Multiple Wh-Fronting

Günther Grewendorf

This article argues that overt multiple wh-fronting in languages like Bulgarian consists of moving a single wh-cluster to [Spec, CP]. The formation of wh-clusters is motivated by the assumption that wh-elements can act as landing sites for wh-movement due to morphological properties of wh-words. I further argue that languages such as Japanese constitute covert instances of this process of wh-cluster formation, accounting for intricate constraints on multiple wh-questions such as the so-called “additional-wh effect.” Another central claim of the article is that despite appearances, multiple wh-questions in German equally involve the formation of wh-clusters, which are shown to consist of one visible and one or more invisible wh-elements. This analysis provides a new account for the lack of “short” and the presence of “long” superiority effects in German.

Keywords: multiple wh-questions (Bulgarian, Japanese, German), wh-cluster, single-level theory of movement, additional-wh effect, superiority

1 Introduction

Natural languages display distinct strategies in their wh-question formation. Among languages with overt wh-fronting we find languages like English and German that show obligatory overt fronting of only one wh-phrase and leave other wh-phrases in situ, and languages like Bulgarian and Romanian that obligatorily move all wh-phrases to [Spec, CP] overtly.

In this article I suggest an analysis of overt multiple wh-fronting languages like Bulgarian and Romanian according to which they actually pattern with languages like English and German in that they move only one wh-element to [Spec, CP], the relevant differences between the two types of languages resulting from a parametric property of the wh-elements themselves that requires the formation of clusters of wh-phrases in one type but not in the other. Using evidence from Japanese, I further demonstrate that there is an LF-type analogue of multiple wh-fronting languages that in a covert way displays exactly the sort of wh-cluster formation that my analysis assumes to occur as an overt process in languages like Romanian and Bulgarian.

Variants of this article have been presented at Yale University, the XXIII Incontro di Grammatica Generativa (Scuola Normale Superiore, Pisa), the University of Potsdam, and the University of Vienna. I would like to thank the audiences at those presentations, and especially Steve Anderson, Sergey Avrutin, Željko Bošković, Chris Collins, Liliane Haegeman, Hubert Haider, Dianne Jonas, Anoop Mahajan, Gereon Müller, David Pesetsky, Ur Shlonsky, and Yuji Takano for their comments. I owe a particular debt to Eric Fuß, Joachim Sabel, and Andrew Simpson for helpful discussions and extensive comments. Thanks also go to Marina Stojanova for help with the Bulgarian data and to two anonymous LI reviewers whose astute comments led to considerable improvement of this article. This research was supported by DFG grant GR 559/5-1.
A crucial claim concerns the analysis of wh-movement in German. Even though in German multiple wh-questions only one wh-phrase is permitted to move overtly to [Spec, CP], I will show that in German multiple wh-questions as well it is in fact a wh-cluster that moves to [Spec, CP]. However, unlike in languages such as Bulgarian and Romanian, where the wh-cluster consists of visible wh-elements, and languages like Japanese, where it consists of invisible wh-elements, in German the wh-clusters that move to [Spec, CP] are assumed to consist of one visible and several other invisible wh-elements. To establish this claim, I adopt a single-level model of movement along the lines of Groat and O’Neil’s (1996) ‘‘Spell-Out at LF’’ theory and Chomsky’s (2000, to appear) minimalist theory, according to which the distinction between movement before and after Spell-Out can be dispensed with. The alternative nature of this model can thus be seen in the claim that movement (overt as well as covert) always takes place before Spell-Out.

In section 2 I briefly discuss three accounts of overt multiple wh-fronting languages that are based on an Empty Category Principle (ECP) approach, an economy approach, and an optimality approach, respectively. I show that these accounts face conceptual and empirical problems so that the search for an alternative, more adequate theory seems to be justified. In section 3 I develop an alternative theory of multiple wh-fronting according to which multiple wh-fronting does not consist of separately moving individual wh-elements to [Spec, CP] but rather consists of forming a single wh-cluster and then moving it to [Spec, CP]. In section 4 I consider multiple wh-questions in German. The crucial claim established here is that despite appearances, multiple wh-questions in German involve the formation of wh-clusters. In section 5 I suggest an account of multiple wh-questions with overtly extracted wh-adjuncts.

2 Constraints on Multiple Wh-Fronting: The ECP, Economy, and Optimality

In this section I briefly discuss three accounts of multiple wh-fronting that are based on an ECP approach, an economy approach, and an optimality approach, respectively. In discussing these accounts, I point out several conceptual and empirical problems for which the new account developed in section 3 suggests alternative solutions.

According to Rudin’s (1988) ECP-based analysis of multiple wh-fronting in languages such as Bulgarian and Romanian, a parallel can be drawn between the possible (unmarked) orders of fronted wh-phrases in these languages and the effects of superiority in languages like English. In both Bulgarian and Romanian the subject must precede the object if both a subject and an object wh-element are fronted. Rudin takes this to be reminiscent of the superiority effect in English, where when subject and object in a multiple question are both wh-phrases, the subject rather than the object must be moved.

(1) a. Koj kogo e vidjal?
   who whom is seen
   ‘Who saw whom?’

1 For a different view on the putative superiority effect with multiple wh-fronting in Bulgarian, see Bošković 1998, where only movement of the first wh-element is taken to be wh-movement (and subject to superiority) but movement of the remaining wh-elements is taken to be focus movement (and not subject to superiority).
b. *Kogo koj e vidjal?  
(Bošković 1998)

Before continuing with Rudin’s account, I should point out that the order of fronted *wh*-elements in languages like Romanian and Bulgarian is not exclusively determined by purely syntactic factors. As mentioned in Rudin 1988 and especially emphasized in Comorovski 1989 and Richards 1997, fronted *wh*-elements exhibit a relative freedom of order if they are D-linked or constitute echo questions. We should therefore keep in mind that the following discussion of these word order facts is actually restricted to cases where neither a D-linked nor an echo interpretation of fronted *wh*-elements is present.²

Proceeding from the assumption that Bulgarian and Romanian have the same structure for multiple questions at S-Structure that other languages have at LF, Rudin accounts for the restrictions on *wh*-word order in multiple *wh*-fronting languages in terms of the ECP, which she adopts in a split version along the lines proposed by Aoun et al. (1987): *wh*-traces must be head-governed at PF and, in addition, fulfill a local binding condition at LF in the sense of Aoun’s (1986) theory of Generalized Binding.³ A crucial component of this account is based on the assumption, independently argued for, that in languages such as Bulgarian and Romanian (as opposed to *wh*-fronting languages with free *wh*-word order like Polish and Czech), a single *wh*-phrase is substituted into [Spec, CP] and the other fronted *wh*-phrases are right-adjoined to the *wh*-element in [Spec, CP].

Rudin’s ECP account of the fixed order of fronted *wh*-phrases in Bulgarian and Romanian faces empirical problems. One of them has to do with the order of fronted *wh*-arguments and *wh*-adjuncts. If it is true that *wh*-subjects have to precede *wh*-adjuncts in the unmarked case, as pointed out in Rudin 1985 and Cheng 1991, ⁴ this order cannot be derived from Rudin’s account.

2 Richards (1997:111) shows that the relative freedom of ordering in the case of D-linked *wh*-words can be attributed to the fact that only D-linked *wh*-words may move into the specifier position of a Topic projection that dominates CP. An alternative account of D-linked exceptions to the superiority effect can be found in Pesetsky 2000:39ff.

3 Note that in Aoun’s approach the theory of Generalized Binding is considered to be an alternative to an ECP account of locality constraints.

4 Rudin (1988:fn. 21) describes the distribution of *wh*-adjuncts as “subtle and complex” and suggests that nonsubcategorized adjuncts like ‘how’ and ‘why’ are not permitted in Bulgarian and Romanian multiple questions at all. However, my Bulgarian informants have confirmed the generalizations stated in Rudin 1985 and Cheng 1991 according to which *wh*-adjuncts can be fronted but cannot precede a *wh*-argument in multiple questions. They pointed out examples such as (i)–(iv).

(i) Kakvo kak napravi Ivan?  
what how did Ivan
‘How did Ivan what?’

(ii) Kogo kak napadna?  
who how you-attacked
‘How did you attack who?’

(iii) Koj kak se čuvstva?  
who how himself feels

(iv) Koj zašto preporuča knigata?  
who why recommended the-book

Similar examples are mentioned in Bošković 1998.
since, as implied by her analysis, binding of the trace of the fronted adjunct is not possible from the right-adjointed position in [Spec, CP].

Furthermore, the well-known “pure” superiority effects that involve two objects are generally taken to show that the ECP cannot be considered the appropriate theoretical means for an adequate analysis of superiority phenomena. Apart from empirical and conceptual problems associated with an ECP account of multiple wh-fronting, it should be noted that the explanatory claim of such an account refers only to ordering restrictions on fronted wh-elements rather than to the question of what motivates multiple overt movement of wh-elements in languages like Bulgarian and Romanian as opposed to overt movement of only one wh-element in languages like English and German and multiple covert movement of wh-elements in languages like Japanese.

The latter question constitutes the main focus of minimalist analyses of multiple wh-fronting that try to account for the observed superiority effect in terms of an economy approach. Such an economy-based analysis of multiple wh-fronting is suggested by Richards’s (1997) complex theory of multiple wh-questions. Richards assumes that languages such as Bulgarian and Romanian (so-called CP-absorption languages) have multiple [Spec, CP]s while languages such as Serbo-Croatian and Polish (so-called IP-absorption languages) have multiple [Spec, IP]s. Basing his proposal on Chomsky’s (1995) notion of cyclicity, Richards suggests a featural theory of cyclicity according

---

5 Since in a Generalized Binding approach (at least nonreferential) adjuncts (unlike objects) do have an accessible SUBJECT and thus a governing category (Aoun 1986:31–32), they need to satisfy the binding requirements.

6 Two remarks are in order here. First, further empirical problems for Rudin’s account of multiple wh-fronting are pointed out by Comorovski (1989), who therefore defends an ECP approach, which (unlike Rudin’s analysis) adopts a disjunctive version of the ECP and assumes left-adjunction of fronted wh-elements to CP. Since ECP accounts of superiority effects have become obsolete for independent empirical and conceptual reasons, I do not discuss the empirical shortcomings of Comorovski’s analysis.

Second, the attempt to analyze the order of fronted wh-elements in a way analogous to an ECP account of superiority effects has to ignore some exceptions to Rudin’s generalizations where the order of the fronted wh-phrases is not completely fixed (Comorovski 1989, Billings and Rudin 1994). In what follows I will ignore these exceptions as well, assuming that they can be attributed to some interfering factors like free inversion, verb preposing, D-linking, phonological considerations (Bošković 1998), and base generation of different orders (see Muller 1997). It furthermore should be taken into account that the order of fronted wh-elements is also determined by occurrences of clitic doubling as found in Romanian, where the wh-elements care ‘which’, care + NP, and cine ‘who’ in its dative form cui, as well as all D-linked NPs, are obligatorily clitic-doubled so that their distribution in the left periphery of the clause cannot in fact be related to requirements of the ECP (Comorovski 1989:98, 108).

7 Whether or not wh-elements in situ (or their relevant features) move at LF is a much-debated issue. While Huang (1982), Watanabe (1992), and Chomsky (1993), among others, argue for LF movement of wh-in-situ, Baker (1970), Reinhart (1993), and Baek (1995) assume that wh-elements in situ are interpreted in situ. More recently, Garrett (1996) has presented interesting interpretive constraints on wh-in-situ, which he analyzes in terms of Beghelli and Stowell’s (1996) theory of phrase structure and which he takes to clearly show that wh-in-situ do in fact move at LF. The same result is achieved as a crucial consequence of Richards’s (1997) account of Japanese wh-questions. Note that the question of whether or not there is movement at LF takes a different form in a single-level model of movement. In such a model the relevant question is not whether movement takes place before or after Spell-Out but whether movement includes PF features or not. Covert movement then does not take place at a specific level; instead, it is just movement without the PF features, taking place at the same point as any other movement. Pesetsky (2000) describes pronunciation distinctions like the one between overt and covert phrasal movement as essentially phonological, characterizing different pronunciation patterns (e.g., in multiple questioning) in terms of “pronunciation rules.”
to which, after a higher XP moves to [Spec, CP], a lower YP can move to the lower [Spec, CP] as long as XP and YP are moving to check a strong feature of C_0. Richards’s account thus permits multiple attraction by a single attractor and derives countercyclic movement in exactly this case. It thus predicts that paths to multiple specifiers of a single attractor should cross, maintaining the base c-command relations (and base order) among elements that move to them. Multiple attraction by a single attractor is restricted by the ‘‘attractor-oriented’’ condition Shortest Attract, according to which the attractor has to attract the nearest available mover, as well as by the ‘‘mover-oriented’’ condition Shortest Move, according to which the attracted element is moved to the closest available landing site, which in the case of multiple specifiers is the lowest available specifier, to which the moved element then must be ‘‘tucked in.’’8 This theory captures a wide range of crosslinguistic interrogative phenomena and, in particular, correctly predicts that in languages like Bulgarian and Romanian a fronted wh-subject must precede fronted wh-objects and wh-adjuncts.

Multiple attraction is subject to a further constraint that plays a crucial role in Richards’s analysis of multiple wh-questions: the Principle of Minimal Compliance. In a nutshell, this principle states that a certain constraint need be obeyed only once in a clause. If, for example, a certain dependency D obeys constraint C (e.g., the Subjacency Condition), then, for the rest of the derivation, any other dependency D’ can ignore the elements (e.g., bounding nodes) that are relevant for determining whether D obeys C. The Bulgarian example (2) illustrates that the Subjacency violation in (2a) caused (with respect to the [+wh] C) by extraction from a complex NP can be ‘‘neutralized’’ by adding a wh-element that obeys Subjacency (with respect to the [+wh] C).

(2) a. *Koja kniga, otreče senatorat [mālvata če iska da zabrani t₁]? which book denied the-senator the-rumor that (he)-wanted to ban ‘Which book did the senator deny the rumor that he wanted to ban?’
   b. ?Koj senator koja kniga, otreče [mālvata če iska da zabrani t₁]? which senator which book denied the-rumor that (he)-wanted to ban ‘Which senator denied the rumor that he wanted to ban which book?’
   (Richards 1997:242)

The Principle of Minimal Compliance is thus designed to account for so-called additional-wh phenomena that occur overtly in languages like Bulgarian and covertly in languages like Japanese (see section 3).

---

8 For a similar analysis of overt multiple wh-fronting in Bulgarian, see Mulders’s (1997) theory of ‘‘mirrored specifiers,’’ which is based on the idea that a checking configuration between a maximal projection and a head can only be established if the maximal projection occupies a position adjacent to the node immediately dominating the head and its complement. In order to fulfill this requirement, a second specifier has to be inserted into a position below the first specifier, thus mirroring the order derived when each successive specifier is inserted at the root of the structure.
Despite the undeniable explanatory force of such an economy-based analysis of multiple wh-fronting, there are both conceptual and empirical problems associated with it. One of the conceptual problems concerns an important consequence of Richards's notion of Shortest Move, namely, that the higher and the lower specifier can no longer be “equidistant.” Even though Richards points out empirical phenomena that are supposed to provide independent evidence in support of this consequence, undesirably two crucial exceptions need to be allowed. First, equidistance must still be assumed for multiple A-specifiers, as shown by Ura’s (1996) analysis of superraising and Grewendorf and Sabel’s (1999) analysis of scrambling in Japanese, and as entailed by Richards’s own analysis of A-movement phenomena. Second, the application of the Principle of Minimal Compliance to Shortest Move would have the effect of making multiple specifiers equidistant after Shortest Move has been satisfied by one instance of movement to specifier position; it must therefore be stipulated that the Principle of Minimal Compliance does not apply to the condition Shortest Move.

There are also empirical problems associated with the Principle of Minimal Compliance. For example, this principle predicts that an additional wh-word outside an island remedies the island violation not only in cases like (2) but also in cases where the wh-element originating from the island is an adjunct. However, this prediction is refuted by facts from both Bulgarian and Japanese.

Further empirical problems with Richards’s approach are pointed out in section 3, where I argue for an alternative analysis of so-called additional-wh effects.

Finally, I wish to briefly mention two Optimality Theory accounts of multiple wh-fronting proposed by Müller (1997) and Ackema and Neeleman (1998). Muller (1997) suggests a general constraint, PARALLEL MOVEMENT, according to which movement must be order-preserving in the sense that asymmetric c-command relations between arguments must be preserved from one level of representation to the next. This constraint, which belongs to the class of so-called faithfulness constraints, consonant with the idea that movement must be order preserving.
constraints, is said to be violable, universal, and low-ranked. In addition to a wide range of empirical evidence for this constraint accumulated from constructions such as pronominal shift in Danish and pronoun fronting in German, Müller offers an account of \( wh \)-movement in Bulgarian based on the assumption that fronting of \( wh \)-phrases in Bulgarian proceeds via left-adjunction to CP as well as on a particular ranking of the four constraints \( Wh \)-Criterion (which requires overt movement of \( wh \)-phrases), Projection Principle (which rules out adjunction to CP and C), Parallel Movement, and Derivational Economy (which disallows S-Structure movement).

The principal conceptual differences between this account and minimalist approaches are that (a) it proceeds from a theory of levels no longer assumed in the Minimalist Program and (b) it is based on a notion of explanation that differs from the explanatory claim made by minimalist approaches. Thus, conceptual comparisons between an optimality account and the minimalist account advocated here are difficult. Therefore, I just wish to point out an empirical difference: Müller’s account succeeds in correctly deriving the fact that Bulgarian has overt multiple \( wh \)-fronting and that fronted \( wh \)-elements display a certain order reminiscent of superiority effects. However, it is neither able to capture important crosslinguistic generalizations about additional-\( wh \) effects nor able to establish any correlation between parametric properties that determine the overt and covert nature of multiple \( wh \)-fronting and morphological properties of the \( wh \)-elements involved.\(^{11}\)

In what follows I wish to maintain Rudin’s intuition that there is a parallel between overt and covert multiple \( wh \)-fronting without modeling the structure of overt multiple fronting on the familiar pattern of traditional LF derivations. I wish to suggest an analysis of overt multiple \( wh \)-fronting that does not proceed from the idea that \( wh \)-elements move individually to [Spec, CP]. I will argue that it is instead a cluster of \( wh \)-elements that moves to [Spec, CP] in languages such as Bulgarian and Romanian, and that assuming the formation of a \( wh \)-cluster prior to \( wh \)-movement to [Spec, CP] avoids the conceptual and empirical problems of the other approaches mentioned here and enables us to answer the question of what motivates multiple \( wh \)-fronting. Using evidence related to the so-called additional-\( wh \) effect, I will then show that the formation of a \( wh \)-cluster prior to \( wh \)-movement to [Spec, CP] also characterizes multiple \( wh \)-questions of \( wh \)-in-situ languages such as Japanese.

3 A Theory of Multiple \( Wh \)-Fronting

In this section I outline a new theory of multiple \( wh \)-fronting based on ideas suggested in Grewendorf and Sabel 1999 and Grewendorf 1999. I make use of this theory in my analysis of multiple \( wh \)-questions in languages such as Bulgarian, Japanese, and German. The crucial claim is that multiple \( wh \)-fronting does not consist of moving individual \( wh \)-elements separately to [Spec, CP];

\(^{11}\) Ackema and Neeleman’s (1998) optimality account of multiple \( wh \)-fronting is based on a specific ordering of the constraints Q-Marking (which is a constraint about the element to be marked as a \( wh \)-question), \( S \) TAY (which is interpreted in such a way that the greater the distance between chain links, the more violations of this constraint result, thus incorporating the basic idea of the Minimal Link Condition), and \( Q \)-Scope (whose high ranking relative to \( S \) TAY determines overt multiple \( wh \)-fronting).
instead, these individual elements form a *wh*-cluster prior to movement to [Spec, CP], and a single complex *wh*-element therefore undergoes movement to [Spec, CP] in ‘‘multiple’’ *wh*-fronting languages. I assume that the formation of *wh*-clusters and hence the existence of multiple *wh*-fronting is driven by a lexical property of the *wh*-words themselves, in the following way: owing to morphological properties of *wh*-words, or, in terms of feature theory, owing to the presence of a particular feature in their head, *wh*-phrases may establish internal operator positions (as adjoined positions), and the feature in question (in a sense to be specified below) ‘‘attracts’’ *wh*-elements to these positions in multiple *wh*-constructions. In terms of the minimalist theory of Chomsky (1995), this idea can be stated as the ‘‘*Wh*-Cluster Hypothesis’’ in (5).\(^{12}\)

\[\text{(5) *Wh*-Cluster Hypothesis}\]

A particular feature of a *wh*-element acts as a checker for other *wh*-elements (*wh*-arguments as well as *wh*-adjuncts).

For reasons that will become clear later, there are two exceptions to this *wh*-checking potential: *wh*-elements located in [Spec, CP] and *wh*-adjuncts are unable to act as checkers for other *wh*-elements.\(^{13}\)

It is a consequence of the *Wh*-Cluster Hypothesis that *wh*-elements can act as landing sites for *wh*-movement. The crucial questions to be raised at this point are these: what are the morphological properties responsible for this property of *wh*-phrases, and how exactly does the checking mechanism operate that ensures the formation of fronted *wh*-clusters? To answer the first question, I will resort to an observation made by Cheng (1991, 1997), according to which in multiple *wh*-fronting languages, *wh*-words can be used to form indefinite NPs if a particular affix is added, as in the Bulgarian data in (6). I will refer to this observation as Cheng’s Generalization.

\[\text{(6) kój} \quad \text{‘who’} \quad \text{njákoj} \quad \text{‘someone’}\
\quad \text{kudé} \quad \text{‘where’} \quad \text{njákude} \quad \text{‘somewhere’}\
\quad \text{kogá} \quad \text{‘when’} \quad \text{njákoga} \quad \text{‘sometime’}\
\quad \text{kakovó} \quad \text{‘what sort of’} \quad \text{njákakvo} \quad \text{‘some sort of’}\]

Cheng concludes from this observation that the *wh*-words themselves in these languages do not have any inherent quantificational force and in this respect are similar to indefinite NPs.\(^{14}\) She

\[^{12}\text{For reasons of ‘‘economy,’’ having to do with a better outcome with respect to STAY, Ackema and Neeleman’s (1998) optimality account derives the structure of multiple *wh*-fronting in Bulgarian by resorting to a similar idea. They first adjoin the *wh*-object to the *wh*-subject and then move this cluster as a whole to the higher specifier position. However, they restrict the formation of a *wh*-cluster to this particular case and do not derive it from independent properties of the language involved. Independent evidence for the existence of cluster formation as a feature-driven process can be gained from phenomena such as verb incorporation, clitic clusters, and multiple topicalization. These phenomena display fixed order effects that can be accounted for in terms of feature-driven cluster formation, as shown in Sabel 1998:chap. 6.}\]

\[^{13}\text{The fact that arguments are possible adjunction sites seems to conflict with the *θ*-theoretic argument that Chomsky (1986) offers in support of his constraint on adjunction, according to which adjunction is possible only to nonarguments. I will argue in section 4 that there are no good reasons to assume that adjunction to an element in a *θ*-position is impossible. Note furthermore that according to standard assumptions (Saito 1994), a *wh*-phrase adjoined to an element in an *A*-position binds its trace—in contrast to adjunction to an element in an *A*-position such as [Spec, CP], which is standardly taken not to permit binding of a trace.}\]

\[^{14}\text{For a similar view with respect to Japanese, see Nishigauchi 1990.}\]
therefore assumes that the interrogative force of wh-words in multiple wh-fronting languages is determined by a null determiner that bears a wh-feature so that wh-phrases are taken to have the structure represented in (7).

\[
\begin{aligned}
(7) & \quad DP \\
& \quad D' \\
& \quad D^0 \quad \text{NP} \\
& \quad \emptyset_{\text{wh}} \quad \text{wh-word}
\end{aligned}
\]

According to Cheng, this null determiner must be licensed by being in a specifier-head relation with a C[+wh], which forces wh-movement to [Spec, CP].

Regarding the second question, which concerns the checking mechanism, a problem arises in view of the different stages of the minimalist theory, which by themselves cannot be considered fully elaborated theories and in which the view of feature checking has undergone considerable changes. In what follows I will express Cheng’s basic idea in terms of the most recent version of the minimalist theory, the “probe-goal” system of Chomsky (2000, to appear), which dispenses with checking theory altogether and assumes instead that there is an operation called Agree that establishes agreement between uninterpretable features of the former “checker” (now called the “probe”) and the matching features of a local “goal” and erases uninterpretable features of probe and goal under matching. Since matching is understood as feature identity, Case (and lexical category), not being a feature of the functional heads that include a probe, cannot enter into Agree; as a result, operations are not induced by Case-checking requirements. Rather, structural Case is taken to be an uninterpretable feature of the goal which serves to “activate” the goal of a probe, allowing it to implement some operation (Agree, Move), and which too is eliminated under Agree as a kind of “ancillary deletion.” As for the uninterpretable features that serve as the probes of functional heads, Chomsky distinguishes between f-features, which yield the Case agreement system and induce A-movement, and P-features, which concern the peripheral system (e.g., force, topic, focus) and involve A-bar-movement. Movement is induced by the (uninterpretable) selectional feature EPP, which is satisfied by Merge of a phrase P(G) determined by the goal G of the probe. Features cannot move or be attracted.

In this system checking reduces to deletion under matching with an active local goal and “ancillary deletion” of the uninterpretable feature that rendered the goal active. The analysis of wh-movement in terms of the probe-goal system proceeds in a way similar to the analysis of A-movement. By analogy to interpretable f-features of an NP, wh-phrases have an interpretable P(peripheral)-feature [Q] that matches the uninterpretable probe [Q] of a complementizer in the final stage. Furthermore, wh-phrases have an uninterpretable feature [wh], analogous to the un-
interpretable Case feature of an NP, that activates the \textit{wh}-phrase as a goal and deletes along with the uninterpretable feature [Q] of the probe after matching of probe and goal. Again, choice of Move follows from the presence of an EPP-feature in C, which may be redundant as a reflex of the uninterpretable P-feature (Chomsky 2000:144). Even though my analysis of multiple \textit{wh}-fronting mainly refers to the basic ideas of this system, I will sometimes use the traditional terminology (‘‘checking,’’ ‘‘attracting’’) to express things in a more familiar and less technical way.

Against the background of this theory, Cheng’s analysis (7) of \textit{wh}-words in languages such as Bulgarian can be expressed in minimalist terms as follows: it is a parametric property of \textit{wh}-words in a language like Bulgarian that unlike \textit{wh}-elements in a language like English, their D-head is endowed with an EPP-feature, which I take to be identical with the uninterpretable P-feature [Q]. As a consequence, not only interrogative C but also the \textit{wh}-elements themselves function as a target of \textit{wh}-movement, providing an operator position owing to the presence of an uninterpretable feature [Q]. I assume that \textit{wh}-phrases ‘‘attracted’’ by the uninterpretable probe of a \textit{wh}-element adjoin to this \textit{wh}-element, the uninterpretable features of probe and goal being deleted as a consequence of Agree. The internal structure of a \textit{wh}-phrase can then be represented as in (8).

\[
\begin{array}{cc}
\text{DP}[Q_{\text{int}}, \text{wh}_{\text{int}}] & \\
\text{D}[EPP=Q_{\text{int}}] & \text{NP}
\end{array}
\]

(8) is intended as a representation of Cheng’s basic idea, according to which in languages like Bulgarian the interrogative properties of \textit{wh}-elements are provided by their null determiner rather than by inherent lexical properties of \textit{wh}-words as assumed for English in Chomsky and Lasnik 1993:536. Note that there is no contradiction in D’s bearing an interpretable as well as an uninterpretable feature [Q]. The former is the feature that is relevant for semantic interpretation; the latter is a selectional feature that seeks an XP to merge with the category it heads. I suppose that

---

\[\text{The nonfinal stages of successive-cyclic movement are ‘‘indirectly feature-driven’’ in the sense that they are needed to guarantee eventual convergence as determined by the so-called Phase-Impenetrability Condition (Chomsky 2000:108). By analogy to defective T, which bears only the EPP-feature [person], one may assume that intermediate C is endowed with an uninterpretable defective P-feature, which needs deletion but is unable to delete the uninterpretable activating feature of the goal, since this deletion requires a full set of features in the probe. For an analysis of successive-cyclic movement as an instance of directly feature-driven movement, see Fanselow and Mahajan 1999, where the presence of a categorial feature in C is assumed, and Sabel 1998, where successive-cyclic movement is attributed to the presence of a focus feature.}

\[\text{Notice that the relevant operator position cannot be the specifier position of the \textit{wh}-DP, since the specifier position of a DP is usually taken to be an A-position. The fact that Chomsky (2000, to appear) assigns an EPP-feature to C suggests that this feature is not necessarily associated with a thematic requirement. It is therefore not implausible that an EPP-feature in this sense can be checked in an adjoined position. That ‘‘checking’’ is in fact possible in adjoined positions is shown by Grewendorf and Sabel (1999), who capitalize on the relevant differences between adjoined positions and specifier positions. There is no principled difference between the present analysis of \textit{wh}-clusters and an alternative suggested by a reviewer according to which \textit{wh}-cluster formation does form a specifier that is not an A-position.}

\[\text{Recall that the features of a syntactic object are always the features of the label, so that all the features in (8) should be represented as features of the D-head. I have chosen the representation in (8) to make it easier for the reader to correlate this structure with the characterization of \textit{wh}-elements given in Chomsky 2000.} \]
an analogous distribution of [Q]-features will have to be assumed for an interrogative CP. There is a requirement that the probe, which consists of uninterpretable features of the label LB(K) of a syntactic object K, seek a matching goal within the domain of LB(K) (Chomsky, to appear). It should be clear that in the case of (8) the domain of the probe must be the domain of the wh-element rather than the domain of its D-head.

I will show that the relevant lexical property of wh-words may induce visible as well as invisible formation of wh-clusters. Within the most recent version of the Minimalist Program, this difference could be expressed as a difference in pronunciation rules along the lines suggested by Pesetsky (2000), where pronunciation distinctions like the one between overt and covert phrasal movement are described as essentially phonological. In traditional terms we could say that this property occurs in both ‘‘strong’’ and ‘‘weak’’ variants so that a typological distinction can be derived between languages with overt multiple wh-fronting like Bulgarian and Romanian and languages with covert multiple wh-fronting like Japanese.

Let us first consider a language where wh-cluster formation applies in the overt syntax. In Bulgarian multiple wh-questions, wh-elements must undergo overt fronting in such a way that the wh-subject precedes wh-objects and wh-arguments precede wh-adjuncts (Rudin 1985, 1988).18 In the Bulgarian example (9), three wh-phrases have undergone overt fronting in a double object construction, which (as observed in Rudin 1985) typically results in a fixed order.19 According to traditional analyses, cases such as (9) involve movement of the wh-subject to [Spec, CP], followed either by right-adjunction of the other wh-phrases to the wh-subject in [Spec, CP] or by movement of these wh-phrases to other specifier positions of CP. Deferring the technical details of the cyclic process of structure building (as assumed in Chomsky 1995, 2000, to appear) until later, I illustrate in (10) the general idea of a derivation of (9) based on the Wh-Cluster Hypothesis.

18 Bošković (1998) mentions a Bulgarian example in which the fronted wh-phrases occur in the order subject-adjunct-object. I will return to this example (see footnote 29). A reviewer mentions a Bulgarian example with one wh-phrase fronted and two wh-phrases left in situ and points out that it is not possible to front two wh-phrases and leave one in situ. However, my informants told me that Bulgarian wh-phrases in situ are to be interpreted as indefinites and need not be answered since they are not in the focus of the question. This corresponds to the generalization stated by Bošković (1998), who points out that wh-movement is obligatory in Bulgarian in all contexts but that D-linked wh-phrases can remain in situ. I therefore conclude that an analysis of the example given by the reviewer has to refer to one of the interfering factors that I wish to abstract away from (see footnote 6).

19 Bošković (1998) observes that in a ternary wh-question, the second and third wh-words are freely ordered, which Richards (1997:281) accounts for in terms of his Principle of Minimal Compliance: if the principle Shortest (Attract) is fulfilled by attraction of the nearest wh-element, this principle no longer applies to other wh-elements. However, Müller (1997) points out, as do Ackema and Neeleman (1998), that Bošković’s observation could also be explained by assuming different underlying orders of the objects. An explanation along these lines receives independent support from the observation confirmed by all my Bulgarian informants: namely, that the order of fronted indirect and direct wh-objects is also free when these objects constitute the first and second instances of wh-movement.

(i) Na kogo kakvo e dal Ivan?
   to whom what is given Ivan
   ‘What has Ivan given to whom?’
(ii) Kakvo na kogo e dal Ivan?
    what to whom is given Ivan

For reasons of space and coherence, I will not go into the details of the representation of double objects.
Because of the uninterpretable feature [Q] in the D-head of kogo, the direct wh-object kakvo first overtly adjoins to the indirect wh-object kogo. Matching of the [Q]-features leads to deletion of the uninterpretable Q-probe in the ("attracting") D-head of kogo and of the uninterpretable feature [wh] that activates the goal. In the next step the uninterpretable feature [Q] of the wh-subject "attracts" the wh-cluster so formed, matching with the interpretable feature [Q] of the probe koj and the uninterpretable "activating" feature [wh] of the goal. Finally, the uninterpretable feature [Q] of C matches with the interpretable feature [Q] of the wh-subject and "attracts" the cluster consisting of the three wh-phrases to [Spec, CP], with deletion of uninterpretable features proceeding accordingly. Notice that no problem arises regarding the uninterpretable feature [Q] of the D-head of the lowest wh-phrase. In the probe-goal system we can assume that deletion of uninterpretable features induced by the operation Agree affects not only the uninterpretable feature [Q] of the probe and the "activating" uninterpretable feature [wh] of the goal but also an uninterpretable matching feature [Q] of the goal (should there be one), with deletion of involved uninterpretable features thus operating in an "all or none" fashion (Chomsky 2000:125).

20 Since the uninterpretable feature [Q] is a selectional feature like the EPP-feature [person] of T and of expletive there, there is nothing wrong with its ancillary deletion in the goal. The situation can be related to EPP checking of there and T, where either one can in principle have its uninterpretable feature deleted depending on the properties (d-completeness) of the other and on the domain restriction for the goal. However, if deletion of goal features takes place, it has to affect all uninterpretable features. Note that the uninterpretable feature [wh] of the probe is not involved in the operation Agree since only the goal is activated and that activation can only take place if there is a probe "seeking" a goal since computation is driven by a probe.

21 At this point the question may arise whether adjunction to a specifier as assumed in derivation (10) obeys the
Let us now turn to languages where wh-cluster formation applies in the covert syntax. Such a language is Japanese. In Japanese, too, the interrogative reading of a wh-word can alternate with an indefinite reading (Cheng 1991:80).

(11) dare ‘who’, ‘someone’
nani ‘what’, ‘something’

Cheng (1991) concludes from this observation that wh-words in these languages do not have any inherent quantificational force either (for a similar conclusion, see Aoun and Li 1993, Ouhalla 1996). Even though this conclusion may be taken to imply that in these languages, too, wh-phrases contain a null determiner \[ \{D \theta \{+\text{wh}\}\} \] that must be licensed, Cheng does not conclude that these languages should be multiple-fronting languages as well. The fact that Cheng does not draw this conclusion leads her to stipulate without further argument that the licensing of \[ \{D \theta \{+\text{wh}\}\} \] is an S-Structure requirement that takes effect only in languages with overt wh-fronting.

From the fact that wh-elements of languages like Japanese also do not possess any inherent quantificational force, I prefer to draw a different conclusion: namely, that (like wh-phrases in languages such as Bulgarian and Romanian) wh-phrases in these languages should be analyzed as having the structure in (7). In other words, I assume that just like their overt wh-fronting counterparts in Bulgarian and Romanian, wh-elements in languages like Japanese contain an operator position (as an adjoined position) so that in these languages as well, the D-head of wh-phrases acts as a ‘checker’ and ‘attracts’ wh-phrases before moving to [Spec, CP]. Within the framework of Chomsky 1995, the proper way to express the crucial difference between languages like Japanese and overt wh-fronting languages would be in terms of feature strength: we could say that in languages like Japanese the relevant D-feature of the wh-phrases is weak. It is not entirely clear how this difference could be expressed in the minimalistic framework of Chomsky 2000, to appear. The notion of strength, designed to force violations of the principle Procrastinate, appears to have no place in the new system. Overt movement seems to be determined by the presence of the EPP-feature, but covert movement still exists. Following ideas of Groat and O’Neil (1996), Chomsky claims that the nonexistence of an LF component within narrow syntax has the consequence that operations that have or lack phonetic effects are interspersed so that the operation Agree alone, when not combined with Merge, can precede overt operations. As for wh-in-situ constructions, Chomsky (2000:128) still contemplates an analysis in terms of movement along

---

Strict Cycle Condition. As it does not extend the phrase marker, it certainly violates the Extension Condition of Chomsky 1993. But recall that some relaxation of this condition was required on independent grounds since it does not permit head-to-head adjunction either. Solutions to this problem permit head adjunction as well as the adjunction proposed in (10). Bobaljik (1995) shows that the cyclicity problem in question results from the notational conventions associated with phrase markers and that it does not arise if the syntactic computation is solely viewed as creating terms rather than phrase markers, that is, if Merge operates on sets of terms rather than on phrase markers. For a similar proposal see Takano 1998:sec. 4.3 (Takano’s proposal is illustrated in section 4 below). Chomsky (2000) notes an ambiguity of cyclicity (Extension Condition and Local Merge) and suggests a kind of economy condition requiring that operations not change basic relations (sisterhood and c-command) involving the label that projects. Other suggestions with similar effects are made by Ackema and Neeleman (1998:462) and Bošković (1998:17). I therefore conclude that the strict notion of cyclicity assumed in Chomsky 1995:234, which disallows adjunction to specifiers, cannot be sustained. The problem no longer arises in the framework of Chomsky, to appear, where the level of strong phases plays the crucial role for cyclic computation.
lines proposed by Watanabe (1992) in which ‘‘[wh] pied-pipes only the head (overtly or covertly).’’ Nevertheless, it remains unclear whether and how the effects of the former notion of strength ‘‘can be fully captured in minimalist terms or remain as true imperfections’’ (Chomsky 2000:132). As an alternative to feature strength, it may be assumed that the overt/covert distinction is essentially phonological and can be accommodated in terms of pronunciation rules that determine in which position elements of chains are pronounced (for this idea, see Pesetsky 2000). In what follows I tentatively assume that the effects of the former notion of strength can be captured by positing some parametric property of the EPP-feature (Chomsky 2000:109); but in order not to complicate the following discussion with new theoretical concepts that have not yet been worked out satisfactorily, I will continue to refer to feature strength, assuming that my use of covert movement can be translated into an analysis where covert movement is understood as invisible ‘‘overt’’ movement.

There is independent evidence for the proposal that the Wh-Cluster Hypothesis applies in a covert way in Japanese: namely, the fact that islands for covert extraction of wh-adjuncts can be neutralized in certain cases. Consider first the examples in (12), where the occurrence of the wh-adjunct naze ‘‘why’’ within a complex NP island (12a) or an adjunct island (12b) leads to ungrammaticality.

(12) a. *John-wa [NP IP sono hon-o naze katta] hito-o sagasiteru no?  
   John-top that book-acc why bought person-acc looking-for q  
   (q John is looking for [the person [that bought that book why]])  
   (Saito 1994:204)

b. *John-wa [PP IP Mary-ga sono hon-o naze katta] kara okotteru no?  
   John-top Mary-nom that book-acc why bought since angry q  
   (q John is angry [because Mary bought that book why])  
   (Saito 1994:205)

Interestingly, if a wh-argument is added within the island in a position higher than the wh-adjunct, these sentences improve considerably. This is illustrated in (13). 22

(13) a. ??John-wa [NP IP nani-o naze katta] hito-o sagasiteru no?  
   John-top what-acc why bought person-acc looking-for q  
   (q John is looking for [the person [that bought what why]])  
   (Saito 1994:204)

b. ??John-wa [PP IP Mary-ga nani-o naze katta] kara okotteru no?  
   John-top Mary-nom what-acc why bought since angry q  
   (q John is angry [because Mary bought what why])  
   (Saito 1994:205)

22 The marginality of (13a–b) is due to independent causes: namely, to a violation of Subjacency (see Saito 1994: 213).
Assuming that the *wh*-phrases in (13a–b) leave the islands and move to [Spec, CP] of the matrix clause one by one, one would expect extraction of the *wh*-adjunct in these sentences to result in the same sort of ungrammaticality observed in (12). Given that complex NPs and adjuncts are not islands for covert extraction of *wh*-arguments, we may therefore suggest that the improved status of (13a–b) is due to the fact that the *wh*-adjunct in these sentences adjoins to the *wh*-argument prior to movement to [Spec, CP] so that it is in fact the resulting *wh*-cluster that is actually extracted from the island. 23

Further evidence for *wh*-cluster formation in Japanese is provided by Maki’s (1994) observation that a *wh*-adjunct can occur in situ in an embedded declarative clause, as in (14a), but that the addition of a *wh*-subject in the matrix clause leads to ungrammaticality, as in (14b).

   John- NOM Mary- NOM why was fired COMP think Q tell please
   (Please tell me [Q John thinks [that Mary was fired why]])

   b. ?*[Dare-ga [Mary-ga naze kubininatta to] omotteiru ka] osiete kudasai.
      who- NOM Mary- NOM why was fired COMP think Q tell please
      (Please tell me [who thinks [that Mary was fired why]])

(Maki 1994)

Obviously, the presence of the *wh*-subject in the matrix clause has a blocking effect on the covert extraction of the *wh*-adjunct out of the embedded declarative clause.

The *Wh*-Cluster Hypothesis provides an account for the ungrammaticality of (14b). If the *wh*-adjunct *naze* must undergo checking by the matrix *wh*-subject *dare-ga* before moving to the matrix [Spec, CP], the ungrammaticality of (14b) follows from the fact that the adjunct must use the embedded [Spec, CP] as an intermediate landing site to satisfy its usual locality restrictions. But when the *wh*-adjunct passes through the embedded [Spec, CP], it cannot undergo any further adjunction, as it has been argued (see Hoekstra and Bennis 1989, Müller and Sternewald 1993, Grewendorf and Sabel 1994, Sabel 1998) that an element that has been moved to [Spec, CP] would actually undergo a kind of improper movement if subjected to any subsequent adjunction operation. This derivation is therefore ruled out by whatever rules out improper movement. 24

As mentioned earlier, there are two exceptions to *wh*-cluster formation: *wh*-elements located in [Spec, CP] and *wh*-adjuncts are unable to “attract” other *wh*-elements. These two exceptions

---

23 For an alternative analysis of data such as (13a–b) in terms of the Principle of Minimal Compliance, see Richards 1997: chap. 5. But note that such an account of (13) predicts that an example such as (12b) should also be rescued by (permissible) overt extraction of a *wh*-argument out of the island, contrary to fact. Richards’s assumption that the Principle of Minimal Compliance may apply in a “temporary” manner does not solve the problem in a nonstipulative way.

24 The *Wh*-Cluster Hypothesis receives further support from the fact that an alternative analysis of “additional-*wh*” effects—namely, Richards’s (1997) Principle of Minimal Compliance—offers no account for the ill-formedness of (14b). If the creation of well-formed dependencies makes parts of the structure immune to certain constraints for the rest of the derivation, as Richards’s principle states, then this principle predicts (14b) to be well formed, since the well-formed dependency of matrix C and matrix *wh*-element should suffice to ensure grammaticality irrespective of whether *naze* is located in a declarative complement as in (14b) or in a *wh*-island, where it is disallowed as well (see Richards 1997: 291).
follow from independent properties associated with the way the probe-goal system operates in the case of wh-elements. If the uninterpretable feature [Q] of a wh-probe leads to deletion of the uninterpretable ‘activating’ feature [wh] as well as of the uninterpretable feature [Q] of the goal, then such a ‘checked’ goal has lost its ability to act as a probe itself. It follows that a wh-element that has moved to [Spec, CP] (or another operator position) and entered the operation Agree with the uninterpretable feature [Q] of C (or another [Q]-bearing head) is unable to form a wh-cluster by having another wh-element adjoined to it. As far as wh-adjuncts are concerned, we can assume that they lack a D-head that could be endowed with an uninterpretable feature [Q]. Consequently, they cannot function as a probe and thus do not provide an operator position (although they function as a goal owing to the presence of the uninterpretable feature [wh]).

The Wh-Cluster Hypothesis combined with its two exceptions provides a simple account of the so-called additional-wh effect in Japanese, as illustrated by examples (15) and (16) (Watanabe 1992).

(15) *John-ga naze nani-o katta no?
   John-NOM why what-ACC bought Q
(16) John-ga nani-o, naze t; ketta no?
   John-NOM what-ACC why bought Q
   ‘Why did John buy what?’

Example (15) shows that in Japanese multiple wh-questions a wh-adjunct may not occur as the highest wh-element of the clause. In the well-formed example (16) the wh-object nani-o has been scrambled to a position in front of the wh-adjunct.

Under the theory proposed here, (15) is ill formed because there is no way for the wh-object nani-o to have its uninterpretable feature [wh] deleted. Nani-o is not allowed to adjoin to the wh-adjunct, regardless of whether the latter is in its overt position or has undergone covert movement to [Spec, CP]. Furthermore, nani-o is not allowed to cross naze and move to [Spec, CP] first. Such movement would not only violate the Minimal Link Condition (or some kind of ‘intervention constraint,’ according to Chomsky (2000, to appear)), it would also leave the uninterpretable feature [wh] of the wh-adjunct undeleted. The well-formedness of (16) can be accounted for along the following lines. Since the short-scrambled nani-o in (16) does not occupy a ([Q]-) operator position but instead occupies an A-position (Saito 1992, Grewendorf and Sabel 1999), the wh-adjunct naze in (16) is allowed to adjoin to the wh-object nani-o, with the operation Agree inducing deletion of the uninterpretable feature [Q] of the probe nani-o and the uninterpretable features

---

25 This property may be related to the nonreferential nature of adjuncts, which according to Cinque (1990) and Rizzi (1990) prevents wh-adjuncts (as opposed to wh-arguments) from leaving (weak) islands. The fact that a wh-adjunct can leave islands when adjoined to a wh-argument clearly shows that wh-adjuncts can be parasitic on the referential property of wh-arguments. As assumed by Cinque (1990) and Rizzi (1990, 1992), referentiality is the crucial factor that determines extraction from (weak) islands.

26 Empirical evidence for the prohibition against adunction to a wh-adjunct can be derived from the fact noted by Rudin (1988) (see also Cheng 1991, Bošković 1994) that overt multiple-fronting languages do not allow the fronting of more than one adjunct. The only way to have more than one adjunct is to have a preposed conjoined phrase containing two adjuncts.
of the goal naze. Then, the wh-cluster so formed is ‘‘attracted’’ by the uninterpretable feature \([Q]\) of C, with deletion of the uninterpretable feature \([Q]\) of C and the uninterpretable feature \([wh]\) of nani-o.\(^{27}\)

The present theory of covert multiple wh-fronting in Japanese furthermore allows us to account for a clause-boundedness constraint on the additional-wh effect that Saito (1994) has taken to be counterevidence to Watanabe’s (1992) account of the additional-wh effect in terms of his Principle of Relation Preservation.\(^{28}\) The relevant examples are given in (17) and (18).

\[\begin{align*}
\text{(17) a. } & \text{*Naze dare-ga Mary-ni } [\text{CP John-ga sono hon-o katta to}] \text{ itta no?} \\
& \text{why who-NOM Mary-DAT John-NOM that book-ACC bought COMP said Q} \\
& \quad (Q \text{ who told Mary [that John bought that book] why)} \\
\text{b. } & \text{Dare-ni, naze dare-ga } t_i [\text{CP John-ga sono hon-o katta to}] \text{ itta no?} \\
& \text{who-DAT why who-NOM John-NOM that book-ACC bought COMP said Q} \\
& \quad (Q \text{ who told whom [that John bought that book] why)} \\
\text{(18) a. } & \text{*Sono hon-o_i naze dare-ga Mary-ni } [\text{CP John-ga } t_i \text{ katta to}] \text{ itta no?} \\
& \text{that book-ACC why who-NOM Mary-DAT John-NOM bought COMP said Q} \\
& \quad (Q \text{ who told Mary [that John bought what] why)} \\
\text{b. } & \text{?*Nani-o_i naze dare-ga Mary-ni } [\text{CP John-ga } t_i \text{ katta to}] \text{ itta no?} \\
& \text{what-ACC why who-NOM Mary-DAT John-NOM bought COMP said Q} \\
& \quad (Q \text{ who told Mary [that John bought what] why)} \\
\end{align*}\]

As was the case with (15), in (17a) the wh-adjunct naze occurs as the highest wh-element of the clause. In (17b) the wh-adjunct is preceded by the wh-argument dare-ni, which has undergone short scrambling to an A-position (see Grewendorf and Sabel 1999). For (17a) and (17b), the present theory of multiple wh-fronting permits an account analogous to the one suggested for (15) and (16). In other words, (17b) is well formed because the wh-adjunct can adjoin to the

\(^{27}\) The question may arise whether the ‘‘additional-wh effect’’ in English illustrated by the contrast between (i) and (ii) could also be explained in terms of the Wh-Cluster Hypothesis.

(i) *What did who give to Mary?

(ii) What did who give to whom?

On the basis of Cheng’s Generalization, I have to assume that English does not allow wh-cluster formation, so that this contrast has to be explained in a different way; this conclusion is independently suggested by the fact that the ‘‘rescuing’’ wh-element that is added in English must occupy a position lower than the other wh-elements. While Kayne (1984) analyzes the English contrast as a Connectedness phenomenon, Pesetsky (2000) relates it to the nature of English interrogative complementizers and to specific properties of feature movement, viewed as (covert) movement of the wh-morpheme.

\(^{28}\) The Principle of Relation Preservation states that a structural relation (between wh-elements or quantificational expressions) established at a certain point in the derivation must be maintained throughout. As can easily be verified, this principle is fulfilled in (16) but violated in (15). However, as will be seen shortly, the principle cannot account for the clause-boundedness constraint on the additional-wh effect. An account of the additional-wh effect that is based on the same basic intuition as Watanabe’s account can be found in Hagstrom 1998:chap. 3.
short-scrambled object \textit{dare-ni}.\footnote{This analysis of (17b) requires that \textit{dare-ni} ‘‘attract’’ not only the \textit{wh}-adjunct but also the \textit{wh}-subject \textit{dare-ga}. The possibility of a probe ‘‘attracting’’ more than once is due to a parametric property of uninterpretable features that is equally involved in multiple head options for A-movement (Chomsky 2000:148, 149) and that can also be observed with overt \textit{wh}-movement in Bulgarian (Richards 1997; Bošković 1998).}

In contrast, (18b) shows that a \textit{wh}-adjunct cannot be ‘‘rescued’’ when it is preceded by a \textit{wh}-argument that has undergone long scrambling out of a finite clause. Watanabe’s Principle of Relation Preservation predicts (18b) to be well formed since the ‘‘S-Structure’’ configuration, in which the scrambled \textit{wh}-argument c-commands the \textit{wh}-adjunct, is maintained at the level of LF, at which the \textit{wh}-adjunct must first move to [Spec, CP] with the \textit{wh}-object adjoining to it, this in turn resulting in a configuration where the \textit{wh}-adjunct is c-commanded by the \textit{wh}-argument in Watanabe’s theory. Richards’s (1997) Principle of Minimal Compliance fares no better in accounting for (18b). Whatever explanation Richards’s approach offers for the ill-formedness of (17a), the Principle of Minimal Compliance provides no account for the difference between (17b) and (18b) since the relation between the matrix C and the highest \textit{wh}-phrase constitutes a well-formed \textit{wh}-dependency in either case.

The analysis of Japanese multiple \textit{wh}-fronting suggested here provides a simple alternative account of the ungrammaticality of (18b). Since it has been established on independent grounds that Japanese long scrambling from a finite clause into a [\(+\text{wh}\)] clause is obligatorily operator movement (to [Spec, CP] according to Takahashi (1993), to a position adjoined to IP according to Grewendorf and Sabel (1999)), this movement implies deletion of the uninterpretable feature [Q] of the long-scrambled \textit{wh}-argument \textit{nani-\textit{o}} so that the \textit{wh}-adjunct in (18b) cannot adjoin to \textit{nani-\textit{o}}. Consequently, it is impossible for the \textit{wh}-adjunct to have its uninterpretable feature deleted.

We can conclude that the theory of multiple \textit{wh}-fronting, suggested here as an analysis of multiple \textit{wh}-questions in languages like Bulgarian and Romanian on the one hand and Japanese on the other, receives empirical support both from languages in which all \textit{wh}-phrases must undergo overt fronting and from languages in which all \textit{wh}-phrases can remain in situ at the surface. Let us now turn to languages that restrict overt fronting in their multiple \textit{wh}-questions to a single \textit{wh}-phrase and require all of the other \textit{wh}-elements to remain in situ. German is a language of this type.

\section{Multiple \textit{Wh}-Fronting and Superiority in German}

As shown in the paradigm in (19), in German the majority of \textit{wh}-words can also be used as indefinites either without morphological alteration or by adding the prefix \textit{irgend}-.
Examples (20a) and (20b) illustrate the use of "bare" wh-words as indefinite subjects and indefinite objects, respectively.

(20) a. Gestern hat wer seinen Mantel vergessen.

yesterday has someone- NOM his coat- ACC forgotten

'Yesterday, someone forgot his coat.'

b. Gestern hat Peter wen überfahren.

yesterday has Peter- NOM someone- ACC run over

'Yesterday, Peter ran someone over.'

Recalling Cheng's Generalization, we can take this observation to imply that wh-words in German do not possess any inherent quantificational force; as a result, we can assume that the interrogative force of wh-phrases in German has to be attributed to the presence of a null determiner endowed with an EPP-feature (uninterpretable feature [Q]) in much the same way as in languages such as Bulgarian and Japanese. I therefore conclude that German wh-phrases likewise have the DP structure depicted in (8): an empty D-head equipped with an uninterpretable feature [Q] that provides an operator position and is able to "attract" wh-elements. Consequently, wh-phrases in German must undergo the same sort of "checking process" that operates overtly in Bulgarian and covertly in Japanese. Since, unlike Bulgarian and Japanese, German requires only one wh-phrase to move overtly to [Spec, CP] and does not allow more than one wh-phrase to move overtly to [Spec, CP], we have to assume that the feature [Q] is strong in C^0 but weak in D^0 so that we can derive the typological classification of languages represented in (21).[^30]

<table>
<thead>
<tr>
<th>Language</th>
<th>C_{[Q]}</th>
<th>D_{[Q]}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulg./Rom.</td>
<td>strong</td>
<td>strong</td>
</tr>
<tr>
<td>Japanese</td>
<td>weak</td>
<td>weak</td>
</tr>
<tr>
<td>German</td>
<td>strong</td>
<td>weak</td>
</tr>
</tbody>
</table>

[^30]: Ur Shlonsky and Chris Collins (personal communication) point to the instance missing from this classification: a language that has weak C_{[Q]} and strong D_{[Q]}. Such a language would exhibit an overt wh-cluster in a position lower than [Spec, CP]. Languages that in fact have this property are Russian (Stepanov 1998), Hungarian (E. Kiss 1994, Richards 1997), and possibly Persian (Raghibdoust 1994). Note that we cannot predict that the wh-cluster formed overtly in a language of this type should be able to occur in a position lower than a non-wh argument. It may well be that this cluster needs to be fronted to a position lower than [Spec, CP] for reasons other than wh-movement, which is the case with Russian (Stepanov 1998).
At this point a serious problem arises with respect to languages such as German. If only one wh-phrase is permitted to move overtly to [Spec, CP], and if adjunction to a wh-phrase located in [Spec, CP] is barred as one of the exceptions to the Wh-Cluster Hypothesis, how can the uninterpretable ‘‘activating’’ feature [wh] of an in-situ wh-phrase be deleted?

To solve this problem, we can resort to the previously mentioned single-level theory of movement, suggested by Groat and O’Neil (1996) and adopted in Chomsky 2000, to appear. According to this theory, the distinction between movement before and movement after Spell-Out is dispensed with; overt movement can then be characterized as movement that pied-pipes PF features, and covert movement can be characterized as movement that leaves phonological features behind. This view of the difference between overt and covert movement necessitates a reformulation of the nature of chain formation in the sense that the copy left behind after movement is not automatically marked as phonologically null. In principle, it may well be the case that the tail of a chain is visible while the head is invisible. The alternative nature of this model can thus be seen in the claim that movement (overt as well as covert) generally takes place at a single level and that operations that have or lack phonetic effects are interspersed.

The single-level theory of movement offers a solution to the problem concerning wh-cluster formation in German. Adopting such a theory, we can analyze the visible movement of a single wh-phrase in German as movement of a complex wh-phrase consisting of exactly one visible and (potentially) several invisible wh- phrases. If overt and covert wh-movement are distinguished not in terms of temporal precedence but in terms of this property of chains, we can assume that multiple wh-fronting in German takes place as the result of a strong feature [Q] in C0 and a weak feature [Q] in the D-head of wh-elements. Accordingly, a question with two wh- phrases like (22) can be derived as depicted in (23). The derivation in (23) is based on the clause structure assumed in Chomsky 1995. Underlining marks the visible elements of chains. The derivation of (22) proceeds as follows. In step 1 the wh-object wen covertly adjoins to the wh-subject wer to enter a checking relation with the weak feature [Q] of wer, resulting in the deletion of the uninterpretable feature [Q] of the probe wer and of the uninterpretable ‘‘activating’’ feature [wh] (including the uninterpretable feature [Q]) of the goal wen. This deactivation of the goal, of course, affects both the head and the tail of the chain so that the tail is no longer able to undergo wh-movement. In step 2 the overt copy of the wh-object moves to the outer specifier position of the light verb projection vP for reasons of φ- (and Case) checking (of the EPP, according to Chomsky 2000). In step 3 the cluster consisting of the overt wh-subject and the covert wh-object moves to [Spec, TP] for reasons of φ- (and Case) checking, and in step 4 this cluster moves to [Spec, CP], ‘‘attracted’’ by the (strong) uninterpretable feature [Q] of the probe C, with deletion of the uninterpretable ‘‘activating’’ feature [wh] of the goal wer.

Given this view of the nature of chain formation, the question arises of what determines which copy of a chain is pronounced. According to Groat and O’Neil (1996), the answer is provided by the reformulated definition of strong versus weak features: a link of a chain is specified for phonological features if it enters a checking relation with strong features. Consequently, the question of where categories are pronounced is not related to the timing of Spell-Out within the derivation but depends on the nature of the ‘‘attracting’’ feature (for an alternative analysis in terms of ‘‘pronunciation rules,’’ see Pesetsky 2000).
Let me further explain this derivation.

Step 1 involves adjunction to an element in a θ-position. This step obviously is at variance with the θ-theoretic argument on which Chomsky (1986) based his prohibition against adjunction to arguments and the semantic considerations with which he restricts the possibilities of adjunction in Chomsky 1995:322. However, note that the idea that adjunction to an argument is impossible because it would require an increase in the number of θ-roles necessarily rests on the assumption that the adjoined element, the term to which adjunction takes place, and the term resulting from adjunction each actually need a separate θ-role. This assumption is not at all self-evident. It is arguments that need θ-roles, but this configuration does not contain three arguments, only three terms. Thus, it would not necessarily violate the θ-Criterion if we assigned two terms the same θ-role.32 I will therefore keep to the plausible intuition that the adjoined element leaves the target unchanged. Consequently, if Merge is supposed to satisfy selectional requirements of a selector, as argued in Chomsky 2000, we can assume that the feature involved in this selection process remains unaffected by the fact that the selected term has an element adjoined to it.33

32 In his new account of obligatory control, Hornstein (1998) argues that θ-roles are features of verbs and derives the conclusion that movement to θ-positions is allowed.

33 Note that base-generated adjunction to arguments is required in the case of relative clauses if postcyclic insertion of relative clauses (Chomsky 1993) is no longer available, a requirement that further weakens Chomsky’s θ-theoretic argument. As an alternative, Chomsky (2000:n. 54) envisions an analysis of relative clauses in terms of multidimensional structures. As Sabel (1998) shows, dispensing with the ban on adjunction to arguments does not have any undesirable empirical consequences since its relevant part can be subsumed under a universal constraint defended by Grewendorf.
Step 2 is motivated by the EPP and by whatever necessitates overt satisfaction of this principle. Given that objects are overtly raised in German, this is simply in accordance with traditional proposals of Chomsky’s (1995). Crossing of the subject is possible because of equidistance, which still plays an important role in the most recent version of the Minimalist Program (Chomsky 2000, to appear).34 The single-level theory of movement becomes relevant after step 2 in that at this stage it prevents the wh-object in [Spec, vP] from moving directly to [Spec, CP]. Because the single-level theory requires that structure building be cyclic, [Spec, TP] has to be created first to provide a Case (EPP) position for the wh-subject to move to. Note that the possibility of the wh-subject (-cluster) moving directly to [Spec, CP] is ruled out also because the subject first needs to delete its Case feature. Thus, in step 3 the wh-subject (-cluster) overtly moves to [Spec, TP] for reasons of Case checking and EPP. Finally, the wh-cluster consisting of the visible wh-subject and the invisible wh-object moves to [Spec, CP] (where the wh-subject enters into a checking relation with the uninterpretable feature [Q] of C and deletes its uninterpretable feature [wh]).35

A theoretical implementation of this idea is possible in which the problem of cyclicity that may be associated with adjunction to a specifier does not arise. This implementation is based on Takano’s (1996) account of scrambling as Copy + Merge, which is similar in spirit to Bobaljik’s (1995) analysis of head movement (see footnote 21). This account proceeds from the idea that Copy is a costless operation that is freely applicable when and only when Merge applies (Move being treated as a special case of Merge). Accordingly, the formation of a wh-cluster can be analyzed as resulting from a combination of Copy and Merge in the following sense: two wh-elements (e.g., wer ‘who-NOM’ and wen ‘who-ACC’) are selected from the numeration; then Merge...
applies to *wen* and *wer*, adjoining the former to the latter and leaving a copy of the former as indicated in (24).

(24) wer

    wer  wen  wen (copy)

In the case of (22) the derivation proceeds by Merging *wen* (copy) with the verb and the resulting \(V^{\text{max}}\) with the light verb, yielding (25).

(25) vP

    \[
    \begin{array}{c}
    \text{\(V^{\text{max}}\)} \\
    \text{\(v\)} \\
    \text{\(\text{wen} \ \ V\)}
    \end{array}
    \]

If Merge applies to the two objects in (24) and (25), the result is (26).

(26) vP

    \[
    \begin{array}{c}
    \text{\(v\)} \\
    \text{\(\text{wen} \ \ V\)}
    \end{array}
    \]

Then the derivation proceeds as indicated in (23). Note that in this approach the operation Merge is asymmetric by nature and permits feature checking. If *wer* had merged to *wen* in (24), the derivation could not have continued as represented in (25) since the application of Copy is contingent on Merge in the sense that the element that is merged to another element leaves a copy. That is, in that case the copy would have been *wer*, and replacing *wen* by *wer* in (25) would not have led to a convergent derivation.\(^{36}\)

\(^{36}\) The fact that covert *wen* has to be merged and overt *wen* left as a copy is determined by the “weak” nature of the uninterpretable feature [Q] of the “attracting” wh-element. Given the assumptions so far, the analysis seems to wrongly predict that (i) is ungrammatical since the wh-object cannot adjoin to the wh-adjunct.

(i) Warum hat Maria *wen* geküsst?
    why  has Maria who-ACC kissed
    “Why has Maria kissed who?”

For a solution to this problem, see section 5.
The analysis of German multiple *wh*-questions in terms of *wh*-cluster formation receives independent support from the fact that it enables us to provide an account for the hitherto unexplained lack of superiority effects in German.\(^{37}\) Consider the examples in (27).

(27) a. Wen liebt wer?
   who-ACC loves who-NOM

b. Wen hat Maria warum geküsst?
   who-ACC has Maria why kissed

As for (27a), the (relevant part of the) derivation represented in (28) may proceed as follows. In step 1 the *wh*-subject (the goal) covertly adjoins to the visible *wh*-object (the probe) in the outer [Spec, vP], with the consequence that the uninterpretable feature [Q] of the probe *wen* and the uninterpretable “activating” feature [wh] of the goal *wer* are deleted. In step 2 the visible tail of the *wh*-subject chain so formed moves to [Spec, TP], crossing the *wh*-cluster headed by the *wh*-object *wen*. This movement poses no problems in a language with overt object shift (Chomsky, to appear). In step 3 the *wh*-cluster consisting of the visible *wh*-object *wen* and the invisible head *wer* of the *wh*-subject chain undergoes *wh*-movement to [Spec, CP], “attracted” by the uninterpretable feature [Q] of C, with the consequence that the uninterpretable feature [wh] of the *wh*-object *wen* is deleted. Although deactivated elements create an intervention effect, the visible *wh*-subject *wer* does not induce a violation of the Minimal Link Condition since it is a “trace” and traces do not create an intervention effect for movement (Chomsky, to appear).\(^{38}\)

At this point we may ask why the *wh*-object does not adjoin directly to the *wh*-subject in [Spec, vP]. If adjunction to elements in \(\theta\)-position is allowed, the derivation in (28) seems to violate cyclicity. However, immediate covert adjunction of the *wh*-object to the *wh*-subject in [Spec, vP] would have the effect that the feature [wh] of the *wh*-object deletes so that the *wh*-object could

\(^{37}\)Familiar accounts relate the lack of superiority effects to the property of scrambling: if the *wh*-object can be scrambled to an A-position in front of the *wh*-subject before undergoing A-movement to [Spec, CP], the Minimal Link Condition is not violated. However, an account along these lines cannot be correct. First, even though German permits short scrambling rather freely and does not exhibit (short) superiority effects, it does not permit (free) scrambling of *wh*-phrases (Fanselow 1990). A reviewer disagrees with this argument, pointing out that *wh*-phrases, while unable to undergo scrambling in general, may still be able to cross a further *wh*-phrase. Though I do not share this intuition, I wish to point out that this objection would only be relevant if this kind of crossing were an instance of A-movement. Otherwise, independent constraints (Müller and Sternefeld 1993, Grewendorf and Sabel 1994, Sabel 1998) would disallow subsequent *wh*-movement. Moreover, Pesetsky (2000) presents empirical evidence from German separation constructions to the effect that scrambling to the left of the subject does not feed *wh*-movement. Further evidence against the scrambling proposal comes from scrambling languages that, unlike German, also permit scrambling of *wh*-phrases but nevertheless show superiority effects. Such a language is Turkish (Kural 1992).

\(^{38}\)The fact that A-traces do not create intervention effects can be seen from raising constructions with (a) quirky Case moving to the matrix subject and (b) long-distance agreement of the matrix verb with an embedded nominative as in the Icelandic example (i).

(i) Mér, víðast t, [pessir hestar vera göðar],
    me-DAT seem-PL these horses-NOM.PL be-INF good
    ‘These horses seem to me to be good.’
    (Jonas 1997)

In (i) the trace of the raised quirky subject does not intervene between the probe (\(\phi\)-features of matrix verb) and the goal (the matching \(\phi\)-features of the embedded nominative).
no longer be activated again. Consequently, we should only be able to derive the structure in (23). But note that there are two requirements to be met at this stage of the derivation (i.e., after merging of the wh-subject): satisfaction of v’s EPP requirement by moving the object to the outer specifier position and satisfaction of the wh-subject’s requirement to delete its uninterpretable feature [Q] by forming a cluster. If cyclicity is evaluated at the level of strong phases, as argued in Chomsky, to appear, there is no reason why one of those requirements should be fulfilled prior to the other. So a real option exists here as to which requirement is met first: if the requirement of the wh-subject is met first, we get the derivation in (23); if v’s EPP requirement is satisfied first, we get the derivation in (28).\footnote{An analogous option with respect to T’s EPP requirement and the Q requirement of the wh-object in [Spec, vP] would exist if we proceeded from the alternative derivation outlined in footnote 35. I would like to thank a reviewer for helpful remarks on this problem. Takano’s (1996) theory would provide the following implementation: First, the covert wh-subject wer is merged with the wh-object, leaving an overt copy of the former. Next, the cluster so formed is merged with the verb. Finally, the overt copy of the wh-subject is merged with the projection of v.} English wh-phrases, unlike German ones, are unambiguously interrogative. Consequently, it may be assumed that they do not contain a wh-operator position and do not give rise to wh-cluster formation. If this is correct, one would expect that a derivation equivalent to (28) would not be available in English. We can thus conclude that the superiority effect in English arises because the cluster in [Spec, vP], consisting of wh-object and wh-subject,
cannot be formed, so that there is no way for the *wh-object to cross the *wh-subject in [Spec, TP] to reach [Spec, CP].

It should be clear that the derivation of (27b) may proceed along analogous lines, the only difference being that in this case the *wh-adjunct covertly adjoins to the *wh-object in the outer [Spec, vP] before the entire *wh-cluster undergoes *wh-movement to [Spec, CP].^40

Let us now turn to the observation (Büring and Hartmann 1994) that despite the absence of short superiority, there is a ‘long superiority effect’ in German with extraction from embedded finite clauses, as the data in (29) illustrate.

(29) a. *Wen, glaubt wer, dass Hans *t, gesehen hat?
   who-ACC believes who-NOM that Hans-NOM seen has
   b. Wer glaubt, dass Hans wen gesehen hat?
   who-NOM believes who-NOM that Hans-NOM who-ACC seen has
   (Büring and Hartmann 1994)

The Wh-Cluster Hypothesis, combined with the single-level theory of movement, offers a simple explanation for these facts and thus provides further evidence for the proposed analysis of multiple *wh-questions in German. Given that the two *wh-elements in (29a) have to form a cluster before moving to the matrix [Spec, CP], after passing through the embedded [Spec, CP] the extracted *wh-object *wen would have to move to a position where the matrix subject could covertly adjoin to it. However, there is no appropriate position for the embedded *wh-object to move to. After passing through the embedded [Spec, CP], the object is not allowed to adjoin (covertly or overtly) to the matrix *wh-subject since from the embedded [Spec, CP] it is prohibited from undergoing adjunction by the same constraint that applied in the above account of the Japanese example (14b) (Hoekstra and Bennis 1989, Müller and Sternefeld 1993, Grewendorf and Sabel 1994, Sabel 1998). It is standardly assumed in traditional analyses that unlike overt extraction, covert extraction of the embedded *wh-object does not need to use the embedded [Spec, CP] as an intermediate landing site (Aoun and Li 1993), so that the well-formedness of (29b) is correctly predicted.^41

^40 To account for the superiority effect in Bulgarian (Rudin 1988, Ivzvorski 1995, Richards 1997, Müller 1997), it remains to demonstrate that a derivation like (28) is not available in a language with overt multiple *wh-fronting. This demonstration rests on a constraint on adjunction movement according to which movement may not proceed via intermediate adjunction. From this constraint (argued for extensively in Grewendorf and Sabel 1999, Sabel 1998), we can conclude that checking the Case of a pronounced *wh-phrase via movement (EPP checking) has to precede its overt adjunction to another *wh-element, because from an adjoined position it could no longer undergo further movement for Case (EPP) checking. This implies that the only possible derivation of a *wh-cluster consisting of an overt *wh-subject and an overt *wh-object is the one illustrated in (10) (see also footnote 35). In other words, a *wh-object can overtly adjoin to the *wh-subject in [Spec, TP] after Case (EPP) checking in [Spec, vP] has taken place, but a *wh-subject can never overtly adjoin to a *wh-object since this would prevent it from checking its Case (EPP) in [Spec, TP]. Note that for the same reason, a derivation of the type that Takano’s (1996) theory provides is not available either: Merge of the overt *wh-subject with the overt *wh-object (followed by Merge of this cluster with V) would leave a covert copy of the former. If this happened, neither the adjoined overt *wh-subject nor its covert copy would be able to satisfy the requirements of T.

^41 The traditional idea that covert movement of embedded *wh-arguments in situ need not undergo successive-cyclic movement through [Spec, CP] has been treated in different ways in minimalist approaches. Either it has been claimed that these categories undergo no covert movement at all but are subject to unselective binding (Reinhart 1993, Chomsky 1995), or it has been assumed that covert movement differs in nature from overt movement and does not pass through
Notice that a sentence like (29a) with long extraction of the embedded wh-object is grammatical if the wh-phrase in situ constitutes the embedded rather than the matrix subject.

\[(30)\] Wen, glaubt Hans, dass wer t, gesehen hat?
who-ACC believes Hans-NOM that who-NOM seen has

The proposed analysis correctly predicts this since the account of (30) parallels that of (27a).

The exceptions to the Wh-Cluster Hypothesis allow us to predict as well that multiple wh-questions are ungrammatical in German in the presence of two (nonreferential) adjuncts. This prediction is also borne out, as shown in (31).

\[(31)\] a. *Weshalb hat er es wie repariert?
why has he it how fixed
b. *Wie hat er es weshalb repariert?
how has he it why fixed
(Haider 1996)

Further evidence for the proposed analysis is that it enables us to account for some intricate contrasts with multiple wh-questions in German observed in Müller 1995. Consider the contrast in (32), which calls to mind the similar contrast in Japanese illustrated in (14).

\[(32)\] a. Warum, glaubt Hans [CP t, dass der Student die Universität t, verlassen hat]
why thinks Hans-NOM that the student-NOM the university-ACC left has
‘Why does Hans think that the student has left the university?’
b. *Wer, glaubt t, [CP dass der Student die Universität warum verlassen
who-NOM thinks that the student-NOM the university-ACC why left
has

As is familiar from the situation we find in English, (32a) shows that a wh-adjunct can be extracted from an embedded declarative clause. However, as (32b) shows, if a wh-element is added in the matrix clause and fronted there, the wh-adjunct is not permitted to occur in situ in the embedded declarative clause. Again, the ungrammaticality of (32b) can be attributed to the need to form a wh-cluster. In traditional approaches it is assumed that for reasons of antecedent government, wh-
adjuncts—unlike wh-arguments—have to use the intermediate \([\text{Spec, CP}]\) when undergoing long covert extraction (for an alternative reason, see footnote 41). But then the same constraint that we invoked in the account of (29a) takes effect: from \([\text{Spec, CP}]\), the wh-adjunct may not undergo further adjunction because of the restrictions mentioned earlier. Recall that this constraint is not operative if the in-situ wh-element in (32b) is a wh-argument. Thus we correctly predicted the well-formedness of (29b).

Now compare the ill-formed (32b) with (33a), which shows a considerably higher degree of grammaticality.\(^{42}\)

(33) a. ?Wer, glaubt Hans \[\text{CP} t, dass Maria \text{t, warum verlassen hat}\]?
   who-NOM believes Hans-NOM that Maria-NOM why left has

   

b. *Wie, glaubt Hans \[\text{CP} t, dass Maria das Auto \text{t, warum repariert hat}\]?
   how thinks Hans-NOM that Maria-NOM the car-NOM why fixed has

Unlike in (32b), where overt wh-movement takes place in the matrix clause, in (33a) the overtly extracted wh-phrase originates from the embedded clause. On the basis of the single-level theory of movement and the proposed theory of multiple wh-questions in German, we can account for the grammaticality of (33a) as follows. Unlike in (32b), in (33a) the embedded wh-adjunct can adjoin covertly to the wh-object since the latter originates from the same clause. Thus, a wh-cluster consisting of the visible wh-object and the invisible wh-adjunct can be formed within the embedded clause. Then the entire wh-cluster undergoes successive-cyclic movement to the matrix \([\text{Spec, CP}]\). The prediction is again that the occurrence of a wh-adjunct in situ in the embedded clause is ungrammatical whenever the wh-phrase to be overtly extracted from the embedded clause is a wh-adjunct, since in this case the exceptions to wh-cluster formation prevent the wh-adjunct in situ from adjoin to the overtly extracted wh-adjunct. As (33b) shows, this prediction is borne out.

The proposed analysis of multiple wh-questions in German accounts for other intricate wh-extraction contrasts that, to the best of my knowledge, have not yet been considered in the literature.

(34) a. *Wer, glaubt t, nicht, dass Peter wen liebt?
   who-NOM believes not that Peter-NOM who-ACC loves

   

b. Wer, glaubt Peter nicht, dass t, wen liebt?
   who-NOM believes Peter-NOM not that who-ACC loves

(35) a. *Wer, glaubt t, nicht, dass Peter den Computer wie repariert?
   who-NOM believes not that Peter-NOM the computer-ACC how fixes

   

b. Wer, glaubt Peter nicht, dass t, den Computer wie repariert?
   who-NOM believes Peter-NOM not that the computer-ACC how fixes

\(^{42}\) In this respect I disagree with the judgment found in Müller 1995.
In (34a), where overt *wh-*extraction of the subject takes place from the matrix clause, the matrix negation disallows the presence of a *wh*-object in situ in the embedded clause. Interestingly, the example improves considerably when overt *wh-*extraction of the subject takes place from the embedded clause, as in (34b), rather than from the matrix clause. The situation in (35) is analogous to that in (34) except that the embedded *wh*-phrase in situ is an adjunct.

The ungrammaticality of (34a) and (35a) appears to confirm a generalization suggested by Beck (1996) according to which an intervening negation blocks LF movement. However, (34b) and (35b) seem to conflict with this generalization. Pursuing a similar idea, we might account for (34a) and (35a) in terms of the suggestion hinted at in Rizzi 1992 that the specifier position of a NegP may be filled by a negative operator only at LF; but again, the grammaticality of (34b) and (35b) raises a problem for such an account.

If an account of (34a) and (35a) along the lines of Beck 1996 or Rizzi 1992 is basically on the right track, the proposed analysis of multiple *wh*-questions enables us to bring the grammaticality of (34b) and (35b) into line with such an account. Let us express Beck’s generalization in the single-level framework by saying that adjunction of a *wh*-element to another *wh*-element is blocked by an intervening negation. We can then explain (34b) and (35b) as follows. Since overt *wh-*extraction in (34b) and (35b) takes place from the embedded clause, we can assume that the *wh*-element in situ has covertly formed a *wh*-cluster with the *wh*-argument to be extracted overtly without creating any problem for the (visible) tail of the *wh*-in-situ chain. As overt *wh*-extraction in German is not affected by an intervening negation, (34b) and (35b) are correctly predicted to be grammatical. In (34a), on the other hand, the embedded *wh*-element is unable to covertly form a cluster with the matrix *wh*-subject if covert crossing of an intervening negation is disallowed. In (35a) the embedded *wh*-adjunct is prevented from covertly forming a cluster with the matrix *wh*-subject, regardless of the presence of the matrix negation, by the same constraint that was effective in the Japanese example (14b).

In Muller 1996:188, a sentence like (34a) is judged as grammatical. I disagree with this judgment, as do all informants I have consulted. On my intuition, (34a) is possible only with an indefinite interpretation or an echo interpretation of the *wh*-element in situ. A reviewer considers example (i) to be perfect.

(i) Wer hatte nicht gewusst, dass der Peter wen eingeladen hatte?

*Who had not known that Peter had invited who?*

I agree that (i) sounds a little better than (34a), but this seems to me to be due to pragmatic factors since according to my intuition, (i) is to be understood as an echo question (or as an exclamation).

Recall that certain variants of German do not exhibit *that*-effects. For an account of the “Beck effect” in terms of intervention effects for movement of an invisible interrogative “Q-morpheme,” see Hagstrom 1998:chap. 3.

A reviewer questions whether my analysis of multiple *wh*-questions in German could deal with examples such as (i).

(i) Wen würde sie sagen dass er ti liebt wenn sie wer fragen würde?

*Who would she say that he loves if who asked her?*

The alleged problem with this example is that the *wh*-phrase *wen*, which undergoes overt extraction, originates from the complement clause while the *wh*-in-situ element is located in an adjunct clause that clearly modifies the matrix clause; as a result, there seems to be no point in the derivation where the two *wh*-elements could form a cluster. This example...
5 Wh-Adjuncts

In this section I want to briefly discuss multiple wh-questions in which a wh-adjunct has undergone overt extraction. Unlike their counterparts in English (Epstein 1993, Hornstein 1995), examples such as (36a–b) where the wh-phrase in situ is an argument are well formed in German.

(36) a. Weshalb hat wer den Professor kritisiert?
   why has who NOM the professor ACC criticized
b. Weshalb hat der Professor wen kritisiert?
   why has the professor NOM who ACC criticized

The examples in (36) pose the following problem for the proposed analysis of multiple wh-questions in German. If wh-elements have to form a cluster before moving to [Spec, CP], how is this requirement met in (36) if adjunction to wh-adjuncts is not possible? Since wh-adjuncts are not endowed with the uninterpretable feature [Q] in D (probably resulting from the lack of a D-head; see footnote 25), they do not provide an operator position and are unable to function as probes.

In order to solve this problem, let us first recall some important properties of multiple wh-questions with wh-adjuncts. Hornstein (1995) has pointed out that a functional reading is required for the obligatory pair-list reading of multiple wh-questions and that, therefore, one wh-element must contribute the individual domain for a functional reading of the other wh-element in a multiple wh-question. As Hornstein puts it: one of the wh-words has to act as the ‘‘generator’’ of such a domain.47 Unlike inherently D-linked wh-phrases, simple wh-words like who and what can single out an individual domain in multiple wh-questions only when occupying [Spec, CP].48 However, this option is excluded for wh-adjuncts since, as argued by Garrett (1996), a multiple wh-question with a wh-adjunct in [Spec, CP] is ungrammatical on the semantic grounds that a nonreferential wh-adjunct is not the sort of wh-phrase that can contribute a set to be distributed over. The same point is made by Hornstein (1995), who claims that nonreferential wh-adjuncts cannot function as ‘‘generators’’ in multiple wh-questions. If this argument is correct, we can

47 In Chierchia’s (1991) and Hornstein’s (1995) account of multiple wh-questions, this is implemented as the idea that functional readings of wh-questions involve implicit bound pronouns. Accordingly, a question such as (i) is analyzed as in (ii).

(i) Who bought what?
(ii) Who, [i bought [pro, what]]?

48 The idea is that a simple wh-element must be located in [Spec, CP] in order to be D-linked (Hornstein 1995:130). For a similar idea, see Garrett 1996:134.
conclude that in a well-formed multiple \( wh \)-question with simple \( wh \)-words in which an overt \( wh \)-adjunct occupies [Spec, CP], this \( wh \)-adjunct cannot be the only \( wh \)-element that is located in [Spec, CP]. It has to be accompanied by some other \( wh \)-element that is able to contribute the individual domain required for a functional reading of the question.

In order to apply this conclusion to the analysis of multiple \( wh \)-questions developed here, I assume, following an idea defended in Epstein 1993 and Hornstein 1995, among others, that \( wh \)-adjuncts in languages such as English and German can overtly occupy [Spec, CP] only as the result of base generation. But this by itself would not solve the problem, for syntactic as well as semantic reasons. First, the uninterpretable feature [\( wh \)] of the \( wh \)-subject is left undeleted; second, the \( wh \)-adjunct is the only \( wh \)-element in [Spec, CP]. To solve these problems, I will argue that in questions like (36) the \( wh \)-argument has the option that its D-head can move to C. The relevant part of a derivation for (36a) would then be (37). 49

(37)

\[
\begin{array}{c}
C' \\
\downarrow \\
C \\
\downarrow \\
D\text{wer} & \text{wer} & T'
\end{array}
\]

The ‘‘checking’’ mechanism that is at work here can be described along the lines of the mechanism that Chomsky (2000) assumes for the ‘‘checking’’ of expletive \( there \): we can assume that the uninterpretable features of \( D\text{wer} \) are deleted because of the uninterpretable feature [\( Q \)] of C (the probe). However, since \( D\text{wer} \) does not have the full set of features required for deletion of the uninterpretable feature [\( Q \)] of the probe (e.g., it does not contribute visibility, which is needed to satisfy the EPP-feature of C), and since the probe’s features delete in an ‘‘all or none’’ fashion (Chomsky 2000:124), the uninterpretable feature [\( Q \)] of the probe cannot delete. Consequently, Merge of the \( wh \)-adjunct \( weshalb \) provides the goal with which the probe C again enters the relation Agree, thereby deleting its uninterpretable feature [\( Q \)]. 50 Note that the option of moving the D-head of a \( wh \)-element to C still disallows the derivation of a multiple \( wh \)-question that contains only two \( wh \)-adjuncts. Furthermore, it exists only for the highest \( wh \)-element since other options are blocked by intervention effects, and it can only occur with a \( wh \)-element base-generated

49 For a similar idea, see Pesetsky 2000, where it is argued that in German all \( wh \)-in-situ undergo \( wh \)-feature movement (which Pesetsky takes to be covert head movement).

50 Concerning the locality restrictions to be assumed for base generation of \( wh \)-adjuncts in operator positions, examples such as (i) suggest that base generation of \( wh \)-adjuncts is subject to the same locality constraints that are assumed for adjunct extraction.

(i) *Wie hat wer gesagt warum Maria das Auto t i repariert hat? how has who-NOM said why Maria-NOM the car-ACC fixed has

In other words, if a \( wh \)-adjunct base-generated in the matrix [Spec, CP] of a complex sentence is to relate to the embedded clause, a ‘‘trace’’ or copy of the adjunct operator (in the sense of Aoun and Li 1993) is required in [Spec, CP] of the embedded clause.
in [Spec, CP] since movement of a (lower) wh-element to [Spec, CP] is also blocked by intervention effects. Finally, it is only possible in languages whose wh-arguments have a D-head endowed with an uninterpretable feature [Q].

Several pieces of independent evidence support the solution illustrated in (37). First, Wiltschko (1997) has pointed out that German yields a superiority effect in multiple questions with two wh-phrases in which the individual domain of each wh-phrase is the same set containing two individuals (see also Pesetsky 2000).

(38) Context: 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-NOM who-ACC called has
      ‘Do you know who called whom?’
   b. *Weißt du, wen wer angerufen hat?

(39) Context: I have heard that Peter and Mary had an affair. Can you tell me:
   a. Wer hat wen verführt?
      who-NOM has who-ACC seduced
      ‘Who seduced whom?’
   b. *Wen hat wer verführt?

It is obvious that German wh-words can have a reading comparable to that of the Latin wh-element _uter_ ‘who of two’. It follows from the analysis offered in the previous section that with this reading, German wh-words are unable to undergo cluster formation. In other words, we can assume that these wh-words do not have D-heads endowed with an uninterpretable feature [Q]. If this is the case, we can conclude that wh-arguments with this reading should be unable to undergo D-movement to C. Consequently, we can predict that an overt wh-adjunct in [Spec, CP] should not be possible with such a wh-argument in situ. This prediction is borne out, as shown by (40) (with the contexts indicated in (38) and (39), respectively).

(40) a. *Weshalb hat wer wen angerufen?
   why has who-NOM who-ACC called
   b. *Weshalb hat wer wen verführt?
   why has who-NOM who-ACC seduced

Other evidence for my analysis of (36) concerns the assumption that C enters an Agree relation with both the wh-adjunct and the wh-argument in situ. There is a strong semantic argument for this assumption. Recall that only elements that range over individuals can function as generators

---

51 The proposal represented in (37) does not imply that languages with overt multiple wh-fronting should allow a wh-adjunct to precede a fronted wh-argument. We can assume that the option of moving the D-head of a wh-element to C is limited to languages that show verb-second effects. We can then predict that those variants of Yiddish that permit overt multiple wh-fronting also allow a fronted wh-adjunct to precede a fronted wh-argument. According to my informants, this prediction seems to be borne out.

(i) Ikh veys nisht far vos wer hot ongeklungen.
   I know not why who has called
in multiple *wh*-questions. In other words, a nonreferential *wh*-adjunct is unable to function as a generator, that is, to contribute the domain of individuals required for a pair-list reading. However, since being located in the C-system is a necessary condition for simple *wh*-phrases to act as generators and since in my analysis there is another *wh*-element located in the C-system (namely, the invisible *wh*-subject), we can assume that the individual domain is provided by the *wh*-subject, assigning the functional reading to the *wh*-adjunct. The assumption that the *wh*-adjunct does not function as the generator corresponds exactly to what is meant by (36a). The interpretation of (36a) does not imply that there are several reasons, and the question is not which person has criticized the professor for which of these reasons. Rather, the sentence is interpreted as meaning roughly the following: ‘I know that a couple of people have criticized the professor; tell me why each person has criticized him’. Thus, according to the natural interpretation of (36a), the set of critics is given and a reason for each critical statement is being requested. In other words, the proper interpretation of (36a) corresponds to the reading in which the *wh*-subject functions as the generator.

6 Summary

In this article I have shown that overt multiple *wh*-fronting in languages like Bulgarian and Romanian does not consist of separately moving individual *wh*-elements to [Spec, CP]; instead, it consists of first forming a single *wh*-cluster and then moving it to [Spec, CP]. The formation of *wh*-clusters is motivated by the assumption that *wh*-elements can act as landing sites for *wh*-movement owing to morphological properties of *wh*-words. I have further argued that languages such as Japanese exhibit covert instances of this process of *wh*-cluster formation, demonstrating that intricate constraints on multiple *wh*-questions in these languages such as the so-called additional-*wh* effect can be derived from this analysis. Furthermore, I have tried to establish the claim that despite appearances, multiple *wh*-questions in German also involve the formation of *wh*-clusters, which in this language consist of one visible and one or more invisible *wh*-elements; in this respect, then, German turns out to hold an intermediate typological position between languages like Bulgarian and languages like Japanese. Finally, I have argued that the *wh*-cluster analysis also accounts for German multiple *wh*-questions with an overt *wh*-adjunct in [Spec, CP].

References


*Institut für deutsche Sprache und Literatur II*
*Universität Frankfurt/Main*
*Gräfstr. 76*
*60054 Frankfurt/Main*
*Germany*

*grewendorf@lingua.uni-frankfurt.de*