effect in these three language types follows straightforwardly from my hypothesis about the movement-like relations available to languages, in tandem with a typology of specifier requirements that also figures prominently in chapters 2–4. In short, I will argue that my general view of movement helps explain how the syntax of *wh*-questions differs across languages.

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

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 position in syntactic structure. Movement is most easily detected 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 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 here will be overt *wh*-phrase movement of the sort familiar from many languages, including English.

Overt wh-*phrase movement*

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

As linguists, we are lucky that overt phrasal movement exists; if it did not, 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 shows that other pronunciation patterns do exist. For example, Chomsky (1976), May (1977, 1985), 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*. Here I will call it *covert phrasal movement.* Covert phrasal movement is *movement* insofar as it creates a link between positions that 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).

ACD

(2) Mary [$_{VP}$ invited [$_{DP}$ everyone that I did [$_{VP}$ Δ]]].

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).

Interpretation of ellipsis in (2)

(3) Mary [$_{VP}$ invited [$_{DP}$ everyone that I [$_{VP}$ invited t]]].

It is a fact about VP-ellipsis (Hankamer and Sag 1976) that unlike a pronoun, which may take its reference from a contextually salient but unmentioned individual, an elided VP requires a linguistic expression—a pronounced VP—as its antecedent. Thus, although I can say "Thank goodness *he*'s left" as a response to the departure of a salient but unmentioned individual, I cannot say "Thank goodness *I* didn't [$_{VP}$ Δ]" 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*. But if the phrase *everyone that I did* undergoes

covert phrasal movement to a VP-external position, then it leaves behind a VP of the form *invited t*, supplying the appropriate antecedent for the elided VP.

Example (2) after covert phrasal movement

(4) [$_{DP}$ 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 note 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]

Scope/ellipsis correlations

(5) a. John refused to visit [every city Mary did [$_{VP}$ Δ]].
  b. i. John refused to visit [every city Mary visited t]. [ambiguous: compatible with narrow or wide scope of *every city*]
    ii. 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 1993). In the process of semantic interpretation, descriptive material in the 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).

No bleeding of Principle C without ACD

(6) *I [sent him$_i$ [every letter that John$_i$ 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.

Bleeding of Principle C with ACD

(7) I [sent him$_i$ [every letter that John$_i$ expected I would [$_{VP}$ Δ]] → [every letter that John$_i$ expected I would [$_{VP}$ send him$_i$ t]] I [sent him$_i$ t]

Important complications are discussed by Fox (1995, 1998) and Sauerland (1998b) 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 Fox argues. The only point relevant here 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 viewed as a language- or construction-specific variant of overt phrasal movement. For example, many researchers have proposed covert phrasal movement of *wh*-phrases to C as a covert version of the overt phrasal *wh*-movement displayed in (1) (e.g., Huang 1981, 1982; 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*-phrase moves.

Command condition on overt and covert wh-*movement*

(8) 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 (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]

Command condition is not semantic

(9) [Give a book to his$_i$ mother] I can guess who will ____!

If these conclusions are correct, the LF representation of *Who gave what to whom?* may look something like (10b), in which *wh*-movement has

affected all three *wh*-elements. Overstriking indicates the pronunciation pattern.

English multiple questions: LF and pronunciation

(10) a. Who gave what to whom?

b. [*who* ~~*what*~~ ~~*whom*~~ [~~*who*~~ gave *what* to *whom*]][6]

As often noted in the 1980s (Lasnik and Saito 1984; Rudin 1985; Pesetsky 1987), this hypothesis about multiple questions in English 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 (10b) represents a correct analysis for English multiple questions, it differs from its Bulgarian counterpart (11b) only in how its *wh*-chains are pronounced.

Bulgarian multiple questions: LF and pronunciation

(11) a. Koj kakvo na kogo dade?
who what to whom gave
'Who gave what to whom?'

b. [*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 may contain an instance of 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 that resolves ACD is not *wh*-movement (as argued by Hornstein (1994, 1995) and Lasnik (1993), among others). I return to this issue in section 3.2, where I argue that the movement that resolves ACD is indeed *wh*-movement. ((12c) is from Fiengo and May 1994, 242.)

ACD licensed in wh-*in-situ*

(12) a. Which girl invited [which student that John did [$_{VP}$ Δ]]?

b. I need to know who can speak [which languages that Ken Hale can [$_{VP}$ Δ]]?

c. Which spy-master suspected which spy that Angleton did [$_{VP}$ Δ]]?

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 Spell-Out.

Phonological Spell-Out in the Y-model
(13) Pronounce only the highest position in a movement chain.

On this view, covert movement is simply movement that takes place after Spell-Out. Movement after Spell-Out thus creates a situation in which the highest position at LF is higher than the highest position at the time of Spell-Out. 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 (1993, 1995), 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 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 Spell-Out 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 versus 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 I will describe pronunciation distinctions like the one between overt and covert phrasal movement as essentially phonological. Against this backdrop I will argue that (14) and (15) properly characterize the difference between English and Bulgarian pronunciation patterns in multiple questions. Reference to the "first instance" and "secondary instances" of *wh*-movement will be justified shortly.

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 unpronounced in its new position.

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.

1.2 Feature Movement

In his first paper developing the Minimalist Program, Chomsky (1993) suggested that covert phrasal movement is the default style of movement. He implemented this idea 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. In this system a strong feature was a feature that needed to be "checked" 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 developed in Pesetsky 1989,

as incorporated, for example, in Brody's (1995) "radically minimalist" model), or no timing principle at all?

In later work Chomsky (1995) 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 nonconvergent. Movement for any other purpose is banned. 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 (gratuitous instances of this movement strategy being 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 arises when 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, 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 a feature moves, "[it] carries along just enough material for convergence" (1995, 262). Movement of more than just the relevant feature occurs only if "generalized pied-piping" is necessary 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 copies a phrase is motivated by PF considerations and must by its very nature be "overt" movement. Movement that copies only the grammatical features that motivate the movement must 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, and postulation of covert *phrasal* movement was simply an error. Notice that this view, if correct, offers a third slant on the overt/covert contrast—distinct from both the Y-model and single-output syntax traditions. 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 take place in a single syntactic cycle, phonological distinctions arising simply from differences in the material copied by the movement operation. Chomsky (1995) did not develop this consequence of his new view of movement, but it follows straightforwardly, nonetheless. In Chomsky 1998 he takes exactly this step.

Chomsky's (1995) proposal concerning movement is consonant with some arguments for covert movement, but not all. 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). However, it runs afoul 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 Fiengo and May's (1994) binding theory evidence, as explained by Fox (1995)[10]).

Consequently, it does not appear correct to simply reanalyze covert phrasal movement as feature movement. Covert phrasal movement exists. I suspect that this conclusion is not a step backward. 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 needed 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 (1995) idea, only the feature 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 expressions bearing a particular *feature*? I will suggest that the answer is yes. If so, then Chomsky's (1995) discussion of feature movement uncovered something real—but misidentified the discovery. The phenomenon that Chomsky called "feature movement" is not an alternative analysis of covert phrasal movement, as he thought, but 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 existential *there* construction. As is well known, the verb in the English existential *there*-construction generally agrees with a postverbal "associate" DP, which is usually required to be indefinite.

Agreement with the associate DP in the there *construction*

(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, Merge 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" (i.e., delete) its number features (and possibly others). The word *there*, Chomsky suggests (1995, 273), does not bear these features. Consequently, the features must move to T from somewhere inside TP. The associate DP furnishes the necessary features.

Feature movement from the associate DP in the there *construction*

(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.

C-command condition on feature movement

(20) 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.

Ban on superraising with overt phrasal movement

(21) *Several $books_i$ are desirable [for it to be t_i on the table].

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

Ban on superraising with feature movement

(22) *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 a means of detecting phrasal movement and strongly suggests that the answer 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.

Relative clauses compatible with an associate

(23) 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.

Next we should ensure that VP-ellipsis is compatible in principle with the *there* construction. Example (24) shows that it is.

VP-ellipsis acceptable in the there-*associate construction*

(24) Will there be phonologists at the party? Well, there should [$_{VP}$ Δ].

If we now elide the VP in (23), we have our test for ACD. The examples in (25) display the crucial contrast with (23).

ACD impossible in a relative clause modifying an associate[13]

(25) a. *There will be [everyone that there should [$_{VP}$ Δ]] at the party.
b. *There will be [almost no one that there should [$_{VP}$ Δ]] at the party.

Since (24) shows that there is nothing wrong with the ellipsis per se, the unacceptability of ACD in (25) shows 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 (1995, 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 (1986b), in which the movement properties of the associate-T relation were attributed to covert phrasal movement. Though Chomsky did not base an argument for this revision on ACD, it does provide 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 find evidence of a relation between a phrase pronounced in situ and a higher position. But the two constructions behave differently under a test for specifically *phrasal* covert movement. Whereas 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 Chomsky's (1995) sense. Later, however, I will also examine some possible variants of 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.

I am now in a position to sketch the goals of the book in some detail. If the discussion so far is on the right track, we might expect to find "minimal triplets" in which a particular head establishes 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 my 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.

In the remainder of this book I present and investigate 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 the discussion of English multiple questions with malice aforethought, since Bulgarian provides some of the foundations of the argument I will present. After giving evidence in favor of covert phrasal movement for *wh*-in-situ, I noted (in the footsteps of Lasnik and Saito (1984)) that on this view Bulgarian simply shows overt phrasal movement where English shows its covert counterpart (see the summary of this difference in (14)–(15)). In the next chapter I present certain other facts about Bulgarian multiple questions and offer partial explanations (borrowed from Richards 1997). In chapters 2–4 I show that a seemingly distinct set of puzzles concerning English multiple questions actually represents the same phenomena observed in Bulgarian. If this is true, I can claim that the pronunciation distinctions in (14) and (15) constitute the *only* relevant difference between multiple questions in the two languages. However, I only reach this conclusion if I assume that in certain cases multiple questions in English (and Bulgarian) involve pure feature movement of *wh*—even though in other cases they involve overt

and covert phrasal movement triggered by the same feature. In chapter 5 I consider syntactic and semantic consequences of this hypothesis, a discussion that helps establish a crosslinguistic typology of *wh*-movement.

1.3 Superiority Effects

The English puzzles involve the observation known as the *Superiority effect*. The Superiority effect arises in a multiple question when more than one *wh*-phrase is relevant to the answering patterns for the question. In such a case 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 responsible for the contrast observed in (26)–(27).

Superiority effect

(26) 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 is accurate.

Kuno and Robinson's constraint

(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*.

Superiority Condition

(29) No rule can involve X, Y in the structure
$\ldots X \ldots [_{\alpha} \ldots Z \ldots\text{–}WYV \ldots]$
where the rule applies ambiguously to Z and Y and Z is superior to [m-commands] Y.

Chomsky's later condition (1995, 280, 296), which I will call *Attract Closest* (AC), 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.

Attract Closest

(30) α can raise to target K only if there is no legitimate operation Move β targeting K, where β 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 the investigation.

1.3.1 D-Linking

One minimal change of this sort involves what I have elsewhere (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 to both 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.

Superiority effect disappears with D-linking

(31) 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 that 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 book. On the other hand, the syntax of the problem is also obscure: What is the structure of D-linked questions in which the Superiority effect appears to be obviated?[16] Are they exceptions to AC, or is something else going on? By the end of the book I hope to have answered *these* questions quite fully.

1.3.2 More than Two *Wh*-Phrases

The next factor is less studied, but equally striking.[17] The Superiority effect is limited to questions with two *wh*-phrases. It disappears in multiple questions with more than two *wh*-phrases.[18]

Superiority effect disappears with more than two wh-*phrases*

(33) a. *What did who give ____ to Mary? [detectable Superiority effect]
b. What did who give ____ to whom? [no detectable Superiority effect]

I will discuss this effect in greater detail later. For now I simply wish to note the important syntactic questions this contrast raises. They are the same as the ones posed previously: 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 AC, or is something else going on?

1.3.3 German

Another minimal change 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.

Superiority effect disappears in German

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

The correct interpretation of this fact is a matter of controversy, which I will turn to in sections 5.3 and 5.4. For now it is sufficient to observe the problem and to ask the same questions posed about the previous two sets of apparent counterexamples.

These are the problems that will occupy the remainder of this book. In the next chapter I present a fuller picture of Bulgarian multiple questions, as necessary background to the discussion of apparent exceptions to the Superiority effect.