the *wh*-phrase is moved out of an embedded clause to the left periphery of an interrogative CP. Thus, for example, (178), in which *nani* 'what' has been fronted from a lower interrogative CP to the left periphery of a matrix interrogative CP, can only be understood as a matrix question:

(178) **Overt *wh*-fronting in Japanese interrogative fixes scope.**

Nani-o John-wa [CP Mary-ga ___ tabeta ka] siritagatteiru no?
what-ACC John-TOP Mary-NOM ate *Q* want-to-know *Q*

Only: 'What did John want to know whether Mary ate.'

Not: 'Did John want to know what Mary ate?'

(Takahashi 1993, ex. (4b))

In all other cases, *wh*-fronting has the properties expected from simple scrambling: "A-movement-like" properties when the fronting is more local (Webelhuth 1989, Mahajan 1990, Saito 1992), and reconstruction properties when the fronting does not target an interrogative CP. For example, (179) is identical to (178) except that the matrix clause is non-interrogative. In this case, the fronted *wh*-phrase *nani* is understood via reconstruction as taking scope in the embedded clause:

(179) **Wh-reconstruction with fronting in a non-interrogative clause**

Nani-o John-wa [CP Mary-ga ___ tabeta ka] siritagatteiru.
what-ACC John-TOP Mary-NOM ate Q want-to-know
'John wanted to know what Mary ate?'

(Takahashi 1993, based on (2b))

The construction behaves as if the following were true of Japanese:¹⁰⁶

- The left periphery of an interrogative CP may not be the target of long-distance scrambling of a *wh*-phrase.
- The left periphery of an interrogative CP may be the target of long-distance *wh*-movement.

Certain facts about this construction are murky. For example, Richards (1997) observes that multiple questions with two *wh*-phrases allow both *wh*-phrases to front, in which case a Bulgarian-style Superiority Effect was observed by Richards' consultant. Some speakers disagree with this judgment, even to the extent of finding (180b) marginally superior to (180a):¹⁰⁷

(180) **Bulgarian-style Superiority Effect (Japanese)**

- a. Dare_i-ni nani_j-o Taroo-ga t_j [Hanako-ga t_j katta to] itta no?
b. * Nani_j-o dare_i-ni Taroo-ga t_j [Hanako-ga t_j katta to] itta no?

(Takako Aikawa, personal communication to Richards 1997)

Takahashi noted the possibility of fronting only one of the two phrases, and noted that a Superiority effect of the English sort arises here too:

(181) **English-style Superiority Effect (Japanese)**

a. Dare-ni John-ga [Bill-ga ___ [Mary-ga nani-o tabeta to] itta to] omotteiru no?
 whom-DAT John-NOM Bill-NOM Mary-NOM what-ACC ate that said that thinks *Q*

b. *Nani-o John-ga [Bill-ga dare-ni [Mary-ga ___ tabeta to] itta to] omotteiru no?

(Takahashi 1993, 664)

Some controversy surrounds these observations as well. For one thing, the sharp cases of the "English-style" Superiority Effect seem to involve situations like (181), in which one *wh*-phrase comes from a higher clause than the other. Other examples show weaker and more variable effects. If so, we might not be dealing with a Superiority Effect at all in (181), but with a simpler contrast between longer and shorter-distance scrambling — much as I suggested for German examples like (147a-b). In support of this is the fact that multiple questions with left-peripheral *wh*-phrases do show an Intervention Effect when a scope-bearing element c-commands an unmoved *wh*-phrase in the question: a German pattern. If this is the case, then our picture of Japanese must be more complex than represented so far. We must actually allow Japanese questions to be introduced by *C_{0-spec}*, *C_{1-spec}* or *C_{m-spec}*: the first for in-situ questions, the second for questions like those in (181), and the third for "Bulgarian-style" questions like those in (180) — with the Complementizer Choice Rule somehow inactive in Japanese. Alternatively, we might continue to view the interrogative complementizer in Japanese as an instance of *C_{0-spec}*, and ask whether the attractor in putative overt *wh*-movement isn't a different entity entirely — perhaps an optional focus head above or below *C_{0-spec}*, which optionally takes multiple specifiers. For lack of clear and decisive evidence, I must leave the matter open.

A similar issue arises in French — a language which overall appears to display the English patterns of movement and pronunciation. It is well-known, however, that in matrix clauses, French allows *wh* to remain in situ rather more freely than does English, without necessarily producing an "echo question" reading:

(182) **Matrix *wh*-in-situ (French)**

Ils ont rencontré qui?

They met who

'Who did they meet?' [non-echo reading possible]

Chang (1997) and Bošković (1997) have observed that *wh*-in-situ in this type of question shows the Intervention effect. The examples in (183) are only acceptable as echo questions, as indicated by the annotation "#" (examples from Chang 1997):

(183) **Intervention effects (French)**

a. #Tous les étudiants ont rencontré qui?

all the students met who

b. #Chaque étudiant a rencontré qui?

each student met who

c. #Il n'a pas rencontré qui?

he didn't meet who

d. #Il admire toujours qui?

he admires always who

f. #Personne n' admire qui?

no one neg admires who

If the Intervention Effect here is a sign that these matrix questions are introduced by C_{0-spec} , we once again need to ask whether the Complementizer Choice Rule is somehow inactive in French as well (with the English-style pronunciation pattern when C_{m-spec} is chosen) — or whether C_{0-spec} is perhaps the only

option for the interrogative complementizer, with some other attractor, optionally absent from matrix questions, responsible for what we have hitherto thought of as *wh*-movement.¹⁰⁸

Alternatively, C_{0-spec} may be a special option related to the semantics of questions like (182). Boeckx (1999a; 1999b) has argued that such questions are interpreted in a manner reminiscent of cleft sentences (e.g. *Who is it that you met*) — with a strong presuppositional reading. *Personne* 'no one', for example, is not a felicitous answer to (182). If this is the case, we may suppose that French is in general like English in its use of interrogative complementizers, except for the special possibility of using C_{0-spec} as a complementizer in strongly presuppositional single questions. This fact in turn might be compared to the fact that English (and French) allow C_{I-spec} as the complementizer for D-linked multiple questions, as discussed earlier in this monograph. Perhaps questions that are strongly context dependent quite generally allow complementizers with smaller than expected specifier potentials. I leave this line of thinking for further research.

Also left for further research is the possible existence of *pronunciation* patterns for multiple phrasal *wh*-movement besides those of English and those of Bulgarian. For example, we noted that the use of C_{0-spec} to introduce a *wh*-question yields a question whose *wh*-phrases are pronounced in situ because they are in situ. The test of this hypothesis was the Intervention Effect. But there are other possible ways to derive a "*wh*-in-situ" language. One might imagine a language whose interrogatives could be introduced by C_{I-spec} or C_{m-spec} — just like English — but which pronounces *all wh*-phrases in their pre-*wh*-moved position. (This is essentially the traditional proposal concerning *wh*-in-situ languages.) The sign of a "superficial" *wh*-in-situ language of this sort would be the absence of the Intervention Effect in simple *wh*-questions. As described by Aoun and Li, Chinese is such a language (Aoun and Li 1993b).¹⁰⁹

Though open questions like these remain, the present study has offered a series of arguments that remain (I believe) unchallenged by the open questions. These include our arguments for the existence of overt and covert phrasal movement as well as the relation I have called feature movement — though other views of this relation were also considered. The evidence for this position came from several sources,

including ACD, Intervention Effects, and the typology of *wh*-constructions with which we have concluded this study.¹¹⁰ If successful, this study has filled in an important piece of the puzzle posed by *wh*-constructions — a puzzle whose solution illuminates the variety and nature of movement and its kin.

FOOTNOTES

¹In his most recent work, Chomsky (1998) has suggested a view of "feature movement phenomena" according to which movement *per se* is not part of the relation. I touch on this proposal later in this study.

²Though not always. See Pesetsky (1997, 1998) for discussion of cases in which pronunciation targets more than one position in a movement construction.

³I use the term *phrasal movement* as a cover term for movement of any syntactic unit that is word-sized or larger. Thus, what is traditionally called "head movement" is an instance of phrasal movement in my sense.

⁴Larson and May's examples involve the matrix verb *want*. The phenomenon is clearer with negative verbs like *refuse* that make the scope of the quantified phrase entirely obvious.

⁵In light of Müller (1996) and Müller and Sternefeld (1993), we must suppose that the c-command property is apparent because VP-fronting and *wh*-movement are in some fashion the same process. Where two movement processes are sufficiently different, the interaction with c-command is not observed (e.g. when scrambling and *wh*-movement yield a configuration in which a trace is not c-commanded by its antecedent).

⁶The order of moved *wh*-phrases given here (*who what whom*) anticipates later discussion, but is irrelevant at this point.

⁷A third approach, represented by Kayne (1998), reanalyzes apparent covert phrasal movement as overt movement masked by compensating overt movement of other constituents. On this view, for example, the derivation of *What did John give to whom?* might involve overt leftward movement of *what* to a position below *who* followed by leftward movement of the remnant IP containing traces of *what* and *to*

whom. I will not investigate this alternative here. There is probably no straightforward translation of our proposals in this paper into this approach.

⁸ This proposal assumes that it is a property of the attracted element that causes movement to be phrasal rather than purely featural. Chomsky 1998 suggests an alternative: that it is a property of the attractor that causes movement to be phrasal (a property anachronistically called the "Extended Projection Principle" (EPP) feature; Chomsky 1981). Our proposal in the end will more closely resemble that of Chomsky 1998, but this detail is not important at the present stage of our argument.

⁹Chomsky's proposal leaves us with only a single style of phrasal movement, but does not explain why this style of movement involves the "overt" pronunciation pattern, as opposed to some other. Nunes (1995) and Wilder (1999) (among others) have offered proposals, based on Kayne (1994), that fill this gap in Chomsky's picture. Of course, if the picture is incorrect (as this monograph argues), then Nunes and Wilder's proposals must be modified in some way. Both Nunes's and Wilder's work explore interesting consequences of their hypotheses beyond the simple prediction of overt movement patterns, so it would be desirable to explore alternatives that retain these interesting results.

¹⁰ Brody (1995) developed independently a notion of "expletive-associate" chain whose properties are close to those one would need to attribute to the operation of Feature Movement in Chomsky's later proposal. Unlike Chomsky, Brody (pp. 114-127) suggested a treatment of ACD compatible with his proposal that responds to some (but not all) of the evidence taken to support a treatment in terms of covert phrasal movement. In particular, he offers a proposal that responds to the scope evidence, but not to Fox's interpretation of the binding evidence in (7) (not available at the time Brody's work was written) – nor, obviously, to the new evidence presented in the present study. He also expresses some skepticism (p. 120) about the validity of the core fact — though a claimed counterexample (Brody's (50)) actually follows from Fox's proposal.

¹¹ I owe these thoughts to a suggestion of Norvin Richards. See also discussion in Brody 1997.

¹²The proposal has some interesting similarities to proposals within the GPSG-HPSG tradition (Gazdar 1981; Gazdar, Klein, Pullum and Sag 1995; Pollard and Sag 1994), which treat a wide variety of syntactic relations as the result of extremely local feature movement.

¹³ Kai von Stechow (personal communication) suggests that the effect may be weaker in *there* constructions that do not involve the verb *be* (judgment indications mine, but tentative):

(i) [??]There arose every problem that there could.

(ii) [??]On Tuesday, there didn't occur any of the disruptions that there did on Monday.

If this is correct, it may perhaps be related to the obligatory focus interpretation of the associate in a *there* construction without *be*. Conceivably the associate is actually extraposed from VP overtly, in the manner of Heavy NP Shift.

¹⁴It is also worth noting that the copula does not block ACD. For example, the US Army's recruiting slogan *Be all that you can be!* could have been offered in a form with ACD: *Be all that you can!*

¹⁵Covert head movement might also fit the bill, except that overt cases of head movement generally obey strict locality conditions, while the movement discovered here (as discussed below) obeys very weak locality conditions. See Hagstrom 1998 for discussion, and some competing considerations.

¹⁶Because most *wh*-expressions may be understood as D-linked, it is hard to tell which among the *wh*-phrases in a multiple question must be D-linked for the Superiority Effect to disappear. My own impression of speakers' judgments is that either *wh*₁ or the *wh*-phrase that moves overtly must be D-linked, but opinions differ. (For example, Comorovski 1996 claims that it is crucial for both the overtly moved phrase and the in situ phrase to be D-linked.) I will try to sidestep this issue by referring sloppily to "D-linked questions" or "questions with D-linking", even though technically it is *wh*-phrases that are D-linked. Any attempt to explain the semantic sources of the syntactic properties of D-linking will require more conclusive investigation of this issue.

¹⁷The effect seen in (33) was first noted, to my knowledge, by Kayne (1983), who offered an account in terms of his "Connectedness" principle — a proposed revision of the ECP. His principle covers cases in which the highest *wh*-in-situ is not lexically governed, i.e. is a subject. The ameliorating effect of a "third" *wh*-phrase is not limited, however, to those cases of the Superiority Effect in which the highest *wh*-in-situ is a subject:

(i) ??What did you persuade whom to give __ to Mary?

(ii) What did you persuade who to give to whom.

The variant of Connectedness proposed by Pesetsky (1982), unlike Kayne's proposal, did cover these cases, though the observation was not made explicitly. Our observations about questions with three or more *wh*-phrases have been independently made by Fiengo (1998).

¹⁸I have encountered a few speakers who find (33b) less than perfect. All speakers acknowledge a contrast.

¹⁹The double-starred examples in (35), (40) and (41) improve when the *wh*-in-situ is D-linked. I return to this fact below.

²⁰Some speakers of some Slavic languages report stronger effects of (37) than do others. There may be individual variation on this point — or cross-language variation. Empirical work more fine-grained than I have undertaken is necessary.

²¹ Though it is tempting at this point to imagine that clauses introduced with C_{I-spec} uniformly receive a single-question interpretation, and that clauses introduced with C_{m-spec} uniformly receive a

multiple-question interpretation, the facts will turn out to be somewhat more complex. In particular, certain multiple questions (in English and in other languages) appear to be introduced with C_{I-spec} .

²²Of course, one might continue to maintain that only a gradient (37) influences the judgment, by positing a non-linear relation between number of violations and actual judgments. (An anonymous reviewer reports that comparable judgments in Serbo-Croatian are detectable gradient, without the major acceptability break reported here for Bulgarian.) In section 4.3 below, I argue that English multiple *wh*-questions with three or more *wh*-phrases are also instances of the phenomenon in (35). This argument will help settle the matter in favor of (36) and a non-gradient (37).

²³This section mostly summarizes the results of Richards (1997), whose work — along with Beck (1996) — provided the inspiration for the research reported in this study.

²⁴As discussed later, the contrasts in (43) and (44) weaken in questions with D-linking. I have also encountered a few speakers who simply do not agree with the contrasts in (43) and (44), either finding all orders acceptable, or reporting a hunch that particular intonation contours render the (b) examples acceptable, independent of D-linking. I have not been able to understand what factors distinguish these speakers from those who detect the judgments discussed here, nor have I been able to get to the bottom of the reports of an intonational difference.

Billings and Rudin (1994) identify several factors that license exceptions to the Superiority Effect among their Bulgarian consultants. These include animacy and avoidance of phonetic identity. My informants and Richards' have not assented to the judgments reported by Billings and Rudin. Nonetheless, it is not impossible to imagine ways in which their data could be understood within the framework of hypotheses reported here, if considerations of animacy and phonetic identity are allowed to override the effects of Shortest Move. I am grateful to Barbara Citko for bringing Billings and Rudin's article to my attention.

²⁵Müller (1998), developing an optimality-theoretic approach to these phenomena, encodes this property directly as a constraint requiring "parallel movement" — movement that minimizes reversals of c-command. As far as I can tell, the approaches are empirically indistinguishable in this case.

²⁶It is worth noting that we might extend Chomsky's requirement to cover all instances of movement, both overt and covert, if the Y-model is abandoned, as suggested above.

²⁷The facts of Bulgarian island effects are somewhat more complex than represented here, as discussed in detail by Richards (1977). I have presented the subjacency argument simply as a further demonstration of the PMC. The point most relevant for us here, however, is the effect of the PMC on AC, as described above. Here again I have found some variability among speakers.

It should also be noted, as Günther Grewendorf (personal communication) has reminded me, that the condition that bars extraction from islands of adjuncts (or perhaps non-NPs) is not subject to the PMC. This may provide an argument that this condition has a non-syntactic basis.

²⁸Bošković (1998; in press) offers an alternative account of the ordering of Bulgarian triple *wh*-questions that does not extend to the subjacency data just considered, but is compatible with the rest of this monograph. He, like Richards, assumes that multiple movement to the same head involves "tucking in". Like Richards and many others, he assumes that Superiority effects arise from a closeness condition on the operation Attract, an operation in which the featural needs of an attracting head are satisfied by copying some expression that can satisfy those needs.

Unlike Richards, however, Bošković suggests that when multiple *wh*-movement to the same C-system is observed, only the *first* instance of movement results from attraction of a *wh*-feature by an interrogative C. (An antecedent for this idea can be found in Cheng's 1991 theory that the first instance of *wh*-movement is special because only it "types" the sentence as interrogative.) All other instances of movement to specifier of that C are the result of featural needs intrinsic to the *wh*-phrases themselves (Focus, perhaps) — not needs intrinsic to C. Thus, in a question that involves three *wh*-phrases, the first instance of *wh*-movement to C is a case of "Attract", but the other two instances are cases of "Move".

When two *wh*-phrases have a need to move to SPEC,CP in a clause in which a third *wh* (*wh*₁) has already been attracted to C, no principle regulates which of the two remaining *wh*-phrases moves first. That is why both *wh*₁ *wh*₂ *wh*₃ and *wh*₁ *wh*₃ *wh*₂ orders are possible. As far as I can tell, most of the results reported here are compatible with this hypothesis. Note, however, that *wh*-movement and Focus movement must both target SPEC,CP in Bulgarian. If the focus position were lower than C, the closest *wh*-phrase to C would be the highest *wh*-phrase in SPEC,Focus, and we would not expect to find the Superiority Effect. This raises questions about focused non-*wh*-phrases, which are not freely interspersed among the fronted *wh*-phrases of a multiple question, as would seem to be predicted by this approach.

²⁹This generalization clearly undermines a traditional argument for the segregation of covert movement into a component distinct from overt movement: their differing levels of obedience to subjacency. (The argument for cases of single-*wh* questions in languages like Chinese and Japanese, which also do not seem to show island effects. For a reanalysis of these cases, see Hagstrom 1998.) Despite this, Richards develops his PMC in the context of the Y-model, and presents other arguments in support of that model. I will not evaluate these arguments here.

³⁰ I will not derive this difference between English and Bulgarian from deeper properties of the grammars of these two languages.

In presenting this material, I have occasionally encountered the view that the absence of such a connection is a drawback of the proposal. As a spur to further research, such an objection can be useful. It is always possible that we will discover a cross-linguistic correlation between following (56) vs. (57) and some other property of grammar. Furthermore, we will never discover such a correlation unless we look for it.

Nonetheless, as an a priori view, the objection is misplaced. Languages do differ, as a consequence of the role of linguistic experience in language acquisition. This fact means that certain differences among languages are virtually guaranteed to be irreducible. It is certainly possible (though not inevitable) that this is

true of the language-particular choice of (56) vs. (57). The fact that the Germanic language Yiddish and the Romance language Romanian both show the Slavic pattern of (57) might support this view, by hinting that obedience to (57) does not depend too closely on other properties of syntax that distinguish the Slavic family of languages from its neighbors.

³¹There is some controversy concerning whether string-vacuous *wh*-movement is possible (George 1980; Gazdar 1981; Chung and McCloskey 1983; Chomsky 1986). I assume that it is. The opposite assumption would require non-lethal but substantial reformulations of our proposals.

³²Kai von Fintel (personal communication) suggests that this phenomenon may be related to an observation of Irene Heim's (personal communication to von Fintel). I can offer a visitor a cup of coffee by asking "Would you like a cup of coffee?" Suppose the visitor finishes that cup and I want to offer him another. I cannot, in this context, simply repeat "Would you like a cup of coffee?", even though a standard semantics for indefinites allows this. I need to explicitly signal the fact that the question is a repetition by asking "Would you like *another* cup of coffee?" Repeated VPs (whether elided or not) may require explicit markers like *too* or *all* as part of the same phenomenon. The real puzzle becomes the absence of this requirement in relative clauses modifying quantifiers, as in familiar examples of ACD like (67a), where the addition of *also* actually makes the relative clause relatively unacceptable.

³³Lasnik (taking up a suggestion of Baltin 1987) suggests that overt extraposition also resolves ACD in certain types of constructions. To simplify the discussion, I put this possibility aside.

³⁴On this view, the impossibility of ACD in the associate of a *there* sentence, seen in (25), might be a consequence of the assignment of a kind of case to the associate that does not require raising out of the VP — for example, what is sometimes called *inherent* Case. This possibility was suggested by an anonymous reviewer of this monograph.

³⁵If there is any phrasal movement for Case, we expect at least the availability of a reading in which Case-motivated movement of the outer bracketed DP permits ACD resolution in which the entire direct object

is a variable, e.g. for (71b), the interpretation *Melander requested copies of most of the tapes Larsson requested*. This is generally unavailable — which would be a good argument for the stronger position (that there is no phrasal movement for Case) were it not that QR also might be expected to allow this reading. If the outer bracketed DP undergoes QR (followed by secondary QR of the inner bracketed phrase), it creates a VP of the sort that could yield the unattested interpretation.

Though Kennedy discusses this issue at length — treating the possibility of double QR as a problem — Sauerland (1998) notes that under certain circumstances this type of reading *is* available (perhaps marginally), especially when the inner and outer DPs contain identical material. *Bill visited a city near the city Mary did* thus seems to have two readings: 'Bill visited a city near the city that Mary visited a city near', and 'Bill visited a city near the city that Mary visited'. Space does not permit me to sketch Sauerland's proposal here, but its logical structure is quite compatible with the proposals of this study.

³⁶This conclusion dovetails with the observation of an anonymous reviewer for MIT Press that even predicates — expressions that seem not to need case — may contain ACD, e.g. *I consider Bill everything that John doesn't*.

³⁷Perhaps the best argument I know in favor of this proposition is Wurmbrand's (1998) discussion of "restructuring" phenomena in German. She argues persuasively that infinitival complements to restructuring predicates are significantly smaller than CP or TP, and may be bare VPs. She then shows that although a bare VP complement to a restructuring predicate may contain a verb and its object, the licensing of object case does not take place within this domain, but instead depends on the case properties of the restructuring predicate. The link between the "smallness" of the embedded VP and the absence of object case licensing in that domain forms a good argument that object case licensing depends on VP-external structure.

³⁸ A reviewer notes that if Case movement is feature movement, we face the task of explaining a set of differences between the feature movement posited for objective case checking and the feature movement

posited between T and the associate NP in the existential *there* construction. Several papers in Lasnik (1999) offer evidence from binding theory, scope and other phenomena that objective case-marked NPs in English are interpreted outside the VP in which they originate. The same is not true of the nominative case-marked NPs in the *there* construction. Its scope is confined to the VP (as in *There do not seem to be many linguists in the room*, where *many* unambiguously has scope lower than negation), as are its binding possibilities (**There seem to themselves to be some linguists in the room.*). I will not try to resolve this difficulty here, but will leave the question open.

³⁹ The only slender hope left for "case movement" theories of ACD, arises if we can argue that object Case checking in general involves phrasal movement (overt or covert) — but that for some reason the pattern of *wh*-movement under study restricts object Case checking to feature movement. I will not explore this possibility here.

⁴⁰ As Kai von Stechow (personal communication) notes, the analyses of (74) and (75) actually coincide if one adopts the proposal by Ludlow, Larson, and den Dikken (1998) (reminiscent of generative semantics) that posits clausal structure in the complement position of (75).

⁴¹ One weakness in this argument is the possibility that the case position for the embedded subject lies outside the scope of *want* and licenses ACD, but the *de dicto* reading in (74a) results from either reconstruction or the scope-neutrality of feature movement for case reasons. (The latter is a suggestion of Kai von Stechow's, personal communication.) The reconstruction option in (74b) might be blocked by the ACD itself. The argument from (61) does not have a comparable weakness.

⁴² In my (1987) theory of the Superiority Effect, the very existence of D-linked counterexamples constituted an argument for the absence of covert movement of *wh* *I*, since that paper assumed that the Superiority Effect arises from illicit configurations of covert movement, rather than from illicit choices for overt movement.

⁴³Chomsky's (1995, chapter 4) theory captures the inviolability of AC by building the condition into the definition of the operation "Move".

⁴⁴An important terminological note: I will continue to use the informal term "*wh* 1-in-situ" to designate the highest of a group of *wh*-phrases when it has not undergone overt phrasal movement — regardless of whether *wh*-feature movement has taken place from it. "In situ" in this study means "pronounced in situ".

⁴⁵Covert phrasal *wh*-movement is not clause-bound. If it were clause-bound, examples like (i) would violate the Multiple-Specifier requirement, contrary to fact, and would have the same status as (ii):

(i) Who did Sue think bought what?

(ii) *What did Sue think who bought?

This result conflicts with a claim by Baltin (1986), who uses data from ACD as part of an argument that *wh*-in-situ does not undergo full phrasal movement of a sort that could resolve ACD. The example that he cites is (iii) (his example (14)):

(iii) Who thought that Fred read how many of the books that Bill did .

According to Baltin, the ellipsis in (iii) may take the lower VP (...*read t*) as its antecedent, but not the higher VP (*...*thought that he read t*).

I believe that Baltin's factual claim is not correct (even though it is assented to by Kennedy 1997, 670, who suggests an account). True, there may be a preference for lower-clause interpretation of the ellipsis in examples like (iii), but the upper-clause interpretation does not seem to me to be excluded. Furthermore, no sensation of deviance arises when the context *forces* an upper-clause interpretation of the ellipsis — as in (iv) and (v), where the modal stranded before the ellipsis site matches a matrix modal and does not match material from the lower clause:

(iv) Which student will claim that you visited which woman that Mary (also) will ?

(v) Who might claim that he visited which city that Mary (also) might?

It seems to me that the correlation between wide scope and the possibility of wide-scope ACD resolution holds up even in these cases.

⁴⁶ Barss (1999) claims that D-linked questions that (appear to) violate the Superiority Condition lack the pair-list reading otherwise available to multiple questions. If he is correct, for example, a question like *Which book did which person buy?* can only seek for its answer a single person-book pair, such that the person bought the book — while the corresponding question that does not resemble a Superiority violation *Which person bought which book* may seek a set of person-book pairs as its answer (*Mary bought this book, Bill bought that book, etc.*). Barss's claim, if correct, would support my contention that apparent Superiority violations are possible only in the presence of a special type of interrogative complementizer, since we would be able to attribute a special semantics to this complementizer. Unfortunately, both types of multiple questions seem to me to have a pair-list reading — particularly when contrasted with examples considered later in this monograph (section 5.1) that strongly exclude the pair-list reading.

⁴⁷ It would be useful to acquire evidence that could confirm or refute our claim that *wh*-in-situ in Bulgarian examples like (80a) (on the acceptable parse) does not undergo covert phrasal movement. To test this, we would have to develop evidence for something like VP-ellipsis in Bulgarian (which is not inconceivable: Stjepanović argues for the existence of VP-ellipsis in closely related Serbo-Croatian), and tests that can ensure that scrambling of complex NPs out of the VP has not taken place overtly. I have not undertaken this work.

⁴⁸This possibility makes the speakers who do *not* report any residual Superiority Effect into the unexplained case, but perhaps these speakers are simply less sensitive to (37) than are some others.

⁴⁹ Roumyana Izvorski (personal communication) notes that the contrast is even clearer if the optional interrogative complementizer *li* is added:

- (i) **Na kogo koj li dade kakvo?
- (ii) **Kakvo koj li dade na kogo?
- (iii) *Na kogo kakvo li dade koj?

⁵⁰ Curiously, if *koj* is placed to the left of the verb *dade* in (84), the result — while still unacceptable — improves once again:

- (i) a. ?*Na kogo kakvo koj dade?
- b. ?*Kakvo na kogo koj dade?

Perhaps the examples in (i) are more acceptable than (83) because they, like (84), have a parse in which *koj* has not undergone phrasal *wh*-movement, but remains (for pronunciation purposes, at least) in SPEC, TP:

- (ii) a. ?*[CP Na kogo kakvo C_{m-spec} [TP koj dade __ __]]?
- b. ?*[CP Kakvo na kogo C_{m-spec} [TP koj dade __ __]]?

The difficulty with this proposal is the fact that verb movement to C is normally obligatory in Bulgarian *wh*-questions, as in English:

- (iii) a. *Na kogo kakvo Ivan dade.
to whom what Ivan gave
- b. Na kogo kakvo dade Ivan.
to whom what gave Ivan

All things being equal, this should render the examples in (i) *worse* than (84) — not better. I leave this as a puzzle. I am grateful to Roumyana Izvorski for her judgments and to an anonymous reviewer for raising the issue.

⁵¹The final *wh*-phrase in (85b) is included only to balance the corresponding phrase in (85a), but the judgment should remain constant even if it is eliminated, e.g. *What did who order [who that Mary (also) did] to buy* ___ .

⁵² As a reviewer for MIT Press pointed out to me, a proposal similar to this has been recently advanced by Barss (1999). Barss suggests that the *wh*-feature on interrogative ("*wh*") phrases is optional, so that when a lower *wh*-phrase moves to SPEC,CP over a higher *wh*-phrase left in situ, it is only because the higher phrase lacks the *wh*-feature. Barss does not discuss what ties this possibility to D-linking, but one might imagine extending his proposal as suggested in this (straw-man) paragraph.

⁵³ In the present context, **Bošković's** (1998; in press) alternative to the PMC (see footnote 28) will do just as well.

⁵⁴It is in this limited sense that AC is "violable" even in our theory, though formally the "violations" actually accord with a general format for constraint satisfaction.

⁵⁵ For me, *whom* is the obligatory form for object *wh*-in-situ, while *who* is the form for both objects and subjects when moved. As far as I can tell, the choice of *who* and *whom* does not make any difference in

this paradigm, except that for any given speaker, the morphological rules followed by that speaker should be used in constructing examples.

⁵⁶It goes without saying that the Principle of Minimal Compliance is conceptually surprising in its own right. Obviously, we would like to understand why the PMC holds. Nonetheless, for present purposes it is sufficient to view the PMC as a placeholder for a future explanation of the phenomenon. As long as the placeholder correctly identifies a real phenomenon, the significance of this section lies in the link it draws between otherwise distinct sets of facts in English and Bulgarian.

⁵⁷Of course, it could be argued that the notion of constraint ranking in Optimality Theory (OT) is simply a different notion of what it means to be an "apparent" violation of a constraint. The issue, in any case, is finding the correct account of the apparent violations. In lectures at the 1997 LSA Linguistic Institute at Cornell, I presented an OT account of many of the phenomena subsequently analyzed as they are in the current study. The proposal was a precursor to the proposals in this study, and included many of the components of the present discussion, including the Multiple Specifier requirement (with a D-linking exception) and AC. The proposal also posited a "MOVEMENT FAITHFULNESS" constraint which favors derivations whose patterns of *wh*-movement and non-movement match those of an "input" derivation. (In OT phonology, among the constraints governing pronunciation are some that favor "Faithfulness" to an input form - in just this fashion. Only constraints ranked higher than a given Faithfulness constraint can enforce a deviation from the input with respect to the property singled out by the Faithfulness constraint.) MOVEMENT FAITHFULNESS strongly resembles the proposal of Legendre, Smolensky and Wilson (1998).

The ranking posited was:

MULTIPLE SPECIFIER REQUIREMENT >> MOVEMENT FAITHFULNESS >> ATTRACT CLOSEST

"MOVEMENT FAITHFULNESS" occupies the same logical place in this OT proposal as the feature movement option occupies in the proposal presented in this study. It is what licenses apparent violations of

ATTRACT CLOSEST, while not tolerating violations of the MULTIPLE SPECIFIER REQUIREMENT. The system handled the data we have discussed as follows:

- Given an input derivation that violates the MULTIPLE SPECIFIER REQUIREMENT (e.g. **What did who buy?*, with no covert movement of *who*), the MULTIPLE SPECIFIER REQUIREMENT will reject it in favor of those alternative derivations in which at least two *wh*-phrases undergo movement, and ATTRACT CLOSEST will chose the one in which the highest *wh*-phrase moves first.
- Given a binary question in which the MULTIPLE SPECIFIER REQUIREMENT is satisfied in the input (e.g. *Who bought what?* with covert movement of *what*; or **What did who buy?* with covert movement of *who*), MOVEMENT FAITHFULNESS will allow the input and all alternatives that share the same pattern of movement and non-movement. ATTRACT CLOSEST will favor the one in which the highest *wh*-phrase moves first, even if that property was not met in the input.
- Given a non-binary multiple question in which the MULTIPLE SPECIFIER requirement is satisfied, but *wh*₁ is not moved (e.g. *what did who give to whom*, with covert movement only of *whom*), FAITHFULNESS will exclude derivations that satisfy ATTRACT CLOSEST, since they involve a movement pattern distinct from the input, and ATTRACT CLOSEST, by the logic of OT, will be violated by the winning candidate.

To explain the facts just discussed in the main text, it was necessary to stipulate that ATTRACT CLOSEST is non-gradient. If presented only with candidates in which *wh*₁ remains in situ, so that the closest *wh*-phrase has not moved, the ATTRACT CLOSEST constraint "gave up", and enforced no preference among other movement possibilities. This property of ATTRACT CLOSEST could not be brought together with data handled by Richards' PMC, which I take to be a notable disadvantage of this OT theory in comparison to its non-OT sibling defended in this study.

⁵⁸We are not, of course, providing a general argument against ranking and violability within the theory of movement — merely suggesting that one potential argument from Superiority violations does not support an OT view. Consequently, it might turn out that movement theory invokes a system in which constraints are ranked and violable that also countenances *wh*-feature movement as the first instance of movement in an apparent Superiority violation. Apparent Superiority violations would just not play a special role in arguing for the OT character of this system.

⁵⁹It is probably important to emphasize that the operation of *wh*-feature movement was not posited simply in order to maintain the view that AC is inviolable. It is also crucial to remember that the PMC considerations discussed in this section also converge on this conclusion.

⁶⁰One might add to the list semi-archaic *whence/thence* and *whither/thither*. I am grateful to Morris Halle for discussion of the morphological and phonological issues connected with the English *wh*-words. The parallelisms noted in (91) are, of course, not an original observation, but have been frequently noted in philological and other studies.

⁶¹The phonological remarks here should be understood as suggestions that the relation between putative underlying form and the surface form could be the result of a natural process. I am aware that serious phonology imposes a higher standard of proof than my speculations meet.

⁶²The forms with *so*- are demonstratives distal with respect to speaker. The forms with *a*- are distal with respect to speaker and hearer.

⁶³This view might help us understand Chomsky's observation that certain instances of feature-movement take other features along with them (as "free riders", in his terminology). The packaging of features in this way may reflect the distribution of morphemes in the language. Thus, for example, if case and number are packaged together, as they typically are in Indo-European languages, we might expect that feature movement of case also moves number.

⁶⁴The *wh*-morpheme discussed here seems to be quite distinct from the "Q-morpheme" on *wh*-phrases discussed at length by Hagstrom (1998). It is Hagstrom's hypothesis that the Q-morpheme is a quantifier over choice functions which moves to C to generate the observed interpretation for interrogative clauses. He argues that the Japanese morpheme *-ka*, usually identified as an interrogative complementizer, is actually this morpheme after overt movement from a *wh*-phrase. The properties he identifies as properties of *-ka*, however, are quite orthogonal to the properties required of our *wh*-morpheme. For one thing, it appears to be present on only one *wh*-phrase in a multiple question, with some evidence that it appears only on the *lowest* of a group of *wh*-phrases when a multiple question receives a "pair-list" reading. It is strikingly hard to identify a semantic function for the *wh*-feature discussed here, at least within the picture painted by Hagstrom. I leave this key issue for future work.

⁶⁵ The mechanism by which the *wh*-feature percolates is unclear. In languages like Quechua and Basque, where clauses undergo pied piping, the feature apparently percolates to larger constituents than is possible in English. Furthermore, the English conventions appear to be *sui generis*, which makes it hard to understand in terms of deeper principles. It appears as though linear peripherality plays an important role. The basic rule for questions in English seems to be the following:

(i) **Rule for identifying an interrogative *wh*-phrase.**

A *wh*-phrase has the form: [(Adv) (P) *wh*...].

If the condition in (i) is met, the *wh*-feature can be embedded indefinitely far in the structure:

- (ii) a. Mary wondered [*whose* brother's cousin's shoes] we had found in the playground.
b. *Mary wondered [the cousin of *whose* brother's shoes] we had found in the playground.

- (iii) a. Mary wondered [in *whose* brother's cousin's shoes] we had been walking.
b. *Mary wondered [in the cousin of *whose* brother's shoes] we had been walking.

- (iv) a. Mary wondered [*how* many people's rights] the government had trampled on.
b. Mary wondered [on *how* many people's rights] the government had trampled.
c. Mary wondered [exactly *how* many people's rights] the government had trampled on.
d. *Mary wondered [the rights of *how* many people] the government had trampled on.

- (v) a. Mary wondered [*which* violinist's cadenzas] Kremer intended to play.
b. *Mary wondered [the cadenzas by *which* violinist] Kremer intended to play.

Perhaps the *wh*-feature adheres to a syntactic phrase by virtue of occupying the left periphery of the first phonological phrase of that constituent (on the assumption that leading prepositions and adverbs may have clitic status and do not need to belong to this phrase).

⁶⁶In Optimality Theory terms, we might think of the two principles as tied constraints (Pesetsky 1997, 1998).

⁶⁷ For the details of the relevant conditions, Ochi follows Takahashi (1994). He argues, for example, that adjunction to individual members of multi-member chains is forbidden by a general "uniformity" requirement on operations that affect chains. If subjects in SPEC,IP have moved from a lower position, this requirement will predict the islandhood of subjects. Adjunct constituents that are semantically interpreted via coordination, including VP-adjuncts and relative clauses, also bar adjunction. Thus, the islandhood of VP-adjuncts and relative clauses is also predicted.

On one key point, Ochi's hypotheses and ours are incompatible. Ochi follows Takahashi in arguing that the constraint on generalized pied piping that requires extremely local landing sites is "mover-oriented" rather than "attractor-oriented". He suggests, in fact, that the relevant constraint is Shortest Move. If this were the case, we would expect the PMC to *never* license violations of island conditions, for the same reason the PMC never licenses violations of the "tucking in" requirement (as discussed in the final paragraph of section 2.2). There is no "tax relief" from tucking in because Shortest Move is not a constraint governing the distance between an Attractor and Attractee, but rather a constraint that chooses the exact position near a given Attractor to which copying takes place. There is, as we have seen, "tax relief" from island conditions. Consequently, the two phenomena cannot both be traced back to Shortest Move.

⁶⁸If we were to substitute Agree for Feature Movement in our account of *wh*-configurations, the distribution of AC and subjacency effects might then be explained as follows:

1. Feature Matching obeys AC, except insofar as the PMC allows exceptions.
2. Movement is subject to island conditions (subjacency), except insofar as the PMC allows exceptions.
3. Movement of α to a specifier of β is possible only if a feature of α has been matched with a feature of β .

⁶⁹Needless to say, contexts in which the "good" examples are natural are often rather *recherché*. For example, (99a) requires a context in which the speaker knows that certain sensitive issues are to be avoided when speaking with certain diplomats,

⁷⁰This observation is a counterpart to a very similar point made by Beck (1996) for German.

⁷¹There is also the (possibly distinct) functional reading, which I will ignore here. (It will ultimately be important to investigate these results in light of Chierchia's hypothesis that the pair-list reading is a species of functional reading.)

⁷² For very similar data and discussion, see Barss (1999). Barss goes on to argue that the "wide-scope answering pattern" in examples like (108) does not include a pair-list reading for the two *wh*-phrases. Instead, the two *wh*-phrases are claimed to have an independent interpretation which, were it not for the universal quantifier *every*, would invite a single pair response. This is in keeping with Barss's claim (disputed in footnote 46 above) that apparent Superiority violations of all sorts generally lack the pair-list reading -- a claim at odds with the much sharper contrast between pair-list readings and their absence found with the Intervention Effect.

⁷³ The illuminating contrast between *every* and *almost every* is adapted from Beck (1996), though she does not offer the explanation suggested here.

⁷⁴ My statement contains the phrase "a clause interpreted as a question", because if the question is embedded in further structure, wide scope readings are available that do end up expressing requests for incomplete information or requests for no answer at all. Example (104) provides a good example. On the other hand, the observation in (111) extends beyond the content of SPEC,CP. For example, (i) cannot be a request for a partial list of people:

(i) #In part, who did Mary invite to the party?

—but (ii) may be a description of a partial answer:

(ii) John told me in part who Mary invited to the party.

⁷⁵ One might imagine a question that requests an absence of an answer in a discourse in which it was conjoined with other questions. For example, one might imagine being able to use (i) below to express the though otherwise expressed by *Tell me which book Mary gave to John, but don't tell me which book anyone from Paris gave to John.*

- (i) #Which book did Mary give to John, but which book did no one from Paris give to John?

The absence of this possibility is therefore of interest. Another unaskable question of the same vein is:

- (ii) Which problems did no one except Mary solve?

The question in (ii) can be a request for the names of the problems such that Mary solved them and no one else solved them. This is the narrow-scope reading for *no one except Mary*. The wide-scope reading for *no one* would amount to a request to keep quiet about who besides Mary solved problems, while providing the names of problems that Mary (possibly along with others) solved. A report of such a question, which would show *no one except Mary* taking higher-clause scope, might be something like (iii):

- (iii) He was only interested in Mary's performance on the exam. He asked which problems no one except Mary solved.

Thinking about (iii) helps bring out the interest of the missing reading in (ii).

⁷⁶I embed the question under *tell me* because *each NP* does not easily allow wide scope over *wh* in a matrix question:

- (i) Which book did each student buy __?(? with a pair-list reading)

In this respect, it behaves a bit like the quantifiers in (113). Perhaps "aggressively requesting" complete information within a question clause is not allowed, just as requests for incomplete information are not allowed. What the notion of "aggressively requesting" might be, such that it distinguishes *every* from *each* in the correct way, is unclear.

⁷⁷ Counterevidence to this conjecture comes from Bulgarian, where the one consultant I have asked (Roumyana Izvorski) reports neither marked unacceptability nor loss of the pair-list reading when a D-linked

wh-in-situ is separated from interrogative C by negation. Recall that Bulgarian allows *wh*-phrases to appear in situ when they are D-linked. We argued in section 4.2 that these phrases are truly in situ, in the sense that their link to C involves feature movement, rather than phrasal movement. Izvorski does report that (i) is somewhat worse than (ii):

(i) [?]Koj ne dade kakvo na Stefan?
who not gave what to Stefan

(ii) Koj dade kakvo na Stefan?

— but also reports that (iii) is a bit worse than (iv), making it impossible to confidently attribute the contrast between (i) and (ii) to the Intervention Effect discussed in the text:

(iii) [?]Koj kakvo ne dade na Stefan?
who what not gave to Stefan

(iv) Koj kakvo dade na Stefan?

It is conceivable that Bulgarian reveals a genuine flaw in our characterization either of the special effects of D-linking or of the Intervention Effect. It is also possible that other factors are masking the expected effect. For example, the degradation observed in (iii) might indicate that negation in Bulgarian has scopal properties (e.g. obligatory widest scope) that, on the one hand, interfere with multiple questions altogether and, on the other hand, by-pass the Intervention Effect. I leave this as an open problem for the hypotheses advanced in this work.

⁷⁸ If the "Agree" alternative to Feature Movement is adopted, then it is the mere fact that the restriction remains within TP (the "nuclear scope"), rather than the existence of separation, that produces the effect. All our generalizations remain intact.

⁷⁹ Beck also observes an intervention effect with the German "partial movement" construction. The partial movement construction is a form of interrogative whose SPEC,CP contains an invariant *was* 'what', and the *wh*-phrase that would otherwise occupy that SPEC,CP is instead found in some lower SPEC,CP, for example (i):

- (i) Was glaubst du mit wem Hans gesprochen hat?
what believe you with whom Hans spoken has
'Who do you believe that Hans spoke to?'

The intervention effect can be shown with examples like the following, from Beck:

- (ii) *Was glaubst du nicht mit wem Hans gesprochen hat?
what believe you not with whom Hans spoken has
'Who do you not believe that Hans spoke to?'
- (iii) ??Was glaubt niemand wen Karl gesehen hat?
what believes nobody whom Karl seen has
'Who does nobody believe that Karl has seen?'
- (iv) ??Was glaubt fast jeder wen Karl gesehen hat?
what believes almost everyone whom Karl seen has
'Who does almost everyone believe Karl has seen?'

Cheng (1997; 1999) has argued that partial movement in German involves phrasal movement followed by feature movement of the *wh*-feature (a conclusion that may dovetail with Cole and Hermon's 1998 demonstration that the relationship between the *wh*-phrase and the invariant interrogative marker obeys

island conditions in Malay). If we adapt this hypothesis to our overall picture, we might handle the effects seen in (ii)-(iv) in the same way as the effects discussed in the text.

⁸⁰The well-known constraint on separation of *combien* 'how many' from its restriction discovered by Obenauer (1984) is presumably an instance of the Intervention Effect in a separation construction (see also Rizzi 1990):

- (i) [Combien de véhicules] a-t-il conduit ___?
how-many of cars did he drive
- (ii) Combien a-t-il conduit [___ de véhicules]?
how-many did-he drive of cars
- (iii) [Combien de véhicules] n'a-t-il pas conduit ___?
how-many of cars did-he not drive
- (iv) *Combien n'a-t-il pas conduit [___ de véhicules]?
how-many did-he not drive of cars

As has been noted by Longobardi (1986), among others, similar effects can be detected in English, when attention is paid to the relative scope of the *many NP* portion of *how many* questions with respect to scope markers. This shows that the Intervention Effect is sensitive to reconstruction phenomena, not a surprise for a condition on LF.

⁸¹I am extremely grateful to Jim McCloskey for collecting these data, which represent the judgments of a number of West Ulster English speakers gathered in San Francisco. He writes (personal communication):

"The effect is weakest for simple negation as in [(130)a]. Everyone was very clear about [(130b)], (d) and (e), but about [(130a)] there was more doubt and more debate (people began making up elaborate scenarios about shopping lists etc). [(130b)], (d) and (e)]are uninterpretable, but interpretations can be constructed for [(130a)]. The interpretation which is available is, in my opinion, one in which the *wh*-phrase is strongly d-linked.

The interpretation that is available for [(130c)] is, in my opinion anyway (though the people I was working with couldn't articulate this), one in which the subject quantifier has wider scope than the *wh*-phrase (i.e. for every relevant person, give me a complete listing of the things they got on Christmas morning). The other scope (*wh*>*every*) is incoherent with [(130c)]."

As McCloskey (1997) notes, the phenomenon of *wh-all* separation is of more general interest, especially because the *all* may be stranded in what appear to be intermediate specifiers of CP. The reader is referred to McCloskey's work for more discussion.

⁸²As already noted, one also wants to know why the Intervention Effect exists. Some considerations advanced by Reinhart (1998) suggest a possible line of attack, though some rethinking is necessary. According to Reinhart, Pesetsky (1986) proposed that all D-linked *wh*-in-situ are interpreted in situ, rather than by covert phrasal movement. She then proceeds to show that on the simplest assumptions about in-situ interpretation of *wh*-phrases, this suggestion fails. In particular, in a downward entailing environment, if we interpret the restriction on a *wh*-phrase as a conjunction at the lowest possible level, multiple questions should allow all manner of silly answers, e.g. *Lucie will invite Donald Duck, and Bill will invite Mickey Mouse* in (i)-(ii):

(i) Who will be offended if we invite which philosopher?

WRONG:

For which $\langle x, y \rangle$, if we invite y and y is a philosopher, then x will be offended?

Answer: Lucie - Donald Duck; Bill - Mickey Mouse; ...

i.e. Lucie will be offended if [DD is a philosopher and] we invite DD.

RIGHT:

For which $\langle x,y \rangle$, y is a philosopher, and if we invite y , x will be offended.

(ii) Which linguist didn't invite which philosopher?

WRONG:

For which $\langle x,y \rangle$, x is a linguist & [it is not the case that y is a philosopher and x invited y]

Answer: Lucy - Donald Duck; Bill - Mickey Mouse

RIGHT:

For which $\langle x,y \rangle$, x is a linguist & y is a philosopher & [it is not the case that x invited y]

On the one hand, Reinhart is not correct about the details of my (1986) proposal. In that paper, I argued that D-linked *wh may* (not "must") be interpreted in situ. Nonetheless, one wonders if — in those cases where the only parse for a multiple question involves *wh*-feature movement — the deviance attributed to the "Intervention Effect" might actually be attributable to the "Donald Duck" problem. It is necessary to imagine that questions which fail to control their answers are actually felt as deviant — which is, of course, the sticking point here. Nonetheless, if this conjecture is correct, Reinhart's observation may have (semi-inadvertently) provided the key to understanding the Intervention Effect. That is, she may be right about the interpretive difficulties of in-situ restrictions, but wrong to suppose that this poses a problem. It might instead offer a solution to a problem: why the Intervention Effect holds. Unfortunately, this proposal does not cover cases of the Intervention Effect involving interveners like *every*, which are not downward entailing for their nuclear scope. If these cases are of a piece with the others, Reinhart's suggestion is not apropos here. Honcoop (1988) offers an alternative that might be investigated in its place.

⁸³In principle, any sort of Antecedent-Contained Anaphora could be used to discover whether a restriction on a quantifier or *wh*-phrase exits its VP by means of covert phrasal movement. For example, one might hope that antecedent-contained clausal anaphora as in (i) below can be taken as a probe of the appropriate sort, where *that* seems to take a clause of the form *PRO invite t* as its antecedent:(i) John wanted to invite exactly those people who didn't want him to do that.If this type of anaphora could provide a test akin to ACD, we could take this test into languages like German that do not allow VP-ellipsis. Unfortunately, this type of anaphora differs from VP-ellipsis in allowing a variety of contextually salient antecedents, including entirely non-linguistic antecedents. For example, one can utter (ii) with *that* taking a visually presented action as its antecedent:(ii) I wish you wouldn't do that!As discussed earlier, the inability of VP-ellipsis to take antecedents not present in the linguistic environment is crucial to the link between ACD and covert phrasal movement.In fact, such a link seems to be absent for clausal anaphora. An argument that ACD could only be resolved by movement came from the facts from Larson and May (1990) discussed in (5). In (5), it was crucial that the elided VP in (iii) could not take the higher VP headed by *refuse* as its antecedent without the *every*-phrase taking wide scope:

(iii) John refused to visit [every city Mary did [vp]]

cannot mean:

John refused to do the following thing: visit every city Mary refused to visit.

Since the referent of clausal anaphora does not have to be present in the linguistic environment, it can be merely "evoked" by the linguistic environment. Consequently, we do not seem to find the link between scope and antecedent-contained anaphora seen in (iii):

(iv) John refused to call every employee who wanted him to do that.

can mean:

John refused to do the following thing: call every employee who wanted John to refuse to call him.

⁸⁴The observation is due to Beck (1996), who does not, however, develop an account of the contrast with English.

⁸⁵Beck (1996) offers a more nuanced description of the unacceptability judgment: "The '??' means that the data are incomprehensible (uninterpretable) rather than simply ungrammatical. I would accordingly ask native speakers to try and interpret the sentences, not simply judge whether they 'sound bad'." Some English speakers describe much the same effect. Though Beck (1996) does not mention the improved acceptability with a single pair reading, it was Beck (personal communication) who first mentioned the fact to me, in the context of German.

⁸⁶If the "Agree" variant is correct, they have in common with Separation constructions the presence of the restriction on quantification inside the nuclear "scope" (the main body of the clause). See footnote 78.

⁸⁷For example, we might be tempted to argue that the presence of separation constructions in German that are not allowed in English indicates a greater overall "separability" of the components of *wh*-phrases. If feature movement or morpheme movement is favored over phrasal movement, this greater overall separability might yield (138). As far as I can tell, however, languages and dialects differ substantially in the availability of separation constructions in ways that do not correlate with (138). The Slavic languages, for example, show a wide variety of separation constructions, yet allow multiple phrasal movement, as we have seen in Bulgarian.

⁸⁸I state (141) informally because its place in the overall scheme of things is underdetermined by the facts at our disposal. The rule of complementizer choice might be part of the theory of Merger in a bottom-to-top derivation, in which case, a more careful formulation might be the following:

Consider a TP that contains n instances of wh not moved to its scope position. An interrogative complementizer with specifier potential m may merge with only if the lexicon contains no interrogative complementizer ' with specifier potential p , such that

$$m < p - n.$$

This formulation has the odd result of ruling out extraction of a wh -phrase from a non-multiple embedded question (but not from an embedded multiple question) — that is, it attributes one, but not all, of the cases of the wh -island condition to the principle of complementizer merger. If this consequence is false, one should look for alternative hypotheses.

For example, the rule could also be understood as a filter on wh -movement, along the following lines:

An interrogative complementizer with specifier potential m may attract n instances of wh only if the lexicon contains no interrogative complementizer ' with specifier potential p , such that $m < p - n$.

I leave the investigation of these possibilities and others for further research.

⁸⁹ The Complementizer Choice Rule as stated in (141) is compatible with the treatment of D-linked questions offered in (79). According to (79), the central peculiarity of D-linked wh lies in the fact that feature movement from a D-linked wh can satisfy a requirement of C_{m-spec} that is otherwise satisfiable only by phrasal movement. I also considered an alternative to (79) according to which the exceptional feature of questions with D-linking is different: they allow a multiple-question interpretation with C_{l-spec} . If the alternative is on the right track, we do not need (79), but D-linked questions are some kind of exception to (141). As noted earlier, more work is necessary before we can understand (rather than merely describe) why

wh with the semantics of D-linking has the syntactic peculiarities that it does. (I owe these points to discussion with Norvin Richards.)

⁹⁰ The truth is more complicated. In addition to cases that we turn to in the next paragraph, certain types of *wh*-phrases do seem to yield a Superiority effect. For example, as Armin Mester (personal communication 1986) pointed out to me, phrases of the type that Pesetsky (1987) called "aggressively non-D-linked" yield something like a Superiority Effect. Judgments differ, but many speakers report a contrast between examples like (i) and (ii) — a discovery independently reported by Wiltschko (1997), from whom I take these examples:

(i) Wer zum Teufel hat wen gesehen?

who to-the devil has whom seen

'Who the devil saw whom?'

(ii) ?*Wen zum Teufel hat wer gesehen?

whom to-the devil has who seen

Conceivably it is wrong to argue, as I have, that German lacks C_{m-spec} entirely. Instead, it is possible that German reserves C_{m-spec} in multiple questions for explicitly non-D-linked questions (with C_{I-spec} as a default) as a mirror image of the fact that English reserves C_{I-spec} in multiple questions for explicitly D-linked questions (with C_{m-spec} as a default).

There are a few other ways to bring out a Superiority effect in German. Haider (1998) notes that it is unacceptable to overtly move a *wh*-phrase over another *wh*-phrase when the two are identical. The verb *beten* 'ask', unlike *versprechen* 'promise', takes an accusative object:

- (i) Wen hat er denn *wem* versprochen [davon __ abzuhalten] ?
whom-ACC has he prt whom-DAT promised from-it __ to keep away
'Who did he promise to keep whom away from it.'
[translation reverses the movement pattern]

- (ii) *Wen hat er denn *wen* gebeten [davon __ abzuhalten] ?
whom-ACC has he who-ACC asked from it to keep away

I have no account for this fact.

^wiltshko (1997) offers another interesting problem for this conclusion. Consider a multiple question with two *wh*-phrases in which both phrases are taken to range over an identical set containing two individuals. In German, such a question yields a Superiority Effect:

- (i) Peter is walking his stubborn dog on the leash. The dog is dragging really hard in the direction of his favorite tree.
- a. Wer führt denn hier *wen* an der Leine?
who lead (prt) here whom on the leash?
'Who is leading whom here on the leash?'
- b. *Wen führt denn hier wer an der Leine?

(ii) I have heard that Peter and Mary had an affair. Can you tell me:

a. Wer hat wen verführt?

who has whom seduced

'Who seduced whom?'

b. *Wen hat wer verführt?

(iii) I am sure that Peter and Mary must have talked to each other on the phone:

a. Weißt du wer wen angerufen hat?

know you who whom called has?

'Do you know who called whom?'

b. *Weißt du wen wer angerufen hat?

One can imagine ways in which such questions might differ syntactically from other questions to yield these effects, but the solutions that come to my mind are mechanical. For example, it can be readily observed in English that the *wh*-phrase that moves overtly in other multiple questions has a special status as the "sorting key" for the answer (Kuno 1982). Thus, a person who asks *Which person bought which book?* normally expects an answer that exhausts some domain of people, but not necessarily any domain of books, while the person who asks *Which book did which person buy?* normally has the opposite expectation. No such difference can be imagined in Wiltschko's cases, where the expected answer naturally exhausts the relevant set no matter what. Suppose it is the case that the requirement of a *wh*-specifier needs to be satisfied as early as possible in the derivation, except when delaying the requirement makes an interpretive difference. Such a condition on the patterns of movement might handle Wiltschko's observation, since the (b) examples, if we are correct, show feature movement preceding phrasal movement.

⁹²Previous proposals link the absence of Superiority effects with subject *wh*-in-situ to the absence of *that*-trace effects in German (for example, Haider, as cited above; see also Müller 1995). The proposal presented here does not make this link.

⁹³I am grateful to Irene Heim (personal communication) for this observation. Furthermore, if a higher clause is added to (147) so that even the (a) example involves *wh*-movement from an embedded clause, three of five speakers polled found that the difference disappeared. Two speakers still felt that the (b) example remained less acceptable, but one volunteered the observation that *wh*-extraction from one finite clause (as in the (a) example) is in general more acceptable for her than extraction from two finite clauses (as in the (b) example):

- (i) a. ??/*Wer denkst du, dass __ glaubte, dass der Peter wen vorstellte?
 who think you that believes that the Peter whom introduced
- b. *Wen denkst du, dass wer glaubte, dass der Peter ihr __ vorstellte?

A particularly interesting case might be provided by Dutch, if the status of the Superiority Effect for Dutch speakers can be cleared up. For some speakers, at least, Dutch presents judgments on Superiority and the Intervention Effect that are identical (or nearly so) to those reported for German, with three exceptions. First, Dutch lacks a scrambling process that places objects to the left of subjects (Johnson and Tomioka 1997). Second, interrogative *wh*-phrases move freely from embedded finite clauses. Third, Dutch fails to display the contrast in (147) (Sjef Barbiers, personal communication). For these speakers, one might argue that scrambling does not provide the explanation for the phenomena discussed here. Koster (1986, 204) and Aoun, Hornstein, Lightfoot and Weinberg (1987; ex. (61)) do note a contrast between Dutch equivalents of *who saw what* and *what (did) who see* that looks like a Superiority Effect. While the judgment reported by Aoun et al. is assented to by van de Koot (1988), Koster himself expresses some doubts about the strength of the effect (p. 236, fn. 12). Concerning (ii) below, he writes "I find it difficult to make up my mind...The sentence does not seem entirely ungrammatical, which again casts doubt on the generality of the

proposed explanations [in terms of an Superiority Effect]". The absence of a star is Koster's, though earlier he stars a comparable example:

- (ii) Wat heeft wie gekocht.
 What has who bought

As Günter Grewendorf (personal communication) has pointed out, it is also worth noting that Bulgarian shows Superiority effects despite also allowing scrambling, as perhaps (according to Grewendorf) does Turkish. Japanese (discussed in the final sections of this work) provides a similar case. Thus there does not appear to be a correlation between the availability of scrambling and the apparent absence of Superiority effects.

⁹⁴ Beck (1996) notes that *wh* may appear to the left of a subject if it is quantified—perhaps a (mysterious) instance of the otherwise forbidden scrambling of *wh*-phrases, as suggested by Sauerland 1996. If so, it is a counterexample to the generalization given in the text:

Wer hat gesagt, dass wen niemand mag?
who has said, that whom-ACC nobody likes
'Who said that nobody likes whom?'

⁹⁵ The translations of (150a-b) and (151a-b) are, of course, not fully acceptable English, since they show the Superiority Effect — but there are no fully acceptable alternatives.

⁹⁶ Kai von Stechow (p.c.) reports that extraction from the embedded clause *without* separation is degraded, to a level better than (b) but worse than (a). I do not know why this is the case. It is also worth noting that the *who of NP* construction does not display the same contrast as that seen in (152) and (153), this might be due to the possibility of taking the PP to modify the clause ('Among the musicians,...') in clause-initial position.

⁹⁷ I quote Fanselow's data here. Some speakers dislike extraction from the object of certain verbs, including *kritisieren*. Such speakers should judge comparable examples with a variety of main verbs.

⁹⁸ Günther Grewendorf (personal communication) notes that many German speakers entirely lack Weak Crossover effects with *wh*-phrases, so a strong burden would fall on the claim that they are generally stronger with feature movement. Wiltschko (1997, 125) notes speaker variation in weak crossover judgments. I will not tackle this problem here.

⁹⁹ Also relevant are examples like the following, which should be compared to the series of examples discussed in section 3.1:

?*I need to know [which girl that Mary did] Sue ordered which boy to congratulate ____.

(*i.e.* I need to know the girl-boy pairs such that both Sue and Mary ordered the boy of the pair to congratulate the girl of the pair.)

The example cannot be ruled out by any failure of ACD, per se, since the deletion site has undergone overt movement. The problem seems to be the fact that *which boy* fails to c-command a variable that it must bind within the deletion site, and feature movement seems to fail to allow this binding.

¹⁰⁰ Beck and Kim (1996) provide similar examples from Korean:

(i) *Amuto nuku-lul po-chi anh-ass-ni?
anyone who-ACC see-ing Neg-do+Past-Q

(ii) *Scrambling eliminates the violation*
Nuku-lul Amuto po-chi anh-ass-ni?

¹⁰¹Miyagawa notes that the choice of quantifier makes a difference here. In particular, *minna* 'every' does not produce the Intervention Effect — one of many differences between *minna* and other universal quantifiers discussed by Miyagawa. I will not attempt to account for this fact in this study.

¹⁰² The order of fronted *wh*-phrases matters. If the order is switched --

Nani-o dare-ni [John-ka Bill]-ga __ __ ageta no?
[single-pair only]

-- only a single-pair reading is possible. This is expected if one of the scrambled *wh*-phrases must reconstruct, but it is unclear why that would be the case. Perhaps if *nani-o* had to be parsed as an instance of "A-bar scrambling" like instances of long-distance scrambling (Webelhuth 1989, Mahajan 1990, Saito 1992), reconstruction would be forced, as suggested by Paul Hagstrom (personal communication). I leave the problem unresolved here.

¹⁰³ Günter Grewendorf (personal communication) has pointed out that *wh*-in-situ appears to generate an Intervention Effect in German separation constructions:

- (i) Was hat der Professor dem Studenten [__ alles] geraten
what has the professor the students all advised
'What-all has the professor advised the students?'
- (ii) *Was hat wer dem Studenten [__ alles] geraten
what has who the students all advised

As he notes, this fact is a problem for our hypotheses concerning Japanese. If all Japanese *wh*-phrases associate with C via feature-movement, then the higher of two *wh*-phrases in a multiple question might be expected to block the existence of the lower, just as *wer* in (i)b blocks the separation of *was* from *alles*. I leave this issue open; perhaps the set of interveners is not quite the same in Japanese as it is in German — as argued by Beck and Kim (1996) with respect to Korean and German.

Another open question noted by Grewendorf concerns subjacency. I observed in section 4.7 that our hypothesized *wh*-feature movement in English does not show island effects of the sort usually attributed to the Subjacency Condition and does not pay a "Subjacency Tax" for later instances of phrasal *wh*-movement. If Japanese *wh*-in-situ always associate with interrogative C by *wh*-feature movement, one might expect them also to disobey subjacency and not to pay a "Subjacency Tax". This is true up to a point. *Wh*-phrases quite generally fail to show Complex Noun Phrase Constraint effects, but they do show *wh*-island effects, as documented by Watanabe (1992).. In addition, they do pay a "Subjacency Tax"(governed by the PMC) for *wh*-island violations, as Watanabe also documented. *Wh*-feature movement of *dare-ni* in (ii) to the matrix complementizer *no* permits subsequent feature movement from *nani-o* to the matrix complementizer to avoid the *wh*-island effect seen in (i):

(i) ***Wh*-island effects**

??John-wa [Mary -ga nani -o katta ka dooka] siritagatte-iru no?

John TOP Mary NOM what ACC bought whether know-want Q

‘What does John want to know whether Mary bought?’

(ii) **Paying the *wh*-island tax**

John-wa [Mary-ga nani -o katta ka dooka] dare -ni tazuneta no?

John TOP Mary NOM what ACC bought whether who DAT asked Q

‘Who did John ask whether Mary bought what?’

It is not clear how we should understand the contradiction between (ia-b) and our observations concerning feature movement in English. It is possible that this contradiction reveals some fundamental problem with the hypotheses advanced here. It may also be the case that the contradiction simply reveals how the PMC individuates syntactic relations. In (i)b we observe (if our ideas about Japanese are correct) an instance of feature movement rescuing another instance of feature movement from subjacency effects. In (95)-(97), by contrast, we were examining cases in which an instance of feature movement might have been expected to rescue an instance phrasal movement. This might make a difference, but I will leave the question open for now.

¹⁰⁴Beck (1996) and Beck and Kim (1996) suggest a different cross-linguistic predictor for the Intervention Effect. They note that both Korean and German (the languages that they discuss) allow scrambling to remove a *wh*-phrase from the scope of a potential intervener, while English does not. They suggest that if a language has a scrambling process that can eliminate an Intervention Effect, it must use it — and speculate that the absence of scrambling correlates with the absence of the Intervention Effect. For them, the Intervention Effect is an island condition on traces of movement that have not been formed as early in the derivation as possible. It thus restricts "LF traces" in a language like German or Korean where scrambling can yield an overt-syntax trace in the same environment, but remains silent about comparable "LF

traces" in a language like English, where no overt process can create the trace earlier in the derivation. (As Beck and Kim note, the logic here follows Pesetsky's 1989 "Earliness Principle"; cf. also Diesing 1992.) They do not note, however, the Intervention Effect with which we began our discussion: the English effect found in multiple questions that appear to violate AC. One can imagine variants of their approach that might accommodate the Intervention Effect in English, but none of these (I think) straightforwardly extends to the other properties of these constructions discussed earlier.

¹⁰⁵This issue was brought to my attention by Hisatsugu Kitahara. I am also grateful to him for discussing the issue with me at length.

¹⁰⁶Hasegawa (1994) shows that a similar paradigm hold for the interaction of scrambling and negation with the negative polarity suffix *-sika* 'only, as well as for *-mo* 'also'. One wonders if a language could have the properties identified by Hasegawa without the property identified by Takahashi, and vice versa.

¹⁰⁷Disagreement of this sort is not unknown among Bulgarian speakers either, as noted earlier. Shigeru Miyagawa (personal communication) suggests that the contrast emerges most clearly when one focuses on the pair-list interpretation for the two examples. In his judgment, (180b) and (181b) have only single-pair readings.

¹⁰⁸A similar speculation has been advanced by Shigeru Miyagawa in recent unpublished work.

¹⁰⁹I owe this suggestion to Norvin Richards (personal communication). Aoun and Li relate the Japanese/Chinese difference to morphological differences in their *wh*-words. A footnote notes an important difficulty: Korean patterns with Chinese morphologically (by allowing bare *wh*-words to serve non-interrogative functions), but patterns with Japanese in showing the Intervention Effect. Our speculation about Chinese entails that Chinese *wh*-phrases, unlike their Japanese counterparts, undergo covert phrasal movement. This contradicts the conclusions of Aoun and Li (1993a) and other researchers, such as Tsai (1994).

¹¹⁰ It will be interesting to investigate similar typologies in other domains. There are some intriguing possibilities. In work still in progress, Guerzoni (1999) investigates whether the licensing of negative polarity items (NPIs) in English and Italian is, like the licensing of *wh*-phrases, accomplished sometimes by (covert) phrasal movement to the licenser, and sometimes by feature movement. Suggestive evidence in favor of this hypothesis comes from the fact that long-distance licensing of an NPI over an intervener (such a closer instance of negation) obeys island conditions that other instances of licensing do not, as noted by Kurata (1991). For example, licensing of an NPI by the nearest negation (or other licenser) does not display any subject/non-subject asymmetry.

- (i) a. John doesn't think that Mary loves anyone.
- b. John doesn't think that anyone loves Mary.

When an NPI has a choice between a nearer licenser and a further licenser, the facts are different. In (ii) below, the embedded object *anyone* may be licensed by either *impossible* in the middle clause, or else by negation in the higher clause. The two readings are paraphrased by Kurata as given:

- (ii) I don't believe it to be impossible that John saw anyone.

narrow scope: 'I believe it is possible that John saw at least one person.'

wide scope: 'I believe of each person that it is possible that John saw that person.'

By contrast, in (iii), the embedded subject *anyone* has only a reading in which it is licensed by the nearest instance of negation (the negative predicate *impossible*):

- (iii) I don't believe it to be impossible that anyone saw John.

narrow scope: 'I believe it possible that at least one person saw John.'

wide scope: *'I believe of each person that it is possible that that person saw John.'

Guerzoni suggests that the only way to allow an NPI to have wider scope than an intervening negation is through covert phrasal movement (a suggestion supported by evidence from ACD) to a position close to the licenser of the NPI (perhaps the specifier of NegP). It is this type of movement, she suggests, that displays subject/object asymmetries (as argued extensively in work of the 1980s, e.g. the papers in Kayne 1984). When scope is not assigned across an intervener, the licensing relation can be established by feature movement, with no subject/object asymmetry — as seen in (i).

REFERENCES

Aoun, Joseph, Norbert Hornstein and Dominique Sportiche. 1981. Aspects of Wide Scope Quantification. *Journal of Linguistic Research* 1:67-95.

Aoun, Joseph, Norbert Hornstein, David Lightfoot and Amy Weinberg. 1987. Two types of locality. *Linguistic Inquiry* 18: 537-578.

Aoun, Joseph and Yen-hui Audrey Li. 1993a. *Wh*-elements in situ: syntax or LF?. *Linguistic Inquiry* 24: 199-238.

Aoun, Joseph and Yen-hui Audrey Li. 1993b. On some differences between Chinese and Japanese *wh*-elements. *Linguistic Inquiry* 24: 365-372.

Baltin, Mark. 1987. Do Antecedent-Contained Deletions Exist? *Linguistic Inquiry* 18: 579-595.

Barss, Andrew. 1999. Minimalism and asymmetric *wh* interpretation. in *Step by Step: Essays in Honor of Howard Lasnik*. Cambridge, MA: MIT Press.

Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language Semantics* 4:1-56.

Beck, Sigrid and Shin-Sook Kim. 1996. On *wh*- and operator scope in Korean. *Journal of East Asian Linguistics* 6: 339-384.

den Besten, H. 1983. On the interaction of root transformations and lexical deletive rules. In Abraham, W., ed. *On the formal syntax of the Westgermania*. Amsterdam: Benjamins.

Billings, Loren and Catherine Rudin. 1994. Optimality and superiority: a new approach to overt multiple-*wh* ordering. In Jindra Toman, ed. *Proceedings of FASL 3: the College Park Meeting*. Ann Arbor, MI: Michigan Slavic Publications.

Bobaljik, Jonathan. 1995. *Morphosyntax: the syntax of verbal inflection*. Doctoral dissertation, MIT, Cambridge, MA.

Boeckx, Cedric. 1999a. Properties of French questions. Ms., University of Connecticut, Storrs, CT.

Boeckx, Cedric. 1999b. Decomposing French Questions. in UPenn Working Papers in Linguistics 6. Department of Linguistics, University of Pennsylvania, Philadelphia, PA.

Bošković, Željko. 1995. On certain violations of the superiority condition, AgrO, and economy of derivation. *Journal of Linguistics* 33: 227-254.

Bošković, Željko. 1997. Sometimes in SpecCP, sometimes in-situ. Ms. University of Connecticut, Storrs, CT.

Bošković, Željko. 1998. Multiple *wh*-fronting and economy of derivation. In *Proceedings of the West Coast Conference on Formal Linguistics* 16. Stanford University, Stanford, CA.

Bošković, Željko. To appear. Multiple *wh*-fronting and economy of derivation. *Proceedings of WCCFL* 16.

Bošković, Željko. In press. On multiple feature-checking: Multiple *wh*-fronting and multiple head-movement. In *Working Minimalism*, ed. Samuel Epstein and Norbert Hornstein. MIT Cambridge, MA:MIT Press.

Bouton, L. 1970. Antecedent-Contained Pro-Forms. *Papers from the Sixth Regional Meeting, Chicago Linguistic Society*. Chicago Linguistic Society, University of Chicago, Chicago, IL.

Brody, M. 1995 *Lexical-Logical Form*. Cambridge, MA: MIT Press.

Brody, M. 1997. Perfect chains. In Liliane Haegeman, ed. *Handbook of syntax*. Dordrecht, Netherlands: Kluwer.

Büring, Daniel. 1994. The dark side of wh-movement. *Linguistische Berichte* 149: 56-74.

Cheng, Lisa. 1991. *On the typology of wh-questions*. Doctoral dissertation, MIT. [reprinted by Garland Publishing, New York, NY.]

Cheng, Lisa. 1997. "Partial" wh-movement. In *UCI Working papers in Linguistics 3*, ed. Luther Chen-Sheng Liu and Kazue Takeda, 27-50. University of California at Irvine, Irvine, CA.

Cheng, Lisa. 1999. Moving just the feature. ms., University of California at Irvine. To appear in ed. U. Lutz, G. Müller and A. von Stechow. *Wh-scope marking*. Amsterdam: Benjamins.

Chierchia, Gennaro. 1993. Questions with quantifiers. *Natural Language Semantics* 1: 181-234.

Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. Stephen Anderson and Paul Kiparsky 232-286. New York: Holt, Reinhart and Winston.

Chomsky, Noam. 1976. Conditions on rules of grammar. *Linguistic Analysis* 2: 303-351; [reprinted in *Essays on Form and Interpretation*. New York: Elsevier North-Holland.

- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht, Netherlands: Foris.
[reprinted: Dordrecht, Netherlands: Mouton de Gruyter]
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press
- Chomsky, Noam. 1998. Minimalist inquiries. unpublished ms., MIT.
- Chomsky, Noam and Howard Lasnik. 1977. Filters and Control. *Linguistic Inquiry* 8: 425-504.
- Chung, Sandra and James McCloskey. 1983. On the interpretation of certain island facts in GPSG. *Linguistic Inquiry* 14: 704-713.
- Cole, Peter and Gabriela Hermon. 1998. The typology of wh movement. *Syntax* 1.
- Comorovski, Ileana. 1996. *Interrogative Phrases and the Syntax-Semantics Interface*. Kluwer: Dordrecht.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press
- Engdahl, Elisabet. 1980. *The syntax and semantics of questions in Swedish*. Doctoral dissertation, University of Massachusetts at Amherst, Amherst, MA.
- Engel, Ulrich. 1972. *Regeln zur Satzgliedfolge: zur Stellung der Elemente im einfachen Verbalsatz*. Düsseldorf: Schwann.
- Fanselow, Gisbert. 1990. Scrambling as NP-movement. In Günther Grewendorf and Wolfgang Sternefeld. 1990. *Scrambling and Barriers*. Amsterdam: John Benjamins.

Fanselow, Gisbert. 1991. *Minimale Syntax*. Habilitation thesis. University of Passau.

Fanselow, Gisbert. 1997 Minimal Link effects in German (and other languages), unpublished ms.
University of Potsdam. [Related 1996 handout at <http://www.ling.uni-potsdam.de/~fanselow/mlc.htm>]

Fiengo, Robert. 1998. How to ask multiple questions — some simple ways: a footnote to Austin. Ms.
Queens College and the Graduate Center of CUNY, New York.

Fiengo, Robert and Robert May. 1994. *Indices and Identity*, Cambridge, MA: MIT Press.

Fox, Danny. 1995. Condition C Effects in ACD. In ed. Robert Pensalfini and Hiroyuki Ura. *Papers on Minimalist Syntax, MIT Working Papers in Linguistics 27*: 105-120. MIT, Cambridge, MA.

Fox, Danny. 1998. *Economy and Semantic Interpretation - A study of scope and variable binding*.
Doctoral dissertation, MIT, Cambridge, MA.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12: 155-184.

George, Leland. 1980. *Analogical Generalization in Natural Language Syntax*. Doctoral dissertation,
MIT.

Grewendorf, Günther and Wolfgang Sternefeld. 1990. Scrambling theories. In Günther Grewendorf and Wolfgang Sternefeld. 1990. *Scrambling and Barriers*. Amsterdam: Benjamins.

Grimshaw, Jane. 1997. Projection, Heads and Optimality. *Linguistic Inquiry* 28: 363-422.

Groat, Erich and John O'Neil. 1996. Spell-out at the LF interface. in: Werner Abraham, Samuel D. Epstein, Höskuldur Thraínsson and Jan-Wouter Zwart, eds. *Minimal Ideas*. Amsterdam: Benjamins.

Grohmann, Kleanthes. 1998. Syntactic Inquiries into Discourse Restrictions on Multiple Interrogatives. Ms., University of Maryland, College Park, MD.

Guerzoni, Elena. 1999. Phrasal and feature movement: syntactic conditions on NPI licensing. Ms., MIT, Cambridge, MA.

Hagstrom, Paul. 1998. *Decomposing Questions*. unpublished Doctoral dissertation, MIT. [available from MIT Working Papers in Linguistics, mitwpl@mit.edu]

Haider, Hubert. 1986. Affect : a reply to Lasnik and Saito, "On the nature of proper government". *Linguistic Inquiry* 17, 113-125.

Hankamer, Jorge. 1974. On WH Indexing. *Papers from the Fifth Annual Meeting of the North Eastern Linguistic Society*. Department of Linguistics, Harvard University, Cambridge, MA.

Hankamer, Jorge and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7: 391-428.

Hoji, Hajime. 1995. *Logical form constraints and configurational structures in Japanese*. Doctoral dissertation, University of Washington, Seattle, WA.

Honcoop, Martin. 1998. *Dynamic Excursions on Weak Islands* (Holland Institute of Generative Linguistics/Landelike Onderzoekschool Taalwetenschap dissertation). the Hague: Holland Academic Graphics.

Hornstein, Norbert. 1994. An argument for Minimalism: the case of antecedent-contained deletion. *Linguistic Inquiry* 25: 455-480.

Hornstein, Norbert. 1995. *Logical form: from GB to Minimalism*. Cambridge, MA: Blackwell.

Huang, C. T. James . 1981. Move *wh* in a language without *wh* movement. *The Linguistic Review* 1: 369-416.

Huang, C. T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*, Doctoral dissertation, MIT. [reprinted by Garland Publishers, New York]

Johnson, Kyle and Satoshi Tomioka. 1997. Lowering and Mid-Size Clauses. Ms. University of Massachusetts at Amherst, Amherst, MA, and UC San Diego, San Diego, CA.

Johnson, Kyle. 1992. Object Positions. *Natural Language and Linguistic Theory* 9: 577-636.

Kayne, Richard. 1983. Connectedness. *Linguistic Inquiry* 14: 223-249; reprinted in Kayne (1984).

Kayne, Richard. 1984. *Connectedness and Binary Branching*, Foris, Dordrecht.

Kayne, Richard 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press

Kayne, Richard. 1998. Overt vs. covert movement. Ms. New York University, New York.

Kennedy, Christopher. 1997. Antecedent-Contained Deletion and the Syntax of Quantification. *Linguistic Inquiry* 28: 662-688.

Kiss, Katalin É. 1986. Against the LF-Movement of WH-Phrases. Ms., Hungarian Academy of Sciences, Budapest.

Koizumi, Masatoshi. 1995. *Phrase structure in minimalist syntax*, unpublished Doctoral dissertation, MIT. [available from MIT Working Papers in Linguistics, mitwpl@mit.edu]

Koot, Hans van de. 1988. The vacuous movement hypothesis, superiority and the ECP. In ed. Ger de Haan and Wim Zonneveld. *Formal parameters of generative grammar IV: Yearbook 1988*. Dordrecht: ICG Printing.

Koster, Jan. 1987. *Domains and dynasties*. Dordrecht: Foris Publications.

Kuno, Susumu. 1982. The focus of the question and the focus of the answer. In R. Schneider, K. Tuite and R. Chametzky, eds. *Papers from the Parasession on Nondeclaratives*. Chicago Linguistic Society. University of Chicago Dept. of Linguistics, Chicago.

Kuno, Susumu and Jane J. Robinson. 1972. Multiple WH-Questions. *Linguistic Inquiry* 3: 463-487.

Kurata, Kiyoshi. Doctoral dissertation, University of Massachusetts, Amherst, MA.

Larson, Richard and Robert May. 1990. Antecedent containment or vacuous movement: Reply to Baltin. *Linguistic Inquiry* 21:103-122.

Lasnik, Howard and Mamoru Saito. 1984. On The Nature of Proper Government. *Linguistic Inquiry* 15: 235-290.

Lasnik, Howard. 1993. *Lectures on Minimalist Syntax*, MIT Working Papers in Linguistics, MIT, Cambridge, MA.

Legendre, Geraldine, Paul Smolensky, and Colin Wilson. 1998. When is Less More? Faithfulness and Minimal Links in wh-Chains". in Barbosa, P., D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky. *Is the Best Good Enough?* MIT Press: Cambridge, MA

Longobardi, Giuseppe. 1986. L'estrazione dalle isole e lo scope dei sintagmi quantificati. in K. Lichem, E.Mara, S.Knaller , eds. *Parallela 2, Atti del terzo Incontro Italo-Austriaco di Linguisti*. Tübingen.

Longobardi, Giuseppe. 1991. In defense of the correspondence hypothesis: island effects and parasitic constructions in logical form. in C.-T. James Huang and Robert May, eds. *Logical structure and linguistic structure*. Kluwer: Dordrecht.

Mahajan, Anoop. 1990. *The A/A' distinction and movement theory*. Doctoral dissertation, MIT, Cambridge, MA.

May, Robert. 1977. *The Grammar of Quantification*. Doctoral dissertation, MIT, Cambridge, MA.

May, Robert. 1985. *Logical Form*. MIT Press: Cambridge, MA.

Miyagawa, Shigeru. 1997. class handout, MIT, Cambridge, MA.

Miyagawa, Shigeru. 1988. *WH chains and quantifier induced barriers*. unpublished ms., MIT.

Müller, Gereon. 1995. *A-bar syntax: a study in movement types*. Berlin: Mouton de Gruyter.

Müller, Gereon. 1996. A constraint on remnant movement. *Natural Language and Linguistic Theory*. 14: 355-407.

Müller, Gereon. 1998. Order preservation, parallel movement, and the emergence of the unmarked. unpublished ms. Universität Stuttgart. [<ftp://rucss.rutgers.edu/pub/OT/TEXTS/archive/275-0798/275-07982.pdf>]

Müller, Gereon and Wolfgang Sternefeld. 1993. Improper movement and unambiguous binding. *Linguistic Inquiry* 24: 461-507.

Nunes, Jairo. 1995. The Copy Theory of Movement and Linearization of Chains in the Minimalist Program. Doctoral dissertation. University of Connecticut, Storrs, CT.

Obenauer, Hans-Georg. 1984. On the identification of empty categories. *Linguistic Review* 4: 153-202.

Ochi, Masao. 1988. Move or Attract? In *Proceedings of the West Coast Conference on Formal Linguistics* 16. Stanford University, Stanford, CA.

Perlmutter, David. 1972. Evidence for shadow pronouns in French relativization. In *The Chicago Which Hunt: Papers from the Relative Clause Festival*, ed. P. Peranteau, J. Levi, and G. Phares. Chicago Linguistic Society. Chicago, IL.

Pesetsky, David. 1989. Language-Particular Rules and the Earliness Principle. Ms. MIT, Cambridge, MA.

Pesetsky, David. 1982. *Paths and Categories*. Doctoral dissertation, MIT, Cambridge, MA.

Pesetsky, David. 1987. Wh-in-situ: Movement and Unselective Binding. E. Reuland and A. ter Meulen, eds., *The Representation of (In)definiteness*, 98-129. Cambridge, MA: MIT Press

Pesetsky, David. 1997. Optimality Theory and Syntax: Movement and Pronunciation. Diana Archangeli and D. Terence Langendoen (eds) *Optimality Theory: An Overview*. Oxford: Blackwell.

- Pesetsky, David. 1998. Some Optimality Principles of Sentence Pronunciation. in Barbosa et al. (1998). Pilar Barbosa, Danny Fox, Martha McGinnis, Paul Hagstrom and David Pesetsky (eds.). 1998. *Is the Best Good Enough?*, MIT Working Papers in Linguistics. and MIT Press.
- Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*, RuCCs Technical Report #2, Rutgers University Center for Cognitive Science, Piscataway, New Jersey [to appear, Cambridge, Mass.: MIT Press].
- Reinhart, Tanya. 1997. Quantifier scope: how labor is divided between QR and choice functions. *Linguistics and Philosophy* 20: 335-397.
- Richards, Norvin. 1997 *What Moves Where in Which Language?* Doctoral dissertation, MIT, Cambridge, MA.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press
- Ross, John R. 1967. *Constraints on Variables in Syntax*. Doctoral dissertation, MIT, Cambridge, MA.
- Rudin, Catherine. 1985. *Aspects of Bulgarian syntax: complementizers and wh constructions*. Columbus, OH: Slavica.
- Rudin, Catherine. 1988. On multiple questions and multiple WH fronting. *Natural Language and Linguistic Theory*. 6: 445-502.

- Sag, Ivan. 1976. *Deletion and Logical Form*. Doctoral dissertation, MIT, Cambridge, MA.
- Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1: 69-118.
- Saito, Mamoru. 1994. Additional *wh*-effects and the adjunction site theory. *Journal of East-Asian Linguistics* 3: 195-240.
- Sauerland, Uli. 1996. The interpretability of scrambling. In *Formal Approaches to Japanese Linguistics 2, MITWPL 29*, ed. by Masa Koizumi, Masayuki Oichi, and Uli Sauerland. Cambridge, MA: MIT Working Papers in Linguistics.
- Sauerland, Ulrich. 1998a. *The Meaning of Chains*. Doctoral dissertation, MIT, Cambridge, MA.
- Sauerland, Ulrich. 1998b. Erasability and interpretation. ms. Kanda University, Makuhari Kaihin, Japan.
- Stjepanović, Sandra. 1998. On the placement of Serbo-Croatian clitics: Evidence from VP-ellipsis. *Linguistic Inquiry* 29: 513-515.
- Takahashi, Daiko. 1994. *Minimality of Movement*. Doctoral dissertation, University of Connecticut, Storrs, CT.
- Tanaka, Hidekazu. 1999. *Conditions on logical form derivations and representations*. Doctoral dissertation, McGill University.
- Tsai, Wei-Tien Dylan. 1994. On nominal islands and LF extraction in Chinese. *Natural Language and Linguistic Theory* 12: 121-175.

Ura, Hiroyuki. 1996. *Multiple feature checking: a theory of grammatical function splitting*. Doctoral dissertation, MIT, Cambridge, MA.

Wachowicz, Krzystina 1974. Against the Universality of a Single WH-Question Movement. *Foundations of Language* 11: 155-166.

Watanabe, Akira. 1992. Subjacency and S-structure movement of *wh*-in-situ. *Journal of East Asian Linguistics* 1: 255-291.

Wilder, Chris. 1999. Right node raising and the LCA. *Proceedings of the West Coast Conference on Formal Linguistics*. Department of Linguistics, Stanford University. Stanford, CA.

Webelhuth, Gert. 1988. Doctoral dissertation, University of Massachusetts at Amherst, Amherst, MA.

Wiltschko, Martina. 1997. D-linking, scrambling and superiority in German. In Werner Abraham, ed. *Groninger Arbeiten zur germanistischen Linguistik*, 41. Germanistisch Instituut, Rijksuniversiteit Groningen, Netherlands.

Wurmbrand, Susanne. 1998. *Infinitives*. Doctoral dissertation, MIT, Cambridge, MA.