CONSTRAINTS ON DISPLACEMENT
A PHASE-BASED APPROACH

Gereon Müller
Universität Leipzig

www.uni-leipzig.de/~muellerg
gereon.mueller@uni-leipzig.de

Contents

Introduction

What This Book Is About

Locality constraints on displacement as they have been proposed over the last decades (since Chomsky (1964) and Ross (1967)) have been highly successful from an empirical point of view. Constraints like the A-over-A Principle, the Subjacency Condition, the (Generalized) Minimal Link Condition ((G)MLC), the Condition on Extraction Domain (CED), or Relativized Minimality impose significant restrictions on movement in natural languages. However, recent developments in the minimalist program (see Chomsky (1995; 2001b; 2008)) have led to the postulation of strict meta-requirements on syntactic constraints, and closer inspection reveals that even the most widely adopted constraints on displacement are at variance with these requirements. Throughout this monograph, I assume that 'good' syntactic constraints must conform to the meta-requirements in (1).[1]

(1) a. Constraints should be as simple and general as possible.
    b. Constraints should not be complex.
    c. Constraints are of type (i) or (ii).
        (i) principles of efficient computation
        (ii) interface conditions
    d. Constraints do not require massive search space.
    e. Constraints do not give rise to redundancies.

The requirements in (1) do not only exclude transderivational constraints like the Fewest Steps condition or the Shortest Paths condition, or constraints like the A-over-A Principle or the Subjacency Condition; I will argue that they also make it impossible to adopt constraints like the (G)MLC, the CED, and Relativized Minimality. A look at the more recent minimalist literature suggests that this conclusion might be uncontroversial for the CED; but I will argue that the conclusion is just as valid for the

---

[1] I will extensively motivate these meta-requirements in the first two chapters of this study. At first sight, it might seem that some of the requirements may overlap; however, as will be shown below, they are all separate.

(G)MLC and Relativized Minimality, where it would at first blush seem to be more controversial (since constraints of this type are regularly employed in current minimalist approaches to displacement). Against this background, the task for syntactic theory is to derive the effects that can be attributed to constraints on displacement like the (G)MLC and the CED, from independently motivated constraints that adhere to the demands imposed by (1).

A crucial role will be played by the Phase Impenetrability Condition (PIC) which is introduced in Chomsky (2000; 2001b) as a constraint that contributes to efficient computation, and which meets all the requirements in (1); see (2).

(2) *Phase Impenetrability Condition* (PIC):
    The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

The PIC requires all instances of movement (that are not extremely local) to take place successive-cyclically, via intermediate phase edges (simplifying a bit, the intervening phase heads' specifiers). The status of a phase is originally motivated only for the 'propositional' categories vP and CP in Chomsky's work; but various extensions have been proposed over the last decade, including, e.g., TP, DP, and PP. I will assume throughout that *all* XPs are phases. If so, the PIC demands that successive-cyclic movement applies in a very local manner, from one phrase edge to the next phrase edge higher up, yielding a representation of displacement dependencies that is comparable to the SLASH feature percolation approach developed in Gazdar (1981). As a result, the PIC's contribution to efficient computation and minimization of search space (see (1-cd)) is further enhanced.

The PIC states that movement *must* take place via phase edges. Assuming that all movement operations must be forced by appropriate structure-building ('edge') features (because of a general Last Resort constraint that prohibits free application of displacement, as it could be derived under Chomsky's (1981) concept of Move α), it then needs to be ensured that intermediate movement steps to phase edges *can* be carried out. Chomsky has suggested that intermediate movement steps can be brought about by edge features that are inserted on phase heads in the course of the derivation if certain conditions are met. In Chomsky (2000, 109) and Chomsky (2001b, 34) (also cf. Chomsky (2008, 149)), the condition for edge feature insertion is stated as follows (I refer to this condition as the Edge Feature Condition).

(3) *Edge Feature Condition* (EFC):
    The head X of phase XP may be assigned an edge feature *after* the phase XP is otherwise complete, *but only if that has an effect on outcome*.

I think there is reason to take issue with two aspects of the EFC in (3), and I have highlighted these two aspects by italicization: On the one hand, assuming that all

syntactic operations are triggered by phase heads, the null hypothesis would seem to be that a phase head is inert, and inaccessible for further modification, once the phase is complete, i.e., once it has "exhaust[ed] the lexical subarray from which it is derived", in Chomsky's (2000) words. In view of this, I essentially suggest to replace *after* with *before*: The phase head must still be active, and thus be able to induce at least one further operation, for edge feature insertion to apply.

On the other hand, the "effect on outcome" requirement in (3) is clearly too vague as it stands; in addition, it suffers from imposing look-ahead capacity on (what should be local) decisions in derivations if taken literally. Following Heck & Müller (2003), I propose that a simple, local way of capturing the gist of the requirement without running into the look-ahead problem centers around the notion of a balanced phase: Basically, "to have an effect on outcome" is to be understood as "to be the only way to lead to a balanced phase", where a phase counts as balanced if, for every structure-building feature in the numeration, there is a matching feature that is potentially available – and a feature is potentially available if it is either part of the workspace, or at the left edge of the current phase. An informal version of the revised Edge Feature Condition would then look as in (4) (see (11) in chapter 4 for a more precise formulation).

(4) *Edge Feature Condition* (EFC, new version):
     The head X of phase XP may be assigned an edge feature *before* the phase XP is otherwise complete, *but only if there is no other way to produce a balanced phase*.

With these modifications of the Edge Feature Condition, two consequences arise that I take to be the two central results of the present monograph: Both (G)MLC effects and CED effects can now be shown to follow from the PIC. In a nutshell, (G)MLC effects follow because the higher one of two items competing for movement to the domain of a movement-inducing head (i.e., the item that is merged later) ensures phase balance without edge feature insertion, which is therefore blocked; and subsequent movement of the lower item violates the PIC. CED effects follow because edge feature insertion cannot take place for an item that is included in a last-merged specifier of a phase head, with the phase head qualifying as inert at this point; subsequent movement of such an item in a last-merged specifier then also violates the PIC.

This approach not only reduces (G)MLC effects and CED effects to independently motivated principles (viz., the PIC and the EFC); it can also be shown to be empirically superior in various domains since it makes correct predictions that the standard approaches in terms of (versions of) the (G)MLC and the CED have nothing to say about. For instance, the new account of (G)MLC effects also covers kinds of intervention effects in German and English where the intervener does not c-command the item whose movement it blocks. Similarly, the new approach to CED effects emerges as empirically superior since it only derives last-merged specifiers as islands for extrac-

tion; non-last merged specifiers may be transparent. This will be shown to yield many new, and empirically correct, consequences, not least of all a hitherto unnoticed effect that I refer to as *melting*, which can be shown to underlie certain cases of transparent inner specifiers in the presence of local scrambling to an outer specifier in Czech and German.

Rizzi's (1990) Relativized Minimality constraint and the (G)MLC are conceptually related, and there is a certain amount of empirical overlap. Still, there are effects that can be covered by Relativized Minimality, but not straightforwardly (as I will show) by the (G)MLC; most notably, this holds for *wh*-island (and related operator island) effects. Here, too, it turns out that the PIC/EFC-based approach to constraints on displacement developed in this monograph can make interesting, and empirically correct, predictions.

In particular, I argue that a consequence of the present approach with respect to the positions of intermediate movement steps makes a new account of operator island (*wh*-island and topic island) effects possible that does not rely on the idea that these islands instantiate some kind of intervention effect: The analysis of displacement in terms of (2) (PIC) and (4) (EFC) gives rise to what I call the *Intermediate Step Corollary*: Intermediate steps of successive-cyclic movement end up in *non*-outermost specifiers (if more than one specifier is available). Given this assumption, it suffices to postulate that certain types of goal features on categories (like *wh*-features) automatically trigger checking once they enter a suitable configuration (as they do with intermediate movement steps); this is an instantiation of a more general concept of feature *maraudage*. As a consequence, operator island effects are derived (as is their variable nature).

Throughout, I adopt a strictly derivational, bottom-up, multiple specifier approach to phrase structure generation in which structure-building by (external) Merge and structure-building by movement (internal Merge) are interleaved (as in Chomsky (2000; 2001b), and much subsequent work), in accordance with the Strict Cycle Condition which demands (among other things) that syntactic operations cannot exclusively target embedded phrases. The clausal spine thus generated is assumed to uniformly consist of CP, TP, vP, and VP. I assume that all syntactic operations are induced by designated features, of which there are basically two types: Structure-building features for external Merge and movement are accompanied by bullets ($[\bullet F \bullet]$), and probe features that trigger Agree operations are embraced by asterisks ($[*F*]$).[2] Also, I adopt a feature checking (rather than a feature valuation) approach, for reasons that will be laid out in chapter 3 below.

---

[2] Chapter 7 will introduce a third type of operation-inducing feature, viz., $[\dagger]$, which triggers deletion.

An indispensable part of the analyses of (G)MLC, CED, and Relativized Minimality effects presented here is the assumption that all syntactic operations apply in a fixed order (this assumption is derivable from the properties and the organization of operation-inducing features on phase heads and a few very general principles but without resort to any extrinsic rule ordering): The *timing* of syntactic operations emerges as crucial. Hence, to the extent that the approach can be viewed as successful, it provides a strong argument for a strictly derivational organization of syntax, and against (even mildly) representational approaches, including those that incorporate Chomsky's (2008) hypothesis that constraint evaluation only takes place at the phase level.

What This Book Is Not About

Given the goal of deriving the effects of three widely adopted constraints on displacement from more elementary assumptions based on the PIC and the (revised) EFC, the present investigation can arguably be viewed as fairly ambitious. However, I would like to emphasize that it is nevertheless quite narrow in scope. On the empirical side, no attempt is made to cover some phenomenon (like, for instance, *wh*-movement in German, or the cross-linguistic distribution of scrambling-related island effects) comprehensively. On the theoretical side, I leave open the question of whether (and, if so, how) various other well-established locality constraints on displacement can also be derived from more elementary assumptions; thus, I will have nothing to say about, e.g., the Coordinate Structure Constraint (see Ross (1967)) and the Left Branch Condition (see Ross (1967)), which also fail the set of meta-constraints in (1).

Furthermore, there are several phenomena that I more or less ignore in this monograph (both as empirical data, and in their relevance for theory construction) even though they are undoubtedly of great importance in any approach that strives to derive locality constraints on displacement. Among these are *resumptive pronouns* (see Pesetsky (1998), Hornstein (2001), Boeckx (2003), Salzmann (2006), and the contributions in Rouveret (2011), among others); *connectedness effects* with multiple movement of the same type (see Kayne (1984) on the Connectedness condition, or Richards (1998) on the Principle of Minimal Compliance); *parasitic gaps* (as well as *Across-the-board movement* (ATB) in coordinate structures) (see Chomsky (1982b), Frampton (1990), Nissenbaum (2000a;b), and Lutz (2001), among many others – but see also Assmann (2010) for an approach to parasitic gaps based on the approach developed in chapter 4 below); and (with the exception of a few scattered remarks) *pied piping* (see Heck (2004; 2009) and Cable (2007) for two recent approaches). Moreover, I will have very little to say about *argument/adjunct* (or related) *asymmetries* as they arise in extraction from islands (see, e.g., Chomsky (1986a), Cinque (1990), Manzini (1992), and Kitahara (1997)); but what little I have to say in chapter 5 may

perhaps eventually lay the basis for a fresh approach to the phenomenon. I omit a detailed discussion of these and other issues only for reasons of space and coherence, and not because I would think that they are not highly relevant to the main issues under investigation here. My hope is that the approach developed in this monograph can be conservatively extended to all these phenomena; but whether this is really the case (as it is with parasitic gaps) remains to be seen.

Overview

I proceed as follows.

Chapter 1 serves as an introduction into the area of inquiry. It sets the stage by showing how the locality constraints on displacement that are the main subject of the present study (viz., the (G)MLC, the CED, and Relativized Minimality) developed out of their predecessors, beginning with Chomsky's (1964) A-over-A Principle and Ross's (1967) island constraints. Some (but certainly not all) of the material in this chapter is textbook knowledge (as it can also be found in, e.g., Riemsdijk & Williams (1986)). However, what is of particular importance here, and what distinguishes this chapter from related material in other studies, is the focus of presentation, which relies on the notion of a "good" constraint, as it can be formulated on the basis of meta-requirements on constraints like those in (1).

In chapter 2, I argue that a successful derivation of both the (G)MLC (as well as the related Relativized Minimality constraint) and the CED in terms of more elementary minimalist principles is outstanding. For the (G)MLC, this amounts to showing that the constraint, though widely adopted nowadays, is incompatible with (1) and basic minimalist tenets. For the CED, it amounts to showing that all existing minimalist approaches that attempt to derive its effects are problematic. Given that neither principles of parsing nor semantic restrictions can systematically account for (G)MLC (plus Relativized Minimality) and CED effects, a new minimalist approach is called for.

Chapter 3 and chapter 4 form the core of the present monograph. In chapter 3, I illustrate how (G)MLC effects can be derived from the PIC if the notion of "having an effect on outcome" in the Edge Feature Condition is replaced with the notion of "being the only way to produce a balanced phase" – a move which I show to be independently motivated.

In chapter 4, it is argued that another independently motivated change in the Edge Feature Condition (viz., that of replacing *after* with *before*) predicts standard CED effects. The analysis is then extended so as to cover freezing effects; and it is shown that it correctly (and even without this extension) predicts the existence of a reverse phenomenon that I call melting effect.

In an appendix to chapter 4, I suggest that this approach to CED effects may shed new light on the occurrence of *movement-related pieces of morphology*, i.e.,

what is often referred to as morphological reflexes of successive-cyclic movement. I show that this phenomenon might perhaps best be viewed as an instance of *pseudo-melting*: An expletive that accompanies an item which would normally qualify as a last-merged item (hence, as opaque for extraction) turns this item into a transparent non-last-merged item. This way, a rationale can be given for the very existence of movement-related morphology.

Chapter 5 addresses the variable nature of operator islands (more specifically, *wh*-islands and topic islands). I show how the Intermediate Step Corollary of chapter 4 makes a maraudage approach to operator islands available, and I suggest various possible extensions that the analysis permits (concerning, inter alia, argument/adjunct asymmetries, movement type asymmetries, and clause-type asymmetries).

Finally, the last two chapters extend the approach developed in the preceding chapters in two separate (but conceptually similar) ways, for two phenomena where CED-type effects that one might initially expect fail to occur. First, in chapter 6 I argue that this approach offers a simple account of a peculiar asymmetry with movement from verb-second clauses in German: Verb-second clauses are islands for movement into verb-final clauses, but there is a lack of CED effects with movement into verb-second clauses again. Based on a reprojection approach to head movement where an in-situ V itself bears the feature triggering verb-second movement (which keeps the V head active at the relevant step of the derivation), this asymmetry will be analyzed in analogy to pseudo-melting effects. Second, chapter 7 shows that standard cases of island repair by ellipsis in sluicing constructions follow without further ado, assuming that deletion is successive-cyclic and brought about by a designated feature that, again, keeps the phase head active at the decisive step of the derivation.

Acknowledgments

very helpful, comments on larger parts of the material on which this book is based; and Anke Assmann, Petr Biskup, Doreen Georgi, Fabian Heck, Stefan Keine, and Marc Richards, who did the same, and who, together with several others, have managed to create a highly stimulating environment for pursuing syntactic theory at the University of Leipzig.

This book is dedicated to my father, Hans Müller, whose crumpled and bruised (though carefully annotated) copies of *Syntactic Structures* and *Aspekte der Syntaxtheorie* sent me off on a fascinating *ezda v ostrov ljubvi* when he gave them to me nearly thirty years ago.

Chapter 1

Locality Constraints

1. Introduction

The goal of the first chapter of this monograph is to set the stage for the remaining chapters. I begin by introducing a taxonomy of constraints in general, and characterizing properties of a "good" constraint. Then I address a number of locality constraints that have been proposed in the last couple of decades; broadly speaking, a locality constraint is a constraint that restricts syntactic operations to certain domains. (I do all of this in a somewhat anachronistic way, by introducing all constraints on the basis of a current derivational model of syntax that conforms to minimalist assumptions.) Research based on these constraints has culminated in the postulation of two highly general, widely accepted constraints that are at the heart of the present study: the (Generalized) Minimal Link Condition (G)MLC and the Condition on Extraction Domain (CED). To end the chapter, I address the issue as to whether these two locality constraints are compatible with elementary minimalist assumptions. A closer look at the relevant literature reveals that the (G)MLC is typically assumed to be compatible with minimalist tenets whereas the CED is not. In line with this, the (G)MLC is typically adopted in current minimalist research, whereas numerous attempts have been made to derive the CED.

2. Constraint Types

2.1 A Taxonomy of Constraint Types

Based on varying degrees of complexity, several types of constraints can be distinguished. The taxonomy in (1) is based on Müller & Sternefeld (2001). According to (1), five core types of syntactic constrains can be recognized (there may well be more, but the ones given here will suffice for present purposes).

(1) *Constraint types*:

    a.   A local *derivational* constraint applies to syntactic (Merge or Move) operations.

b. A local *representational* constraint applies to an output representation.

c. A *global* constraint applies to a whole derivation; it correlates non-adjacent steps in the derivation.

d. A *translocal* constraint applies to sets of output representations; it picks out an optimal output representation among competing output representations.

e. A *transderivational* constraint applies to sets of derivations; it picks out an optimal derivation among competing derivations.

2.2   Local Derivational Constraints

First, local constraints can be separated from non-local constraints. In derivational grammars, there are two kinds of local constraints. Local *derivational* constraints restrict the application of elementary syntactic operations (like Merge and Move); a classical local derivational constraints is the Subjacency Condition that blocks movement operations which cross two bounding nodes (DP and TP, or DP and CP, as options of parametrization); see Chomsky (1977; 1986a) and Rizzi (1982), among many others. The Subjacency Condition is defined in (2).

(2) *Subjacency Condition* (Chomsky (1977)):
     In a structure α ... $[_\beta$ ... $[_\gamma$ ... δ ... $]$ ... $]$ ..., movement of δ to α cannot apply if β and γ are bounding nodes.

The Subjacency Condition has been held responsible for the illformedness of sentences like (3) in English (with DP and TP as bounding nodes), and for the contrast between the well-formed example (4-a) and the ill-formed example (4-b) in Italian (with DP and CP as bounding nodes); bounding nodes crossed by movement operations are marked by underlining.

(3) *How$_1$ does $[_{\underline{TP_3}}$ she know $[_{CP}$ $[_{DP_2}$ which car $]$ $[_{\underline{TP_4}}$ Mary fixed t$_2$ t$_1$ $]]]$?

(4) a. Tuo fratello $[_{CP_3}$ $[_{PP_1}$ a cui $]$ mi   domando $[_{\underline{CP_4}}$ $[_{DP_2}$ che   storie $]$
          your brother   to whom myself I ask   which stories
          abbiano   raccontato t$_2$ t$_1$ $]]$ era molto preoccupato
          they have told   was very worried

     b. *Francesca $[_{CP_3}$ $[_{DP_1}$ che $]$ non immagino $[_{\underline{CP_4}}$ $[_{DP_2}$ quanta   gente $]$ t$_2$
          Francesca   who not I imagine   how many people
          sappia $[_{\underline{CP_5}}$ $[_{PP_6}$ dove $]$ hanno   mandato t$_1$ t$_6$ $]]]$ ...
          know   where they have sent

In (3) and in (4-ab), *wh*-movement crosses a *wh*-phrase in a SpecC position (a *wh*-island effect, see below); but whereas this leads to a violation of the Subjacency Condition in (3) (with two TP crossed) and (4-b) (with two CPs crossed), it does not in (4-a) (where only one CP is crossed, and crossing of two TPs is irrelevant, by assump-

tion).[1]

2.3   Local Representational Constraints

In contrast to derivational constraints like the Subjacency Condition, local *representa-tional* constraints restrict possible output representations (i.e., results of applications of derivational operations). These latter constraints are also often referred to as *filters* (see Chomsky & Lasnik (1977)). The Empty Category Principle (ECP; see Chomsky (1981; 1986a), Lasnik & Saito (1984; 1992)) is a constraint of this type. It demands that traces left by movement have to be marked $[+\gamma]$; a trace can be marked $[+\gamma]$ only if it is in a proper government relation (with either its antecedent or a lexical head: proper government arises under antecedent-government or under lexical government). A (still somewhat simplified) version of the ECP along these lines is given in (5).

(5) *Empty Category Principle* (ECP):
     Every trace t is marked $[+\gamma]$.
     A trace t is marked $[+\gamma]$ iff it is properly governed.
     A trace t is properly governed iff (a) or (b) holds.

     a.   t is antecedent-governed.

---

[1] That said, it is not really clear that a parametrization of bounding nodes as it is proposed in Rizzi (1982) (more generally, leaving options in, by assumption, universal constraints that individual languages need to fix in one way or another) is the correct approach. First, the Italian examples above involve relativization; however, Italian *wh*-movement constructions that closely resemble constructions like those in (3) that are typically chosen to illustrate Subjacency effects in English are also fairly ill formed. The pertinent examples can already be found in Rizzi's (1982) original work; see (i). Second, some kinds of English examples of the general type in (3) (like, e.g., (ii-ab)) are often not judged to be that ill formed after all; see Grimshaw (1986) and Chomsky (1986a).

(i)   a.   *$[_{DP_1}$ Chi] ti   domandi $[_{CP_3}$ $[_{DP_2}$ chi] t_2 ha incontrato t_1] ?
                who yourself you ask   who   has met

      b.   ??$[_{DP_1}$ Chi] non sai $[_{CP_3}$ $[_{DP_2}$ che cosa] t_1 ha fatto t_2] ?
                who not you know   what   has done

(ii) a.   ?*$[_{DP_1}$ Which book] did the students forget $[_{CP_3}$ who_2 t_2 wrote t_1] ?
      b.   ?$[_{DP_1}$ Which car] did John tell you $[_{CP_3}$ how_2 to fix t_1 t_2] ?

For reasons like these, the concept of parametrization of bounding nodes is often considered dubious nowadays, quite in contrast to the significant role it has played in the development of *Government and Binding* theory; see Chomsky (1981), and compare the following citation from Chomsky (1982a, 45):

> Take Luigi Rizzi's paper on *wh*-movement, there is a typical case. You consider the *wh*-island constraint and you look at Italian and it just refutes it, period. It seemed to me that it refuted it, and I was confused and surprised, but it just looked like a correct refutation. You could stop there, and most people would have, but he did not. He showed that, if you modify a few things, not only did Italian not refute the constraint and the proposed explanation for it in terms of subjacency, but in fact it supported it.

A trace is antecedent-governed iff there is no barrier separating it from its immediate chain antecedent.

b.   t is lexically governed.

A trace is lexically governed iff it is in the m-command domain of a lexical category, and no barrier separates it from this lexical category.

Since this constraint is supposed to interact with abstract operations like post-syntactic complementizer deletion and post-syntactic deletion of intermediate traces, it is, in its standard formulation, assumed to apply at the level of logical form (LF); assuming the ECP to apply to every syntactic operation would give rise to wrong predictions. Thus, the ECP rules out complementizer-trace configurations as in (6-a) but does not block minimally different examples like (6-b) if it is assumed that (i) C cannot properly (lexically) govern a subject trace, and (ii) a phonologically zero C does not block antecedent government by an intermediate trace in SpecC whereas an overt C does.

(6)   a.   *Who$_1$ do you think [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$ that] [$_{TP}$ t$_1$ left]]] ?
      b.   Who$_1$ do you think [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$] [$_{TP}$ t$_1$ left]]] ?

Importantly, the complementizer-trace effect disappears if the complementizer intervenes at a later stage of the derivation, and the original subject trace can be antecedent-governed by an intermediate trace, as in (7-a); this latter intermediate trace cannot be marked [+γ], but (unlike the original trace) it can be deleted before the ECP is checked at LF. Similarly, assuming that adjunct traces can only be [γ]-marked at LF (whereas argument traces are [γ]-marked as soon as they arise in the derivation), the non-occurrence of a complementizer-trace effect in (7-b) can be accounted for by postulating complementizer deletion before LF.

(7)   a.   Who$_1$ do you believe [$_{CP}$ that Mary thinks [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$] [$_{TP}$ t$_1$ left]]]] ?
      b.   Why$_1$ do you think [$_{CP}$ t$'_1$ [$_{C'}$ [$_C$ that] [$_{TP}$ Mary left t$_1$]]] ?

Thus, assuming the ECP to be a derivational constraint would give rise to wrong predictions ((7-ab) should be ill formed), and only a representational view (as an LF filter) is compatible with the evidence.

The principles A, B, and C of the binding theory that regulate the distribution of *anaphors* (i.e., reflexive pronouns and reciprocal pronouns), *pronouns* (i.e., non-anaphoric pronouns), and *R-expressions* (i.e., non-pronominal DPs) are also assumed to be representational constraints in Chomsky (1981) and much subsequent work (but see footnote 4 for a qualification).[2]

---

[2] At this point, one may ask whether the Subjacency Condition mentioned in the previous section could also be conceived of as a representational constraint. Chomsky (1981, 302f) (based on Howard Lasnik (p.c.))

2.4   Global Constraints

Next, *global* constraints apply neither to single steps of derivations, nor to output representations thus generated; rather, they apply to whole derivations of sentences where they correlate non-adjacent steps. Global constraints are introduced in Lakoff (1971); Lakoff suggests that order preservation effects observable with quantifier lowering (where the relative scope of quantified items corresponds to their surface order) can be accounted for by a "global rule" according to which the relative order of quantified items that is present at the beginning of the derivation must be maintained throughout later levels of the derivaton, even under quantifier movement.[3]

Another well-known global constraint is Chomsky's (1981) Projection Principle which states that lexically determined selection relations between two items must be preserved throughout the derivation in the sense that they can be locally checked at every level of representation (thereby forcing the presence of traces of movement, among other things). Thus, to find out whether the Projection Principle is violated, it does not suffice to simply look at a given output representation, or at a step in the derivation; rather, one has to compare different output representations (e.g., one output representation in which movement has applied must be compared with an earlier output presentation in which the moved item is still in situ).[4]

---

argues that this is not the case. One argument is based on sentences like (i) (also see Sternefeld (1991a)).

(i)   Who$_1$ did [$_{TP}$ you believe [$_{CP}$ t$_1'$ C [$_{TP}$ Bill to have seen t$_1$]]] ?

Assuming that AcI constructions like (i) must involve deletion of the CP layer (so-called 'S-bar deletion') including the intermediate trace in the course of the derivation in order to make assignment of case from the matrix verb (*believe*) to the embedded subject (*Bill*) possible ('exceptional case marking'), the output representation (after CP deletion) will violate the Subjacency Condition (with t$_1$ separated from *who*$_1$ by two TP nodes). However, CP deletion is unproblematic if the Subjacency Condition is a derivational constraint, with movement preceding CP deletion.

[3] Similar proposals for relative quantifier scope in terms of a global constraint demanding order preservation can be found in other (including more recent) work; see, e.g., Kroch (1974), Reinhart (1983), Huang (1982; 1995), Lasnik & Saito (1992), Müller (2001), Williams (2003), Bobaljik & Wurmbrand (2008), and Wurmbrand (2008), among others. Most of the approaches just mentioned (but not all; see, e.g., Williams (2003)) postulate that the relevant global order preservation constraint operates on the basis of the standard T-model, where LF follows S-structure and the constraint in question is one on quantifier *raising*; this is in contrast to the generative semantics model originally envisaged by Lakoff (1971), where surface structure is derived from a prior (semantically disambiguated) D-structure, and the constraint accordingly is one on quantifier *lowering*. This difference, though fundamental as such, does not affect the claim just made: In both cases, we are dealing with a global constraint.

[4] Interestingly, based on the wellformedness of sentences like (i-ab) (from Italian and German, respectively, where the anaphor has undergone movement out of the c-command domain of its antecedent), (i-c) (where the anaphor finds a binder only after movement of its antecedent), and (i-d) (where the anaphor can find a new binder after an item that contains it moves; see Barss (1986)), it has been argued that principle A of the

2.5   Transderivational Constraints

The remaining two constraint types to be discussed here are *transderivational* and (what I will call, following Müller & Sternefeld (2001)) *translocal* constraints. These constraint types cannot be checked locally or globally; they rely on the construction of a reference set (or candidate set, as in optimality theory; see Prince & Smolensky (2004)) from which they pick out the candidate that is characaterized as optimal according to some appropriately defined evaluation metric. Transderivational constraints are typical of some (mostly earlier) versions of the minimalist program.

2.5.1   Fewest Steps

One such constraint is the Fewest Steps condition, which picks the derivation from a reference set that involves the smallest number of elementary operations; see Chomsky (1991; 1993; 1995), Epstein (1992), Collins (1994), Kitahara (1997), and Bošković (1997), among others. The Fewest Steps condition can be defined as in (8).

(8)   *Fewest Steps*:

   If two derivations D_1 and D_2 are in the same reference set and D_1 involves fewer operations than D_2, then D_1 is to be preferred over D_2.

As an illustration, consider the account of the difference between French and English with respect to V-to-T movement put forward in Chomsky (1991). A standard assumption is that the data in (9) show that French has overt V-to-T movement of finite verbs whereas English does not have such movement (except for auxiliaries).[5]

---

binding theory should be conceived of as an "anywhere" constraint, as in (ii) (see Belletti & Rizzi (1988), Epstein et al. (1998), Sabel (2011) for proposals and discussion).

(i)   a.   [TP [DP$_1$ Questi pettegolezzi su di sé$_2$]   T [VP [V′ preoccupano t$_1$] Gianni$_2$ più   di
                 these gossips   about himself   worry   Gianni more than
            ogni altra cosa]]
            anything else

      b.   [CP dass [TP sich$_1$ gestern   wieder die Kinder$_1$ t$_1$ geschlagen haben]]
                 that   each other yesterday again the children   hit   have

      c.   [CP [TP John$_1$ [T′ T [VP [V′ seems [PP to himself$_1$]] [TP t$_1'$ to be [AP t$_1$ clever]]]]]]

      d.   [DP$_3$ Which portrait of himself$_{1/2}$] does John$_2$ believe that Bill$_1$ criticized t$_3$?

(ii) *Principle A*:
      An anaphor is bound in its binding domain at some point of the derivation.

On this view, principle A is not a local constraint anymore: To find out whether (ii) is violated or not, the whole derivation must be scanned, and if principle A is to be respected, there must be at least one step of the derivation where the anaphor is bound within its binding domain. Hence, (ii) does in fact qualify as a global constraint.

[5] See Pollock (1989), based on earlier work by Emonds (1978). Also see Iatridou (1990) for a thorough critique of Pollock's proposal. The obvious alternative is to approach the data in (9) in terms of variation

(9) a.   Jean embrasse$_1$ souvent [$_{VP}$ t$_1$ Marie]
    b. *Jean souvent [$_{VP}$ embrasse$_1$ Marie]
    c. *John kisses$_1$ often [$_{VP}$ t$_1$ Mary]
    d.   John often [$_{VP}$ kisses$_1$ Mary]

Given (9), two questions arise. First, why do French verbs have to move to T? And second, why must English main verbs not move to T? Chomsky's basic assumption about the locus of parametrization concerns the notion of "strength" of T: French has strong T, English has weak T. Next, there is a constraint sensitive to the strong/weak distinction. This constraint can be formulated as in (10), and we may assume that it qualifies as a local derivational constraint restricting the application of movement operations.

(10) *Strength of T*:
     Strong T tolerates adjunction of all Vs; weak T tolerates adjunction only of "light" Vs (auxiliaries).

This excludes (9-c) in English: Overt V-to-T movement violates Strength of T. In contrast, overt V-to-T movement in (9-a) in French does not violate this constraint. Still, something extra needs to be said about (9-b) in French, which vacuously fulfills Strength of T, just like (9-d) in English does. Thus, the question is: Why is overt V-to-T raising obligatory if it is possible? Chomsky's (1991) background assumption is that inflection is base-generated in T. If V does not raise to T, T must be lowered to V in overt syntax, so as to fulfill another local constraint, the *Stray Affix Filter*, which prohibits inflectional affixes that are not attached to a verbal host. The crucial idea now is that overt T lowering creates an unbound trace that must be undone by LF, via covert raising of the whole V-T complex to the position of the trace of T. The derivations underlying (9-a) and (9-b) in French are given in (11) and (12), respectively.

(11) *The V-to-T derivation in French*:
    a.   Jean T$_2$ souvent [$_{VP}$ embrasse$_1$ Marie]   (overt raising)
    b.   Jean [$_{T_2}$ embrasse$_1$-T] souvent [$_{VP}$ t$_1$ Marie]

(12) *The V-in situ derivation in French*:
    a.   Jean T$_2$ souvent [$_{VP}$ embrasse$_1$ Marie]   (overt lowering)
    b.   Jean t$_2$ souvent [$_{VP}$ [$_{V_1}$ embrasse$_1$-T$_2$] Marie]   (covert raising)
    c.   Jean [$_{V_1}$ embrasse$_1$-T$_2$] souvent [$_{VP}$ t$_1$ Marie]

---

in adverb placement. Still, at least for the sake of the argument, I will assume for now that the idea that different verb movement options underlie the variation in (9) is correct.

Thus, the second derivation (in (12)) has more movement steps than the first one (in (11)), and it can therefore be filtered out as uneconomical by the transderivational economy constraint *Fewest Steps* in (8) – assuming, that is, that the two derivations are part of the same reference set. Various definitions of the notion of reference set have been proposed; a definition that works for the case at hand (but, it can be noted, not for certain other applications of the Fewest Steps condition that have been suggested in the literature) is given in (13).

(13) *Reference Set*:
      Two derivations D_1 and D_2 are in the same reference set iff (a) and (b) hold:
      a. D_1 and D_2 start with the same lexical array.
      b. D_1 and D_2 do not violate local or global constraints.

As a consequence of (13-a), an ordinary declarative sentence like (14-b) cannot accidentally block the *wh*-question in (14-a) even though it involves fewer syntactic operations. As a consequence of (13-b), an ill-formed version with *wh*-in situ as in (14-c) cannot accidentally block (14-a) even though it involves fewer syntactic operations by leaving the *wh*-phrase and the auxiliary in situ – (14-c) violates the (local) constraint that triggers *wh*-movement in English (whatever that is assumed to be).

(14) a.   What$_1$ have$_2$ you t$_2$ seen t$_1$?
      b.   You have seen a car
      c.   *You have$_2$ seen what$_1$?

Although historically it is the very first instantiation of an account relying on the Fewest Steps condition, this particular application of a transderivational constraint in the minimalist program does not seem to have ever been widely accepted, and it certainly is not generally accepted anymore today. Still, it spawned a number of further studies based on the Fewest Steps condition; and it can serve as an illustration of certain recurring properties and problems of transderivational constraints, and of competition-based approaches in general. A first problem noted by Chomsky (1991, 433) is that this system "tends to eliminate the possibility of optionality in derivation. Choice points will be allowable only if the resulting derivations are all minimal in cost ... This may well be too strong a conclusion, raising a problem for the entire approach." The problem can be illustrated with optional topicalization in English; see (15-a) vs. (15-b).

(15) a.   Mary gave a book to John$_1$
      b.   To John$_1$ Mary gave a book t$_1$

Several kinds of solutions can be proposed in view of the dilemma posed by (15-ab) for the Fewest Steps condition. Chomsky (1991) suggests that certain movement operations might be "assigned to some other component of the language system, perhaps

a 'stylistic' component of the mapping ... to PF;" movement operations of this type might then be exempt from the Fewest Steps constraint. A second possibility is to revise the definition of reference set appropriately, such that the two derivations in (15) do not compete anymore. For instance, one might add the requirement that competing derivations must have identical LF representations, as in (16). Assuming that sentences which differ only with respect to whether topicalization has applied must have different LFs, this would yield the desired result.

(16)   *Reference Set* (revised):
       Two derivations D_1 and D_2 are in the same reference set iff (a) and (b) hold:

       a. D_1 and D_2 start with the same lexical array and have the same LF representation.
       b. D_1 and D_2 do not violate local or global constraints.

Third, one might postulate that the PPs (i.e., *to John*) in (15) are in fact not fully identical. The idea would be that there is a [top] feature in the numeration (or lexical array) in (15-b) but not in (15-a); and then there is a local constraint (similar to the constraint triggering *wh*-movement) that forces [top]-marked items to undergo topicalization in English. On this view, the derivation generating (15-a) and the derivation generating (15-b) are not part of the same reference set, and each emerges as best in its own reference set. In particular, topicalization without a [top]-feature (which is string-identical to (15-b)) is blocked by Fewest Steps in the reference set that contains the derivation generating (15-a), and non-occurrence of topicalization in the presence of a [top]-feature (which is string-identical to (15-a)) is blocked by the local constraint triggering topicalization of [top]-marked items.

   A second potential problem related to the Fewest Steps condition (at least at first sight) arises with successive-cyclic movement: It is standardly assumed that long-distance *wh*-movement must be successive-cyclic; otherwise, a locality requirement will be violated. A constraint imposing a fairly strict locality requirement on movement operations is the Phase Impenetrability Condition (PIC) introduced in Chomsky (2000, 108) and Chomsky (2001b, 13), and adopted in much subsequent minimalist work. The PIC will emerge as the core of the present monograph. A restrictive version of the PIC is given in (17).

(17)   *Phase Impenetrability Condition* (PIC):
       The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

The notion of edge can be understood as in (18) (see Chomsky (2001b, 13)).

(18)   *Edge*:
       The edge of a head X is the residue outside of X′; it comprises specifiers of X

(and adjuncts to XP).

Finally, it remains to clarify what a phase is. Chomsky (2000; 2001b; 2008) proposes that phases are the "propositional categories" CP and vP.

(19) *Phase*:
      The propositional categories CP and vP are phases; other XPs (except perhaps for DP) are not.

Thus, it follows from the PIC that movement must apply successive-cyclically, via the phase edges of CP (i.e., SpecC) and vP (i.e., Specv). A string like (20-a) must therefore have the derivation in (20-c), and cannot have the derivation in (20-b), which has one less application of movement and thus looks like it should, ceteris paribus, be preferred by Fewest Steps.

(20) a. What do you want ?
      b. [$_{CP}$ What$_1$ do you [$_{vP}$ want t$_1$]] ?
      c. [$_{CP}$ What$_1$ do you [$_{vP}$ t$'_1$ want t$_1$]] ?

Clearly, successive-cyclic movement then creates a potential problem for the Fewest Steps condition: Successive-cyclic movement as in (20-c) should always be blocked by one-step movement as in (20-b), which violates the PIC. Assuming the PIC, we would then, it seems, incorrectly expect all long-distance dependencies to be ill formed. However, closer inspection reveals that this is in fact not the case under under the notion of reference set in (13) or (16): Only those derivations can compete that respect all local constraints of grammar, i.e., that are otherwise well formed. The derivation that generates the surface representation (20-b) violates the Phase Impenetrability Condition; hence, it cannot compete with the derivation that generates (20-c), and (20-c) is chosen by Fewest Steps because there is no competing derivation that would be more economical. Thus, successive-cyclic movement forced by locality requirements (like the PIC) can only pose a problem for the Fewest Steps condition if one were to dispense with clause (b) in the definition of reference set in (16) (or, for that matter, (13)), or if clause (b) were to be weakened in such a way that some derivations violating local constraints can compete after all. As a matter of fact, it has been argued (e.g., by Chomsky (1993; 1995) and Collins (1994)) that there is evidence for this latter option, and that clause (b) of the definition of reference set (repeated here, in adjusted form, in (21-a)) is indeed to be replaced with a weaker requirement that permits some ungrammatical derivations to compete; see (21-b).

(21) a. Two derivations can only compete if they do not violate local or global constraints.
      b. Two derivations can only compete if they both *converge*.

Here, *convergence* is to be defined in such a way that a violation of some local (or global) constraints leads to both ungrammaticality (or illformedness) and non-convergence, whereas a violation of some other local (or global) constraints, while producing ungrammaticality (or illformedness), still leads to convergence, and thus does not remove the derivation from the reference set, so that it may still block other derivations (despite not producing a good output itself).[6] If "not violating a local or global constraint" is replaced with the weaker requirement of "converging", the derivations generating (20-b) and (20-c) might compete after all (viz., if violation of the PIC does not automatically lead to non-convergence), and the problem of accounting for successive-cyclic movement under Fewest Steps would become real. How can successive-cyclic movement be reconciled with Fewest Steps under such an assumption? An answer is already provided in Chomsky's (1991) original analysis introducing Fewest Steps: The potential problem disappears if successive-cyclic movement is reinterpreted as a single (albeit complex) operation Form Chain, as in (22):

(22) *Form Chain*:
      Move α to its target position and freely insert intermediate traces in appropriate positions.

Assuming Form Chain, the derivations in (20-b) and (20-c) are equally costly from the point of view of Fewest Steps, and (20-c) can be chosen.[7] Still, the concept of Form Chain is far from being conceptually unproblematic. For one thing, it involves the stipulation of a fairly complex syntactic operation of exactly the kind that basic minimalist tenets would seem to preclude. For another, it is at variance with an elementary principle of derivational structure-building: the Strict Cycle Condition (SCC). A version of the SCC is given in (23) (see Chomsky (1973), Perlmutter & Soames (1979)).

(23) *Strict Cycle Condition* (SCC):
      Within the current XP α, a syntactic operation may not target a position that is included within another XP β that is dominated by α.

Insertion of intermediate traces that follows (or accompanies) the movement operation proper violates the Strict Cycle Condition because it qualifies as a syntactic operation that targets a position which is part of an XP domain (the vP, in the case currently

---

[6] To give a somewhat more concrete example: Collins (1994) argues that a derivation that violates the Condition on Extraction Domain (see below) may still successfully block a less economical derivation that does not, by Fewest Steps; thus, such a derivation, while yielding no good output, still converges.

[7] Note that the French case of T-lowering at S-structure followed by V-raising at LF in (12) would still have to involve two applications of Form Chain since Form Chain cannot span two levels of grammar (separated by spell-out) as a single operation.

under consideration) that is included in another ('current') XP domain (the root CP).[8]

2.5.2   Shortest Paths

Another transderivational constraint that has been proposed in various syntactic anal-
yses is the *Shortest Paths* condition (see Chomsky (1993; 1995), Kitahara (1993),
Collins (1994), Nakamura (1998)). The Shortest Paths condition can be defined as in
(24).

(24)   *Shortest Paths*:
       If two derivations D_1 and D_2 are in the same reference set and the movement
       paths of D_1 are shorter than the movement paths of D_2, then D_1 is to be preferred
       over D_2.

The following understanding of the concepts of movement path and path length is
informal but may suffice for present purposes.[9]

(25)   *Movment path* (informal):
       a.   A movement path is the set of nodes that are crossed by a movement oper-
            ation.
       b.   A movement path α is shorter than a movement path β if α has fewer nodes
            than β.

I will discuss two cases where the constraint has been put to use, viz., Collins (1994)
and Nakamura (1998).

       Collins (1994) is concerned with excluding cases of so-called yo-yo movement.
Yo-yo movement characterizes a combination of lowering and raising operations af-
fecting a single item in the course of a derivation. Collins argues that the availability of
yo-yo movement would make a wrong prediction for the West African language Ewe,
and he attempts to derive a ban on yo-yo movement from the Shortest Paths condition.

---

[8] In addition, adopting a Form Chain approach makes it necessary to view the PIC in (17) (more pre-
cisely, an appropriately modified version) as a representational constraint that can be violated throughout
the derivation and is only checked at the end of the derivation. Such a view would be incompatible with the
underlying rationale that has been advanced as an account of why something like the PIC should exist in
the first place; see below.

[9] A more precise definition is given in (i) (see Müller (1998, 130); also cf. Pesetsky (1982, 289), Collins
(1994, 56)).

(i)   *Movement path*:
      The path from X to Y is the set of categories Z such that (a) and (b) hold:
      a.   Z is reflexively dominated by the minimal XP that dominates both X and Y.
      b.   Z dominates X or Y.
      The length of a path is determined by its cardinality.

Ewe is among the languages that show reflexes of of successive-cyclic *wh*-movement
in the C domain. The reflex of successive cyclicity concerns the morphological form
of the 3. person singular subject pronoun in the canonical subject position. The regu-
lar form of the pronoun is *é*; cf. (26-a). The regular pronoun *é* can be replaced by *wo*
in cases of long-distance extraction (focus movement, in the case at hand); cf. (26-b).

(26) a. Kofi gblɔ [$_{CP}$ be   é/*wo ƒo Kɔsi]
          Kofi said   that he   hit Kɔsi

     b. Kofi$_1$ ɛ   me gblɔ [$_{CP}$ (t$'_1$) be   é/wo ƒo t$_1$]
          Kofi Foc I   said   that he   hit

Two assumptions prove important for the account that Collins gives. First, *é* is re-
placed by *wo* iff the local SpecC position is filled. And second, long-distance move-
ment of argument DPs in Ewe may apply either successive-cyclically, via SpecC (in
which case *wo* is obligatory), or in one step (in which case *wo* is impossible).[10]

The basic observation that Collins wants to account for then is that movement
that originates in the matrix CP, and that targets the SpecC position in the matrix
clause, may never trigger a reflex of successive-cyclicity in the embedded CP; i.e.,
a sentence like (27-a) cannot involve lowering of the moved item to the embedded
SpecC position, followed by raising to the ultimate SpecC landing site in the matrix
clause, as shown in (27-b).

(27) a. Kofi$_1$ ɛ   me gblɔ na t$_1$ [$_{CP}$ be   é/*wo ƒo Kɔsi]
          Kofi Foc I   said to   that he   hit Kɔsi
     b. (i)   [$_{VP}$ said [$_{PP}$ to Kofi$_1$] [$_{CP}$ that [$_{TP}$ he hit Kɔsi]]]
        (ii)   [$_{VP}$ said [$_{PP}$ to t$_1$] [$_{CP}$ Kofi$_1$ that [$_{TP}$ he hit Kɔsi]]]   ...
        (iii) [$_{CP}$ Kofi$_1$ Foc [$_{TP}$ T [$_{VP}$ said [$_{PP}$ to t$_1$] [$_{CP}$ t$'_1$ that [$_{TP}$ he hit Kɔsi]]]]]

Collins (1994) observes that the derivation in (27) (call it D$_1$) is blocked by another
derivation (call it D$_2$) via Shortest Paths: As illustrated in (28), D$_2$ proceeds without
intermediate lowering; therefore, its movement paths are shorter.[11]

---

[10] For the sake of presenting this argument, I will leave open the question of how a fatal PIC violation can
be circumvented in the latter derivation.

[11] Three remarks are in place here, two of them related. First, at least at first sight, it looks as though the
illformedness of the derivation in (27) already follows from the Strict Cycle Condition, with no reference
to a Shortest Paths constraint necessary: On this view, the lowering step would qualify as counter-cyclic.
Second, again at least at first sight, one might think that Fewest Steps would also suffice to block D$_1$ in
favour of D$_2$. However, neither the Strict Cycle Condition nor Fewest Steps can successfully rule out D$_1$
if the concept of Form Chain is embraced (assuming, of course, that the very concept of Form Chain can
be reconciled with the Strict Cycle Condition in some way or other): If so, the two movement operations
in D$_1$ can be reanalyzed as a single instance of Form Chain, and they do not violate the SCC any more
than ordinary instances of successive-cyclic movement do. – In addition, note that this account potentially

(28)   a.   [$_{VP}$ said [$_{PP}$ to Kofi$_1$] [$_{CP}$ that [$_{TP}$ he hit Kɔsi]]]
        b.   [$_{CP}$ Kofi$_1$ Foc [$_{TP}$ T [$_{VP}$ said [$_{PP}$ to t$_1$] [$_{CP}$ that [$_{TP}$ he hit Kɔsi]]]]]

As a second illustration of how the Shortest Paths condition can be put to use, consider the case of *wh*-movement in Tagalog that is addressed in Nakamura (1998). The relevant generalization is that (as in other Austronesian languages) only the highest A-position of a given clause (the subject position SpecT) is accessible for *wh*-movement in Tagalog. In constructions in which an agent DP occupies the highest A-position (the so-called *Agent Topic* (AT) constructions), this DP can be *wh*-moved; a DP bearing a different Θ-role that shows up in an object position cannot undergo *wh*-movement. This is shown in (29).

(29)   a.   [$_{CP}$ Sino$_1$ ang [$_{TP}$ t$'_1$ b-um-ili   [$_{VP}$ t$_1$ t$_V$ ng damit$_2$]]] ?
              who   Ang   bought$_{AT}$ dress$_{inh}$
              'Who is the one that bought the dress?'
        b.   *[$_{CP}$ Ano$_2$ ang [$_{TP}$ *si Juan$_1$* b-um-ili   [$_{VP}$ t$_1$ t$_V$ t$_2$]]]?
              what   Ang   Juan$_{abs}$ bought$_{AT}$
              'What is the thing that Juan bought?'

A different marking on the verb signals the so-called *Theme Topic* (TT) construction. Here, the theme DP occupies the structural subject position SpecT; and indeed, only the theme DP can undergo *wh*-movement, blocking movement of the agent DP; compare the pattern in the TT construction in (30) with the reverse pattern in the AT construction in (29).

(30)   a.   *[$_{CP}$ Sino$_1$ ang [$_{TP}$ *ang damit$_2$* b-in-ili   [$_{VP}$ t$_1$ t$_V$ t$_2$]]] ?
              who   Ang   dress$_{abs}$ bought$_{TT}$
              'Who is the one that bought the dress?'
        b.   [$_{CP}$ Ano$_2$ ang [$_{TP}$ t$'_2$ b-in-ili   [$_{VP}$ ni Juan t$_V$ t$_2$]]] ?
              what   Ang   bought$_{TT}$ Juan$_{erg}$
              'What is the thing that Juan bought?'

Nakamura's idea now is that the derivations generating (29-a) and (30-a) compete, as do the derivations generating (29-b) and (30-b). The derivations underlying (29-a) and (30-b) can then block their respective competitors as ungrammatical because of the Shortest Paths constraint. Consider, e.g., *wh*-movement of the theme DP in (29-b) and (30-b). The movement path from the VP-internal object position to the SpecC

---

has non-trivial consequences for the definition of reference sets: The subject pronoun "he" shows up as *wo* in D$_1$ and as *é* in D$_2$. Hence, this difference must not suffice to create different reference sets. This would follow either under the view that the difference is "minor" (in a sense to be made precise), and minor differences in lexical arrays do not count; or under the assumption that morphological realization of function words is brought about by a post-syntactic operation ('late insertion'), with the two feature bundles for the subject pronoun identical in syntax (see Halle & Marantz (1993; 1994), among others).

target position in (29-b) is longer than the path from the subject position SpecT to SpecC in (30-b).[12]

2.5.3 Merge Before Move

The Merge before Move constraint is introduced in Chomsky (1995), and maintained in much subsequent minimalist work (e.g., Chomsky (2000; 2001b; 2005), Frampton & Gutman (1999), Hornstein (2001; 2009)). The definition of the constraint given in Frampton & Gutman (1999) suggests that it is a transderivational constraint; see (31).[13]

(31) *Merge before Move*:
Suppose that two derivations D_1 and D_2 are in the same reference set and respect all local constraints, and $D_1 = <\Sigma_0, ..., \Sigma_n, \Sigma_{n+1}, ...\Sigma_k>$ and $D_2 = <\Sigma_0, ..., \Sigma_n, \Sigma'_{n+1}, ...\Sigma'_{k'}>$. Then D_1 is to be preferred over D_2 if $\Sigma_n \to \Sigma_{n+1}$ is an instance of Merge and $\Sigma_n \to \Sigma'_{n+1}$ is an instance of Move.

The data that originally motivated this constraint come from expletive constructions in English; see (32).

(32) a. There$_1$ seems [$_{TP}$ t$_1$ to be [$_{PP}$ someone$_2$ in the room]]
b. *There$_1$ seems [$_{TP}$ someone$_2$ to be [$_{PP}$ t$_2$ in the room]]

Both derivations have an expletive *there* in the original lexical array. In (32-a), the EPP property of the infinitival TP is satisfied by merging the expletive *there* (and leaving

---

[12] To fully make the analysis work, several additional assumptions must be made. For instance, it must be ensured that two derivations can compete even though they do not have identical lexical material: The Agent Topic and the Theme Topic constructions differ in lexical make-up – with respect to both the *form* of the verb and the *morpho-syntactic features* involved. In view of this, Nakamura suggests that "identical lexical array" is to be replaced by "non-distinct lexical array" in the definition of reference sets, with the latter defined in such a way that two lexical arrays that only differ with respect to functional features do not count as distinct. (Assuming a late insertion analysis, as envisaged as a possibility for the *é/wo* alternation in Ewe in footnote 11, would not help in the case at hand because the morpho-syntactic feature bundles – which would act as items of the lexical array on this view – would still be different in Agent Topic and Theme Topic constructions.)

Another question that arises with the reasoning just given in the main text concerns the relevance of the initial movement steps that move agent and theme DPs to the SpecT positions in Agent Topic and Theme Topic constructions, respectively. The derivation that generates (30-b) may minimize the *wh*-path in comparison with the derivation that generates (29-b), but it increases path lengths in the TP-domain. Thus, if the Shortest Paths account of *wh*-movement options is to be maintained, prior A-movement operations must essentially be ignored for the purposes of this transderivational constraint. To achieve this goal, Nakamura (1998) replaces the notion of "movement paths" in the definition of the Shortest Paths condition with the more specific notion of "comparable chain links".

[13] That said, it should presumably be possible to maintain the gist of the constraint in a formulation that is not transderivational in the strictest sense. Still, a comparison of alternatives is unavoidable.

the DP *someone* in situ); the expletive is then moved to the matrix CP, satisfying the EPP property of the matrix T node. In contrast, in (32-b), the EPP property of the infinitival TP is satisfied by moving the DP *someone* from its in-situ position.[14] After that, *there* is merged in the matrix SpecT position, thereby satisfying matrix T's EPP feature. Nothing seems to be wrong with either derivation as far as local (and global) constraints are concerned; however, Merge before Move successfully blocks (32-b) in favour of (32-a): On the embedded TP level, Merge of *there* is preferred over Move of *someone*.

2.5.4   Other Transderivational Constraints

Several other transderivational constraints have been suggested over the years. Among them are the *Procrastinate* constraint (see Chomsky (1993; 1995)) and the *Preference Principle for Reconstruction* (see Chomsky (1993)).

2.6   Translocal Constraints

Finally, there are translocal constraints. Like a transderivational constraint, a translocal constraint picks an optimal candidate from a reference set of competing candidates (to use the terminology introduced in optimality theory, cf. Prince & Smolensky (2004)). However, the status of the candidates is different: With translocal constraints, candidates are output representations rather than full derivations (as is the case with transderivational constraints).

A typical example of a translocal constraint is the Avoid Pronoun principle of Chomsky (1981). Chomsky proposes the Avoid Pronoun principle as a genuinely grammatical (i.e., non-pragmatic) constraint; and there seems to be no option to interpret it in a local or global way (in other words, reference to other sentences is necessary in order to determine the grammaticality status of a given sentence). The original empirical evidence comes from English gerunds. A background assumption is that all entries in the argument structure (or Θ-grid) of a predicate must be represented as arguments in the syntax. In those cases where no external argument DP is visible, there is a non-overt argument; and in the case of English gerunds, various restrictions conspire so as to ensure that the only non-overt argument that is permitted in the subject position here is the empty category PRO. With this in mind, consider the pair of sentences in (33-ab).

---

[14] This is an option that exists as such. Compare (i) (which is not generated by a derivation that is part of the same reference set as the derivations generating the sentences in (32) because it lacks the expletive *there* in the initial lexical array).

(i)   Someone$_2$ seems [$_{TP}$ t$'_2$ to be [$_{PP}$ t$_2$ in the room]]

(33) a.   John$_1$ would much prefer [PRO$_1$ going to the movie]
    b. *John$_1$ would much prefer [PRO$_{2/arb}$ going to the movie]

As shown by the illformedness of (33-b) and the wellformedness of (33-a), PRO must be obligatorily controlled by a matrix argument if it shows up in the subject position of a gerund in English. This can be taken to follow from a local constraint like (34) (see Manzini (1983)).[15] (33-b) violates (34) because matrix subject and embedded subject are contra-indexed; (33-a) does not because the two subjects are co-indexed, and control is effected.

(34) *Constraint on Control*:
    If PRO is minimally dominated by a declarative clausal complement α, then it must be controlled by an antecedent within the minimal CP that dominates α.

Interestingly, the subject position of English gerunds can alternatively also be realized by a genitive DP or possessive pronoun. However, in this case, co-indexing of embedded and matrix subject is impossible; see (35-a) vs. (35-b).

(35) a.   *John$_1$ would much prefer [$_{DP}$ his$_1$ going to the movie]
    b.   John$_1$ would much prefer [$_{DP}$ his$_2$ going to the movie]
    c.   John$_1$ would much prefer [$_{DP}$ his$_1$ book]

It is a priori unclear why (35-a) is ungrammatical. In particular, note that constraints of the binding theory (more precisely, principle B) cannot be involved. If the pronoun *his*$_1$ were too close to its antecedent *John*$_1$ in (35-a), it would also be too close in the regular (i.e., non-gerund) DP construction (35-c), where, however, co-indexing is unproblematic. In view of this, Chomsky's (1981) proposal is that (35-a) is blocked by the translocal Avoid Pronoun principle. The original formulation is given in (36).

(36) *Avoid Pronoun*:
    Lexical pronouns are blocked by empty pronouns if possible.

Here, the qualification "is possible" is to be interpreted as "if no local or global constraints are violated", and the competition between "lexical pronouns" and "empty pronouns" presupposes that a reference set has to be constructed for the purposes of

---

[15] In the formulation given here, the constraint is representational, but a derivational version of the constraint does not seem impossible. It is also worth pointing out that the question of whether the constraint in (34) should be taken to be primitive, or derivable from some more basic assumptions, is immaterial for the issues to be discussed here (i.e., the working of the Avoid Pronoun principle). In Manzini's approach, (34) is in fact not primitive; it follows from a simpler but much more abstract requirement (viz., that PRO is bound in its D-governed category). See Stechow & Sternefeld (1988, 314) for extensive discussion. Also see Landau (2000), Hornstein (2001), and Boeckx et al. (2010) for some more recent attempts to derive (something like) Manzini's Constraint on Control.

the translocal Avoid Pronoun Principle (just as was the case with transderivational con-
straints), with the competition between sentences with overt pronouns and sentences
with non-overt pronouns resolved in favour of the latter as long as that candidate does
not violate other (local or global) constraints. Thus, a more precise definition of the
Avoid Pronoun principle might look as in (37).

(37) *Avoid Pronoun* (different formulation):
      If two representations R_1 and R_2 are in the same reference set and R_1 uses a
      lexical pronoun where R_2 uses an empty pronoun, then R_2 is to be preferred
      over R_1.

The idea now is that (35-a) (with co-indexing of the two subjects and an overt pronoun
in the gerund) is blocked by (33-a) (with co-indexing of the two subjects and a non-
overt pronoun in the gerund), as a consequence of the Avoid Pronoun principle. As
before, questions arise with respect to the proper understanding of what reference sets
are. For the case at hand, it is clear that two candidates can only belong to the same
reference set if they have an identical semantic representation; in particular, different
indices lead to different reference sets (otherwise, (35-b), with an overt pronoun and
contra-indexing, would be expected to be blocked by (33-a), where the two subjects
are co-indexed and the gerund's subject is a non-overt pronoun). Furthermore, it must
of course be ensured that the overt/non-overt distinction with pronouns cannot create
two different reference sets – otherwise, (33-a) could never block (35-a). A definition
of reference set as in (38) meets these requirements.[16]

(38) *Reference Set* (revised; for representations):
      Two representations R_1 and R_2 are in the same reference set iff:
      a. R_1 and R_2 have the same lexical categories and the same LF representation.
      b. R_1 and R_2 do not violate local constraints.

A further application of the Avoid Pronoun principle is proposed by Haegeman (1995)
as an account of the strong tendency of subject pro-drop languages (where the subject
position of a finite clause may remain empty) to use a non-overt subject pronoun (for
which the empty category *pro* has been suggested in Chomsky (1982b), and is still
widely adopted nowadays) instead of an overt subject pronoun "if possible", i.e., un-
less the subject needs to be focussed (or be rendered prominent in some other way).[17]

---

[16] As an alternative to the confinement to *lexical* (as opposed to *functional*) categories in the initial lexical
array, one might again adopt a late insertion approach according to which overt and non-overt pronouns
may be identical in the syntax (see above).

[17] Also see Grimshaw & Samek-Lodovici (1998) for a somewhat more fine-grained optimality-theoretic
approach that nevertheless shares the same fundamental insight.

The Avoid Pronoun principle is not the only translocal constraint that has been proposed; another translocal constraint is *Economy of Representation*, which is introduced in Chomsky (1991). The constraint demands the minimal structural representation (in a sense that needs to be made precise) that is "possible" (i.e., compatible with local and global constraints).[18] Closer scrutiny reveals that translocal constraints can usually be reconstructed as transderivational constraints. Thus, the Avoid Pronoun principle could in principle be rephrased as a transderivational constraint without too much ado (by referring to two derivations that differ only with respect to the nature of a subject), and the same goes for Economy of Representation (where the derivation is chosen that generates the minimal representation that is possible).

2.7 *Good Constraints*

Having introduced the various constraint types that figure prominently in analyses developed in syntactic theory, let me step back at this point and ask a more general question: What are the properties that a constraint must have so as to qualify as good (where "good" in this context is to be understood as "more likely to be a closer approximation to reality than others", or "more likely to permit a better modelling of the faculty of language than others")? A standard assumption is that if there are two theories that account for the same set of data, the one is to be preferred that is superior with respect to simplicity, generality, and elegance. Against this background, the following meta-constraint on constraints can be formulated; a good constraint adheres to the requirements embodied in (39).[19]

(39)   Constraints should be as simple and general as possible.

Are the constraints mentioned so far simple and general? In at least some of the cases, there is reason to doubt this. The ECP in (5), for instance, with its highly complex inner structure and with its disjunction (by demanding antededent-government or lexical government), certainly does not look simple and general. Similarly, the Constraint on Control in (34) does not seem to be maximally simple and general. It mentions a specific category label (viz., C) and is highly construction-specific because it talks about "declarative clausal complements" to ensure that the restriction on control does not apply in questions or in non-complements (but cf. footnote 15 on the status of (34) as primitive or derived).

---

[18] Also see Haider (1993a; 2000a) for a similar concept.

[19] I have omitted the requirement of elegance in (39) since, in contrast to simplicity and generality, elegance is a property that only holds of theories, not of individual constraints; on the constraint level, the relevant property would be something like "contributing to elegance", which one may thus tacitly add in (39). Also see Grimshaw (1997; 1998), who adopts (39) in the context of an optimality-theoretic approach but also points out that it should be taken to hold independently of the choice of syntactic theory.

What about the Subjacency Condition in (2)? Arguably, this constraint fails requirement (39), too. Even disregarding the issue of parametrization of bounding nodes, the Subjacency Condition (in the form that it takes in a single language) mentions specific category information (e.g., the category features D and T in English); this invariably detracts from generality. Furthermore, the Subjacency Condition requires keeping track of crossed bounding nodes and keeping them in storage – the information that a first bounding node has been crossed must still be available when a second bounding node is crossed (thus, counting must be involved). As pointed out by Riemsdijk (1978) and Koster (1978), other things being equal, a single-node, category-neutral constraint is inherently preferable to a two-node, category-sensitive constraint like the Subjacency Condition vis-à-vis the meta-requirement in (39).[20] That said, the Subjacency Condition succeeds in giving a unified explanation for a number of different phenomena that had so far been treated by separate constraints (see the next subsection for some of these constraints); so, while it may not be a "good" constraint, it is not exactly a "bad" constraint either because it makes general (and elegant) analyses possible.

Another property of "good" constraints suggests itself when the taxonomy of constraint types introduced above is considered. The constraint types can be ordered with respect to increasing complexity as in (40).

(40)   *Complexity of constraint types*:
       derivational constraints, representational constraints < global constraints < translocal constraints < transderivational constraints

---

[20] Riemsdijk (1978) and Koster (1978) also argue that the one-node constraint that is to replace the Subjacency Condition (the category-neutral, derivational Head Constraint and the category-neutral, representational Bounding Condition, respectively) is empirically superior. Here is just one well-known empirical problem for the Subjacency Condition. As will be discussed in some detail below, there is an asymmetry between extraction from subject and extraction from object: Whereas the former is typically impossible, the latter is often permitted. However, the Subjacency Condition fails to distinguish between argument DPs that are external arguments merged in specifier positions (subjects) and argument DPs that are merged in complement positions (objects). Hence, it seems that the constraint is too strong; it rules out examples like (i-ab). Still, it can be noted that it also rules out (i-cd), which as such is a welcome result – (i-c) has another type of embedding predicate, and (i-d) has replaced the indefinite determiner of (i-abc) with a more specficic, definite determiner (a so-called *Specificity* effect).

(i)   a. $[_{DP_1}$ Which author] did $[_{TP_3}$ you read $[_{DP_2}$ a book about t_1]] ?
       b. $[_{DP_1}$ Who$_1$] did $[_{TP_3}$ you see $[_{DP_2}$ a picture of t_1]] ?
       c.   *$[_{DP_1}$ Which author] did $[_{TP_3}$ you destroy $[_{DP_2}$ a book about t_1]] ?
       d.   ?*$[_{DP_1}$ Which author] did $[_{TP_3}$ you read $[_{DP_2}$ the book about t_1]] ?

The notion of complexity used here is a purely linguistic one: A constraint α is more complex than another constraint β if the linguistic object that α applies to is more complex than the linguistic object that β applies to. According to this criterion, local constraints (derivational and representational) qualify as less complex than global constraints: Both the derivational step and the structural representation are less complex objects than a full derivation (i.e., a succession of syntactic representations). Similarly, global constraints emerge as less complex than translocal constraints and transderivational constraints because the latter two constraints require the creation of a reference set of competing candidates, which is more complex than a single linguistic object (even if it is a full derivation, as with global constraints). Finally, transderivational constraints are predicted to be more complex than translocal constraints because the candidates that are collected in the reference set are full derivations in the former case, and simple representations in the latter. Taken together, this gives rise to the order depicted in (40).[21] On this basis, (41) can be assumed as a second meta-constraint on syntactic constraints.

(41)   Constraints should not be complex.

Given (40), transderivational, translocal, and global constraints are excluded, leaving only local (derivational or representational) constraints. This is in line with work carried out in the minimalist program over the last decade or so, which has generally strived to minimize the role of transderivational (and translocal) constraints – ideally, to dispense with them altogether. (Similar conclusions hold for the role of global constraints in the minimalist program.) See, e.g., Müller & Sternefeld (1996), Collins (1997), Epstein et al. (1998), Frampton & Gutman (1999), and Chomsky (2000; 2001b; 2005; 2007; 2008). The underlying reasons for this are typically complexity-related. However, it is worth pointing out that the complexity issue is not the only problem incurred by transderivational (and translocal) constraints. Another recurring problem is that it has proven to be extremely difficult to come up with a uniform

---

[21] Let me emphasize that no claims about *computational complexity* are involved in deriving (40), and I do not think that this notion should be directly invoked in the present context since what is at issue in the present context is the modelling of competence (the faculty of language), not of performance (parsing techniques); also see Graf (2008b, 9), who points out that "allowing transderivational computations in a grammar does not imply that any such computations are actually carried out during processing". Still, it is likely that the list in (40) can to a large extent be mirrored if a purely computational perspective is adopted. Graf (2008a;b) investigates the various constraint classes in a model-theoretic approach to syntax, on the basis of a more formal approach to the notion of complexity. (In doing so, he relies on Rogers's (2003) extension of model-theoretic syntax which allows three-dimensional objects that can encode full derivations.) This yields some refinements and potential qualifications with respect to (40), but the basic differences are confirmed: Local derivational constraints are less complex than global constraints, and transderivational and translocal constraints are most complex.

definition of reference set that works for different transderivational (or translocal) constraints, and with different (as such well-motivated) applications of one and the same transderivational (or translocal) constraint; see Sternefeld (1996a) and Johnson & Lappin (1997).

Furthermore, non-trivial questions arise with respect to the interaction of different transderivational (or translocal) constraints (see Johnson & Lappin (1997)). Let me briefly dwell on this issue. The basic question is this: How does optimality evaluation proceed in the presence of more than one transderivational constraint in the grammar? In this case, conflicts may arise. As a simple, abstract example, suppose that there are two transderivational constraints (T_1 and T_2), and only three derivations (D_1, D_2, and D_3) in the reference set. Suppose further that T_1 prefers D_1 over D_2 and D_3, and that T_2 prefers D_2 over D_1 and D_3; there simply is no competing derivation D_0 that would be preferred by both T_1 and T_2 because such a derivation D_0 necessarily violates a local constraint (or, at least, fails to converge; see above). In such a situation, various possibilities arise, which for present purposes may be labelled (i) *tolerance*, (ii) *ranking*, and (iii) *breakdown*. According to (i) *tolerance*, it suffices for a derivation to be selected by one transderivational constraint to be optimal; consequently, both D_1 and D_2 would be classified as optimal. According to (ii) *ranking*, the conflict among transderivational constraints is resolved by a ranking, such that that derivation is optimal that is preferred by the higher-ranked constraint in the case of conflict.[22] If T_1 is ranked higher than T_2, this would imply that only D_1 is optimal. Finally, according to (iii) *breakdown*, no derivation can emerge as optimal in the case of conflicting instructions made by transderivational constraints. This issue seems to be rarely discussed in the literature, but where it is, it is option (iii) that is adopted: Either all transderivational (and translocal) constraints converge on a single optimal derivation, or there is no well-formed output for a given reference set (see Collins (1994), Sternefeld (1996a), and Müller (2000)). On this view, grammaticality of a sentence can be understood as follows:

---

[22] This may look like an instance of optimality-theoretic grammar (see Prince & Smolensky (2004)), but the correlation is not as direct as one might assume at first sight. The ranked constraints of optimality theory are typically local in nature, not transderivational or translocal (as in the case currently under consideration). In contrast, transderivational (translocal) constraints have a status similar to the central principle of harmony evaluation ('optimality') in optimality-theoretic grammar, in the sense that both transderivational (translocal) constraints and harmony evaluation select an optimal candidate in a reference (or candidate) set, with optimality defined in terms of (mostly) economy concepts like minimization of operations (Fewest Steps) or minimization of path lengths (Shortest Paths) in the former case, and in terms of best constraint profile in the latter. Thus, ranking of transderivational constraints cannot directly correspond to anything in optimality theory since there is only one such constraint (the principle of optimality) here.

(42) *Grammaticality in a minimalist syntax with complex constraints*:
      A sentence is grammatical if the derivation D_i that generates it respects the
      requirements in (a) and (b):

      a. D_i does not violate a local constraint.
      b.   There is no derivation D_j in the same reference set that is preferred over D_i
           by a transderivational or translocal constraint.

To sum up, an interaction of transderivational constraints (with possibly conflicting
choices of derivations) may not pose an unsurmountable problem; but it will invariably
require extra assumptions, and it remains to be seen whether these assumptions may
then lead to inconsistencies (so far, this issue does not seem to have been tackled in
detail in the literature).

   More generally, I will adopt the two meta-constraints on constraints in (39) and
(41) for the remainder of this monograph. Among other things, this implies that con-
straints like the Subjacency Condition, the ECP, the Projection Principle, the Fewest
Steps condition, and the Shortest Paths condition cannot be maintained. Against this
background, I will now turn to the two locality constraints on movement that form
the central topic of the present study, viz., the (Generalized) Minimal Link Condition
((G)MLC) and the Condition on Extraction Domain (CED). I proceed by introducing
some predecessors and alternatives first.

3.   Towards the (G)MLC and the CED

3.1   The A-over-A Principle

The first substantial locality constraint on movement (and other) operations that has
been proposed is the A-over-A Principle introduced in Chomsky (1964); see (43).

(43) *A-over-A Principle*:
      In a structure ... [$_A$ ... [$_A$ ...] ...] ..., an operation can only affect the higher,
      more inclusive category A.

Here, "A" is a variable over category features. The constraint states that movement
of some category α to some landing site δ is impossible if there another category β
with the same category feature that dominates α (in this case, β is *closer* to δ than
α because the path from β to δ is necessarily shorter than the path from α to δ; see
below). The A-over-A Principle is a local derivational constraint. To find out whether
a given derivation respects it or not, each (Move) operation must be checked, by taking
into account the phrase marker constructed so far.

   A first consequence of the A-over-A Principle is illustrated in (44) (with examples
from Riemsdijk & Williams (1986, 21)). (44-a) instantiates a DP-over-DP structure.
As shown by the contrasts between (44-b) and (44-c) (for *wh*-movement), and between
(44-d) and (44-e) (for relativization), only the higher, "more inclusive" DP$_1$ can un-

dergo movement in this configuration; and this is exactly what the A-over-A Principle predicts.

(44) a. $[_{DP_1}$ My letter to $[_{DP_2}$ a friend in Italy]] got lost
        b.   *$[_{DP_2}$ Who] did $[_{DP_1}$ my letter to t_2] get lost ?
        c. $[_{DP_1}$ Which letter to $[_{DP_2}$ a friend in Italy]] got lost?
        d.   *John is the friend $[_{DP_2}$ who] C $[_{DP_1}$ my letter to t_2] got lost
        e.   This is the letter $[_{DP_1}$ which] t_1 got lost

In the same way, the A-over-A principle accounts for the data in (45), where the relation between the two DPs is not as local as in (44) because the lower DP is included in a CP that is a complement of the head noun. The example pair in (45-bc) involves *wh*-movement, the two sentences in (45-de) are instantiations of relativization.[23]

(45) a.   John heard $[_{DP_1}$ a rumour that you had read $[_{DP_2}$ this book]]
        b.   *$[_{DP_2}$ What] did John hear $[_{DP_1}$ a rumour that you had read t_2]] ?
        c. $[_{DP_1}$ Which rumour that you had read $[_{DP_2}$ this book]] did John hear ?
        d.   *This is a book $[_{DP_2}$ which] John heard $[_{DP_1}$ a rumour that you had read t_2]
        e.   This a rumour $[_{DP_1}$ which] John heard t_1

As pointed out in Ross (1967), in the formulation given in (43), the A-over-A Principle is both too strong and too weak (the same goes for the representational version in (i) in the last footnote). The second problem makes the A-over-A Principle look less

---

[23] In its standard formulation, the A-over-A Principle is a derivational constraint. However, it can straightforwardly be reformulated as a representational constraint on outputs; see (i).

(i)   *A-over-A Principle* (representational version):
        *... A_2 ... $[_{A_1}$ ... t_2 ...] ...] ...

Crucially, this formulation relies on the existence of traces. Indeed, this is in fact one of the two main reasons why one would want to postulate traces in the first place (the other main reason being that traces are relevant for semantic interpretation): Traces are needed by representational locality constraints on movement so as to encode earlier stages of the derivation whose properties would otherwise be lost. Thus, one should keep in mind that if the role of representational constraints is reduced in a grammatical theory, so is the evidence for traces as remnants of movement.
Another, extremely local, way of implementing the A-over-A Principle is suggested in Gazdar (1981, 161) and Gazdar (1982, 176-177). In this approach, displacement in syntax is not brought about by movement operations, but by local passing on of SLASH features (/) (that encodes all relevant information about the displaced item) from the base position (in which a trace shows up) to the antecedent, where the SLASH feature is instantiated on a mother if it shows up on a daughter. The A-over-A Principle can then be formulated as a ban on the instantiation of a SLASH feature with category label A on a node with the same category label A, as in (ii).

(ii) *$[_{A/A}$ ...]

plausible; the first problem is potentially severe. The examples in (46) (cf. Riemsdijk & Williams (1986)) highlight a case where the A-over-A principle is too weak, in the sense that it does not rule out an ill-formed example that it should arguably rule out on a par with the ill-formed sentences in (45) and (44): In the DP-over-AP structure, only the more inclusive DP can undergo movement, not the lower AP category. This does not follow from the A-over-A principle because the two categories have different category features (D vs. A).

(46) a. You have [$_{DP_1}$ a [$_{AP_2}$ very intelligent] sister]
      b. [$_{DP_1}$ [$_{AP_2}$ How intelligent] a t$_2$ sister] do you have ?
      c. *[$_{AP_2}$ How intelligent] do you have [$_{DP}$ a t$_2$ sister] ?

Similarly, the ill-formed sentences in (47) are not blocked by the A-over-A principle despite exhibiting the same kind of structural properties as the ungrammatical examples in (44) and (45) (see Riemsdijk & Williams (1986)). The sentences involve a PP-over-DP structure in German, with movement of the lower DP category across the intervening PP (an instance of preposition stranding) leading to ungrammaticality. Again, since the labels of the categories are not identical (P vs. A), the A-over-A principle has nothing to say about these data.

(47) a. Sie spielt [$_{PP_1}$ mit [$_{DP_2}$ dem grünen Auto]]
          she plays   with   the green car
      b. [$_{PP_1}$ Mit [$_{DP_2}$ welchem Auto]] spielt sie t$_1$?
          with   which   car   plays she
      c. *[$_{DP_2}$ Welchem Auto] spielt sie [$_{PP_1}$ mit t$_2$] ?
          which   car   plays she   with
      d. [$_{PP_1}$ Mit [$_{DP_2}$ dem grünen Auto]] spielt sie t$_1$
          with   the green car   plays she
      e. *[$_{DP_2}$ Diesem Auto] spielt sie [$_{PP_1}$ mit t$_2$]
          this   car   plays she   with

Failure to exclude the ungrammatical sentences in (46) and (47) may make the A-over-A principle somewhat less convincing, but it is certainly not fatal: The examples could well be excluded by some other constraint. The seemingly identical pattern would then have to be viewed as a spurious generalization.[24]

---

[24] Given that German has postposition stranding, as in (i-c) vs. (i-b), there might be independent evidence for such a view.

(i) a. Sie spielt [$_{PP_1}$ da$_2$-mit]
         she plays   there-with
     b. [$_{PP_1}$ Wo$_2$-mit   ] spielt sie t$_1$?
             where-with   plays she

Cases where it looks as though the A-over-A Principle is too strong because it rules out well-formed examples are potentially much more severe. In (48-ab), for instance, the lower DP is moved out of a higher DP in a DP-over-DP structure, in clear violation of the A-over-A Principle.

(48) a. $[_{DP_2}$ Who] would you approve of $[_{DP_1}$ my seeing t_2]] ?

b. $[_{DP_2}$ Which author] did you read $[_{DP_1}$ a book about t_2] ?

The same pattern can be seen in (49), where a lower CP is extracted from a more inclusive CP in a CP-over-CP structure. This operation is legitimate in English (see (49-ab)) as it is in German (see (49-cd)). Again, the wellformedness of (49-bd) is unexpected, given the A-over-A Principle.

(49) a. John wouldn't say $[_{CP_1}$ that Mary thinks $[_{CP_2}$ that Bill is nice]]

b. $[_{CP_2}$ That Bill is nice] John wouldn't say $[_{CP_1}$ that Mary thinks t_2]

c. Fritz   hat behauptet $[_{CP_1}$ Maria   würde denken $[_{CP_2}$ dass er nett
$Fritz_{nom}$ has claimed $Maria_{nom}$ would think   that he nice
ist]]
is

d. $[_{CP_2}$ Dass er nett ist] hat Fritz behauptet $[_{CP_1}$ würde Maria denken t_2]

Similarly, VP-over-VP strucures permit extraction of the lower VP in German, in violation of the A-over-A Principle. This is shown in (50), which involves a restructuring infinitive. Assuming restructuring infinitives to be VPs (see Stechow & Sternefeld (1988), Wurmbrand (2001)), (50-b) must involve VP topicalization from VP. (In (50-c), the higher VP undergoes movement, which is also an option.)

(50) a. Fritz   hat $[_{VP_1}$ $[_{VP_2}$ zu arbeiten] versucht]
$Fritz_{nom}$ has   to work   tried

b. $[_{VP_2}$ Zu arbeiten] hat Fritz $[_{VP_1}$ t_2 versucht]
to work   has $Fritz_{nom}$ tried

c. $[_{VP_1}$ $[_{VP_2}$ Zu arbeiten] versucht] hat Fritz t_1
to work   tried   has $Fritz_{nom}$

Interestingly, such a VP that is extracted from a VP can also undergo long-distance topicalization in German. This implies that two (or more) higher VPs can be crossed; violations of the A-over-A Principle can be iterated. Thus, in (51-b), there is in fact a double violation of the A-over-A Principle (with VP_2 moving across VP_1 and VP_0);

c. $[_{DP_2}$ Wo] spielt sie $[_{PP_1}$ t_2 mit] ?
where plays she   with

in (51-c), there is a single violation (with VP_1 only crossing VP_0).[25]

(51) a. Ich [$_{V_3}$ denke] nicht [$_{VP_0}$ t_3 [$_{CP}$ dass er [$_{VP_1}$ [$_{VP_2}$ zu arbeiten] versucht]
          I   think   not   that he   to work   tried
          hat]]
          has

    b. [$_{VP_2}$ Zu arbeiten] denke$_4$ ich nicht [$_{VP_0}$ t_4 [$_{CP}$ dass er [$_{VP_1}$ t_2 versucht]
            to work   think I   now   that he   tried
        hat]]
        has

    c. [$_{VP_1}$ [$_{VP_2}$ Zu arbeiten] versucht] denke$_4$ ich nicht [$_{VP_0}$ t_4 [$_{CP}$ dass er t_1
            to work   tried   think I   not   that he
        hat]]
        has

In view of the evidence that the A-over-A Principle is too strong, there are only two strategies that can be pursued. The first possibility is to give up the constraint, and replace it with some other constraint(s). This is the position taken by Ross (1967). The second possibility is to revise the constraint in such a way that the counter-examples lose their force (while, ideally, supporting evidence can be maintained). A revision of the A-over-A Principle is argued for by Bresnan (1976) (on the basis of earlier work in Chomsky (1973); also see footnote 25).[26]

---

[25] It should be pointed out that long-distance VP topicalization is not judged as fully felicitous by all speakers of German; for many speakers, data of this kind sound clumsy. The remarks in the main text are based on the varieties of German that permit this construction. Also note that (51-c) is a special case in that the two items competing under the A-over-A Principle are initially very far apart. Still, under the formulation of the A-over-A Principle in (43), this does not change things because the distance of the two A-categories is immaterial. What is more, (51-c) in fact continues to instantiate a violation of the A-over-A Principle if it is assumed that the constraint holds only within certain ('cyclic') domains, as assumed in Chomsky (1973, 235); this latter version of the A-over-A Principle is given in (i).

(i) *A-over-A Principle* (Chomsky's (1973) version):
     If a transformation applies to a structure of the form

     [$_\alpha$ ... [$_A$ ...] ...]

     where α is a cyclic node, then it must be so interpreted as to apply to the maximal phrase of the type A.

Again, the initial distance between the two A-items is irrelevant.

[26] Chomsky (1973, 235) remarks that the version of the A-over-A Principle in (i) of footnote 25 "does not establish an absolute prohibition against transformations that extract a phrase of type A from a more inclusive phrase of type A. Rather, it states that if a transformational rule is nonspecific with respect to the configuration defined, it will be interpreted in such a way as to satisfy the condition." This understanding of (i) of footnote 25 captures fundamental aspects of the revision to be discussed momentarily in the main text. However, it does not necessarily follow from the formulation in (i), at least not as long as it is not

Bresnan (1976) develops an elaborate theory of transformations; on this basis, she suggests a revision of the A-over-A Principle that evades the counter-evidence given above (but, as we will see, this comes at the price of losing some of the original evidence in favour of the A-over-A Principle). The basic idea is that the A-over-A Principle should not be sensitive to category information; rather, it is sensitive to the nature of the transformation involved. The central claim is that the item that can be subjected to a given transformation should be as inclusive as possible, independently of category labels (whereas an even more inclusive item that cannot be subjected to the same transformation because it does not fit its structural description can be ignored, even if it shares the same category label). Consider the following revision of the A-over-A Principle taken from Bresnan (1976, 14).[27]

(52) *A-over-A Principle* (Bresnan's revision):
      No transformation T can apply to a structure ϕ under a proper analysis π unless π is a maximal proper analysis of ϕ for T.

This version of the principle relies on the notion of "maximal proper analysis"; it is defined in (53).

(53) *Maximal proper analysis*:
      π is a maximal proper analysis of ϕ for T = $<\mathcal{C},\mathcal{M}>$ iff π assigns maximal values to all target predicates in $\mathcal{C}$.

Abstracting away from much of the technical machinery employed by Bresnan here and in what follows, let me clarify the key notions involved here. A transformation T is an ordered pair $<\mathcal{C},\mathcal{M}>$, where $\mathcal{C}$ is a structural condition consisting of a number of terms, and $\mathcal{M}$ is a transformational mapping affecting these terms. The *wh*-movement transformation, e.g., can be encoded as the $<\mathcal{C},\mathcal{M}>$ pair in (54) (I tacitly adapt parts of Bresnan's original formulation to more current terminology).

(54) *Wh-movement transformation*:
      [$_{CP}$ Q   – W$_1$ – [$_{XP}$ *wh* – W$_2$] – W$_3$]
               1   2   3   4   5 $\Rightarrow$
           ⌢
           3 4   2   Ø   5

---

clarified what exactly it means for a transformation to "apply to a phrase of type A". Indeed, as Chomsky notes, "alternatively, one might interpret the A-over-A constraint as legislating against any rule that extracts a phrase of type A from a more inclusive phrase A," which is the understanding embodied in (43). He concludes that "the former interpretation [...] is perhaps more natural", and therefore "tentatively" adopts it in the remainder of the study.

[27] Bresnan takes this to be essentially a more formal implementation of Chomsky's (1973) basic idea; see the last two footnotes.

Here, Q represents the specifier of an interrogative CP. The transformation moves the *wh*-phrase from its in situ position to Q, as an instance of substitution, and carries out deletion in the base position of the *wh*-phrase. In (54), there are both *variables* (W_1–W_3) and *predicates*; the former stand for any sequence of terms whereas the latter encode restrictions for the transformation – they define the structural condition $\mathcal{C}$. We can distinguish between two types of predicates: First, *target predicates* are the items that the transformation applies to (the item that is moved, in the case at hand). Second, *context predicates* are the non-target predicates of a structural condition (i.e., items that must be present in a phrase structure for the transformation to be able to apply, but which are not directly affected by the operation). Next, a *proper analysis* of a syntactic structure ϕ for a transformation T = $<\mathcal{C},\mathcal{M}>$ is a factorization $<\phi_1,...\phi_n>$ of ϕ (simplifying, a complete partition of the whole initial phrase marker into subtrees) of which $\mathcal{C}$ is true and upon which $\mathcal{M}$ is defined. Finally, a *maximal proper analysis* as in (53) is a proper analysis of a phrase marker that "assigns maximal values to all target predicates in $\mathcal{C}$"; and the value of a target predicate is maximal if it represents the most inclusive category (i.e., the one that dominates all others) among those that satisfy the requirements imposed by $\mathcal{C}$ on the transformation. This, then, is what the revised A-over-A Principle in (52) amounts to: A transformation can only apply to a given structure if the size of the target predicate is maximal vis-à-vis the demands imposed by the structural condition $\mathcal{C}$ of the transformation.

Before going through an example illustrating the working of this version of the A-over-A Principle, it is worth emphasizing that there is a significant shift of perspective: Category information is at the heart of the original A-over-A Principle, but it does not play any role whatsoever here. Two (or more) categories compete for movement not if they have identical labels, but if they both satisfy requirements imposed by the structural condition of a transformation on the form of the target predicate. Bresnan (1976, 21ff) highlights this new property of the A-over-A Principle by focussing on "cross-categorial" transformations, i.e., transformations that are not category-specific (like, arguably, movement to subject position in English is). *Wh*-movement as in (54) is just such a transformation; see (55).

(55) a.   What book did you read ?
      b.   How long is it ?
      c.   How quickly did you read it ?
      d.   How much did it cost ?

Consider now the data in (56) (from Bresnan (1976, 22), with (56-de) added), which are structurally similar to the ones in (46), and which therefore pose a problem for the original version of the A-over-A Principle (in that they show that the constraint is too

weak).[28]

(56) a.   How many feet tall does the girl stand ?
     b.   *How many feet does the girl stand tall ?
     c.   *How many does the girl stand feet tall ?
     d.   *How does the girl stand many feet tall ?
     e.   How many feet are in a yard ?

Bresnan assumes that the pre-movement structure of *how many feet tall* in (56-abc) is roughly as in (57).

(57)   Q the girl [$_{VP}$ stands [$_{AP}$ [$_{NP}$ [$_{QP}$ [$_{DP}$ how] [$_Q$ many]] [$_N$ feet]] [$_A$ tall]]]

The first thing to note is that a proper analysis of the structure in (57) for the *wh*-movement transformation may involve a factorization in which the DP *how* is the moved item;[29] or a factorization in which the QP *how many* is the moved item; or a factorization in which the NP *how many feet* is the moved item; or a factorization in which the AP *how many feet tall* is the moved item. Under all these factorizations, the structural condition of (54) is met: The moved item is an XP with a *wh*-item at its left edge. At this point, the A-over-A Principle becomes relevant and blocks (56-bcd), leaving (56-a) as the sole remaining option. While all the partitions underlying (56-abcd) involve a *proper analysis* of the structure for the *wh*-movement transformation, only (56-a) involves a *maximal proper analysis*; the reason is that only in this latter case is the target predicate's value maximized (i.e., the AP is the most inclusive category that can be moved in accordance with the structural condition of (54)). An even more inclusive category would be the VP that minimally dominates AP in (57). However, movement of the VP *stands how many feet tall* does not meet

---

[28] Parallel examples from German show a slightly different pattern, with the construction that is analogous to (56-b) emerging as the preferred option; see Heck (2008) and (i).

(i) a.   Wieviele   Zentimeter ist sie groß ?
          how many centimeters is she tall
     b.   ?Wieviele   Zentimeter groß ist sie ?
          how many centimeters tall   is she

I will have nothing insightful to say about this apparent instance of parametrization here; but the present logic might lead one to hypothesize that in German, an analysis of the pre-movement string is available according to which *wieviele Zentimeter* ('how many centimeters') emerges as the most inclusive item.

[29] Bresnan does not actually envisage this as a possibility because she assumes *how* to be a non-maximal D item in QP that therefore does not fit the structural condition (which demands an XP in which *wh* is initial). This issue is orthogonal to the current argument, though: On one view (the one pursued in the main text), (56-d) is blocked by the revised A-over-A Principle; on the other view (Bresnan's), (56-d) cannot be generated because the structural condition is not respected.

the requirements of $\mathcal{C}$ of (54) according to which the *wh*-item must be leftmost in the moved phrase. (In a nutshell, this derivation fails because the VP is not a *wh*-phrase.) As noted by Bresnan (1976), there is nothing inherently wrong with moving a category like the NP in (57). If there is no more inclusive XP that could act as a *wh*-phrase, as is the case in (56-e), movement of such an NP is possible.

This general reasoning is quite independent of the exact structure attributed to categories in the adjectival and nominal domain, and of the labelling of the categories (although details of the analysis may differ under other assumptions). Suppose, e.g., that we were to assume (as one of several further options for analysis that have been proposed) (58) instead of (57) as the structure that feeds *wh*-movement.

(58)   Q the girl [$_{VP}$ stands [$_{AP}$ [$_{DP}$ [$_{DegP}$ [$_{Deg}$ how] [$_{QP}$ many]] [$_{D'}$ [$_D$ Ø] [$_{NP}$ feet]]]
        [$_A$ tall]]]

In this case, Bresnan's A-over-A Principle in (52) would imply that AP movement is possible whereas DP and DegP movement are not; movement of *how* alone would be excluded by the requirement that it has to be an XP that undergoes the movement.

Exactly the same kind of analysis can be given for the examples in (46), repeated here in (59) (without commitments to structural analysis).

(59)   a.   You have a very intelligent sister
        b.   How intelligent a sister do you have ?
        c.   *How intelligent do you have a sister ?

Whatever the internal structure of the nominal projection here is supposed to look like, it is clear that movement of the more inclusive *wh*-category in (59-b) respects Bresnan's A-over-A Principle whereas movement of the less inclusive *wh*-category in (59-c) violates it.

However, not all of the the evidence brought forward as part of the argument that the original A-over-A Principle is too weak can be derived under Bresnan's reformulation. For instance, the ban on preposition stranding in German (see (47)) does not follow from (52) either (at least not unless further assumptions are made).

What about the evidence *in support* of the original A-over-A Principle (see (44), (45))? The ill-formed examples here involve illicit *wh*-movement (or relativization) of a *wh*-item (or a relative pronoun) where the more inclusive category has the same category feature but, crucially, cannot act as a *wh*-item (or relative item) itself. These data can therefore not be derived from Bresnan's A-over-A Principle.

Finally, what about the evidence that proved that the original A-over-A Principle is too strong (see (48), (49), (50), and (51))? First, (48) (where movement of a *wh*-DP takes place from a more inclusive non-*wh*-DP, and the result is well formed) ceases to be a problem for exactly the reason that (44), (45) cannot be derived anymore as ungrammatical: Categorial information as such is irrelevant, and movement of a lower

wh-item cannot be blocked by a higher non-*wh*-item. If one compares the ungrammatical examples in, say, (44) with the grammatical examples in, e.g., (48), it is indeed hard to see how any version of the A-over-A Principle could make the right distinctions. The evident difference is one between ill-formed extraction from subject and well-formed extraction from object, but this is a domain that the A-over-A Principle has nothing to say about, in any version. Hence, in light of the wellformedness of data like those in (48), the fact that the revised A-over-A Principle is not capable of excluding data like those in (44) should not be viewed as a shortcoming; the data suggest that some other locality constraint that can distinguish between subjects and objects will account for the difference.

Second, the data in (49) have CP topicalization from a CP. Assuming (as seems natural) that the lower CP has an abstract feature [top] that is required by the structural description $\mathcal{C}$ of an appropriately defined topicalization transformation, and that the higher CP does not have such a feature, the two items do not interact, and Bresnan's A-over-A Principle makes the right predictions; again, categorial identity emerges as irrelevant.

Third, more or less the same analysis can be given for the VP-over-VP contexts that permit topicalization of the lower VP in (50) and (51): The lower VP that undergoes the movement differs from the higher VP (or VPs) in that it bears the [top] feature required by the structural condition of the topicalization transformation.

The version of the A-over-A Principle in (52) meets the demands of a good constraint (as does the original version in (43)): Abstracting away from the technical aspects of Bresnan's analysis, it is simple and general, and it is not complex.[30]

3.2   The F-over-F Principle

From the perspective of current linguistic theory, Bresnan's (1976) revision of the original A-over-A Principle looks very modern. In a theory where movement transformations involve designated features, her approach can straightforwardly be rein-

---

[30] The formulation in (52) is not the definitive version that Bresnan (1976) ends up with. Rather, eventually she suggests a "Relativized A-over-A Principle" in which the notion of a "maximal proper analysis" is replaced with the weaker notion of an "r-maximal proper analysis" ("weaker" in the sense that there are more r-maximal proper analyses than there are maximal proper analyses). The basic idea is that the size of target predicates is not maximized as such; target predicates are only maximized according to a given choice of context predicates. The motivation for this change is based on issues like the increased options of affecting target predicates in deletion constructions (like VP ellipsis); I will abstract away from the modifications of theoretical concepts that these kinds of construnctions may require in what follows, and consequently ignore the minor revision of the A-over-A Principle in (52) that are motivated by it. – Incidentally, as explicitly noted by Bresnan (1976, 16), the version of the A-over-A Principle in (52) already is a "relativized" constraint – relativized with respect to possible items that can be affected by transformations (rather than being rigidly defined in terms of categorial labels).

terpreted as requiring movement of the closest item that bears a feature matching the attracting feature on the head of the landing site. From this perspective, (52) amounts to stating that if there is a head X with a movement-inducing feature $\mathcal{F}$, of all the items $Y_1,...,Y_n$ that are in the c-command domain of X and bear a matching feature $\mathcal{F}$, only the most inclusive Y_i (i.e., the Y_i that dominates all other Y_j's) can be moved to the specifier of X. This reformulation of Bresnan's (52) is a constraint that has in fact more recently been proposed by a number of researchers, as an alternative to the original A-over-A Principle (and, it may be noted, without reference to Bresnan (1976) throughout). Thus, assuming that certain designated features are responsible for triggering various movement operations ([•wh•] for *wh*-movement, [•top•] for top-icalization, etc.), it is clear that these features are not (necessarily) categorial (they are "cross-categorial", in Bresnan's terms).[31] On this basis, it has been proposed that the original A-over-A Principle should be revised as in (60); here and henceforth, I will refer to this minimality-based revision as the *F-over-F Principle*. See Takano (1994), Koizumi (1995), Fukui (1997), Kitahara (1997), Müller (1998), Sauerland (1999), Fitzpatrick (2002), Vicente (2007), and Heck (2008), among others.

(60) *F-over-F Principle*:

In a structure $\alpha_{[•F•]}... [_{\beta_{[F]}} ... [_{\gamma_{[F]}} ...] ...] ...$, movement to [•F•] can only affect the category bearing the [F] feature that is closer to [•F•].

(60) is very similar to (52). There is a difference with respect to the empirical domains that are covered by the analyses employing these two kinds of constraints, though. Bresnan only concerns herself with initial ambiguities in rule application that arise from the fact that more than one category can in principle represent the moved item for one and the same feature. Essentially, all her data center around the question of whether *pied piping* is possible (and the formulation of (54) suggests that leftmost items can pied-pipe other material whereas non-leftmost items cannot).[32] In contrast, in the approaches that rely on (some version of) (60), the empirical domain typically does not involve pied piping but rather two separate *wh*-phrases, one of which domi-nates the other in the pre-movement structure.

Empirical evidence in support of a constraint like the F-over-F Principle comes from the consideration of a restriction on the movement of remnant categories (i.e., categories from which extraction has taken place). Among other things, the constraint blocks certain illicit instances of remnant scrambling in languages like German and

---

[31] For the moment, the • notation for features that trigger operations can be ignored. I will come back to this in chapter 3.

[32] See Heck (2004; 2008) and Cable (2007) (and references cited in these works) for recent approaches to the phenomenon in general, and for the status of this generalization in particular.

Japanese (see Takano (1994), Koizumi (1995), Kitahara (1997), Müller (1998), Sauer-land (1999)). Consider first some data from German. (61-ab) show that a restructuring infinitive from which scrambling has taken place (i.e., a remnant infinitive) can be top-icalized.[33]

(61) a. $[_\alpha$ t_1 Zu lesen $]_3$ hat $[_{DP}$ das Buch $]_1$ keiner t_3 versucht
              to read   has   the book$_{acc}$ no-one$_{nom}$ tried

      b. $[_\alpha$ t_1 t_2 Zu reparieren $]_3$ hat der Frank   dem Matthias$_1$ den Drucker$_2$
              to fix   has the Frank$_{nom}$ the Matthias$_{dat}$ the printer$_{acc}$
          t_3 versprochen
            promised

Next, (62-ab) illustrate that the same remnant infinitive from which scrambling has taken place cannot undergo scrambling itself (scrambling of a non-remnant infinitive is possible, though).

(62) a. *dass $[_\alpha$ t_1 zu lesen $]_3$ $[_{DP}$ das Buch $]_1$ keiner t_3 versucht hat
              that   to read   the book$_{acc}$ no-one$_{nom}$ tried   has

      b. *dass der Frank $[_\alpha$ t_1 t_2 zu reparieren $]_3$ dem Matthias$_1$ den
              that the Frank$_{nom}$ to fix   the Matthias$_{dat}$ the
          Drucker$_2$ t_3 versprochen hat
          printer$_{acc}$ promised   has

The same restriction is illustrated on the basis of data from Japanese in (63). In (63-a), a complement CP has undergone scrambling in front of the subject; (63-b) shows that such scrambling is impossible if DP scrambling has taken place from the complement CP prior to CP fronting.[34]

(63) a. $[_{CP}$ Mary-ga $[_{DP}$ sono hon-o $]_1$ yonda-to $]_2$ Bill-ga $[_{CP}$ John-ga t_2
              Mary$_{nom}$ that book$_{acc}$ read-COMP   Bill$_{nom}$ John$_{nom}$
          itta-to   ] omotteiru (koto)
          said-COMP   think   fact

      b. *$[_{CP}$ Mary-ga t_1 yonda-to $]_2$ $[_{DP}$ sono hon-o $]_1$ John-ga t_2 itta
              Mary$_{nom}$ read-COMP   that book$_{acc}$ John$_{nom}$ said
          (koto)
           fact

This anti-identity restriction on remnant movement (which is referred to as an *Unam-biguous Domination* requirement in Müller (1998)) is not confined to scrambling. As

---

[33] At this point, it does not really matter what category α is. For concreteness, one might assume (following Stechow & Sternefeld (1988) and Wurmbrand (2001), among many others) that it is VP; but other category labels would not affect the analysis.

[34] Scrambling of an (object) DP from a CP as such is unproblematic.

illustrated in (64), the same effect shows up with *wh*-movement in English: Extraction of a *wh*-phrase from a *wh*-phrase makes subsequent *wh*-movement impossible; the effect is much stronger than one would expect if only a typical *wh*-island violation with *wh*-argument extraction were involved.

(64) a. *[$_{DP}$ Which book about t_1]$_2$ don't you know [$_{CP}$ who$_1$ to read t_2] ?
      b. *[$_{DP}$ Which picture of t_1]$_2$ do you wonder [$_{CP}$ who$_1$ she likes t_2] ?

In (62), (63), and (64), there is a stage of the derivation where a movement-inducing feature (like [•Σ•] for scrambling, and [•wh•] for *wh*-movement) on a target head could in principle attract either the more inclusive category or another category dominated by the latter (both bearing, by assumption, a matching feature [Σ] or [wh]). The F-over-F Principle then demands movement of the higher category (β in (60)) first (which is thus not yet a remnant category at this point because extraction from it has not yet taken place), and subsequent movement of a category (α in (60)) from within β to a lower position will invariably be blocked by the Strict Cycle Condition (see (23)), which (among other things) excludes cases of lowering.[35]

I will return to the status of the F-over-F Principle below. For the time being, I will leave it at that, and turn to other predecessors of current locality constraints, viz., the island constraints proposed by Ross (1967), at least partly in reaction to Chomsky's A-over-A Principle.

3.3 The Complex NP Constraint

Ross (1967) does not pursue the same kind of approach as Bresnan (1976) in view of the counter-evidence against the original A-over-A Principle in (43); rather than attempting to revise the original A-over-A Principle, he concludes that the constraint should be abandoned, and replaced by a set of empirically more adequate constraints. One constraint that he suggests as a partial replacement for the A-over-A Principle is the Complex NP Constraint (CNPC) in (65).[36]

(65) *Complex NP Constraint* (CNPC):
      No element contained in a CP dominated by a DP may be moved out of that DP.

The Complex NP Constraint accounts for some of the data that motivated the A-over-A Principle. Ungrammatical examples like those in (45), for instance, now follow

---

[35] Such a derivation will also violate the Condition on Extraction Domain discussed below.

[36] It was a standard assumption until the late eighties that NP dominates DP, not DP NP, as assumed here (and in most current work). Hence, the original Complex NP Constraint is a constraint on movement from NP, not from DP. The constraint is still known under its original name, which is therefore also adopted here, even though "Complex DP Constraint" might be more appropriate.

from the CNPC; see, e.g., (66-ab), where a DP is *wh*-moved from a relative clause CP that accompanies the noun of the next higher DP.

(66)   a.   *[$_{DP_1}$ Which book] did John meet [$_{DP_2}$ a child [$_{CP}$ who read t_1]] ?
       b.   *[$_{DP_1}$ Who] does Mary know [$_{DP_2}$ a girl [$_{CP}$ who is jealous of t_1]] ?

Similarly, *wh*-extraction from a complement CP of a noun is impossible, and again, this follows from the CNPC. Incidentally, whereas the illformedness of (67-a) and (67-c) could also be derived from the original A-over-A Principle (see (45-b)), (67-b) would not follow under either the original version of the A-over-A Principle (unless one assumes *how* to be a DP; see (57) vs. (58) above) or Bresnan's revision (there is only one *wh*-phrase around in (67-b)).[37,38]

(67)   a.   *[$_{DP_1}$ Which book] did John hear [$_{DP_2}$ a rumour [$_{CP}$ that you had read t_1]] ?
       b.   *[$_{XP_1}$ How] did John hear [$_{DP_2}$ a rumour [$_{CP}$ that you had fixed the car t_1]] ?
       c.   *The hat [$_{DP_1}$ which] I believed [$_{DP_2}$ the claim [$_{CP}$ that Otto was wearing t_1]] is red

Is the CNPC a good constraint in the sense of (39) (simplicity, generality) and (41) (non-complexity)? The answer has to be no. Just like the Subjacency Condition, the CNPC is category-specific (since it mentions the labels D and C); and just like the Subjacency Condition, it is a two-node constraint that requires keeping track of a node crossed earlier.[39] What is more, the two-node property is also responsible for the fact that the CNPC is faced with severe empirical counter-evidence. According to (65), extraction from CP to a position *above* DP is predicted to be impossible, but extraction from CP to a position *below* DP should be fine. This is not the case. To

---

[37] Movement from argument clauses (i.e., subcategorized categories) in complex DPs typically yields better results than movement from relative clauses (i.e., non-selected, modifier categories): Data of the type in (67) are typically judged to be somewhat better than data of the type in (66). However, this does not hold for movement of modifiers, which is completely impossible throughout; i.e., sentences like (67-a) are usually perceived as being less ill formed than sentences of the type in (67-b)). For the time being, I will ignore these differences; but I will come back to the question of how they might be accounted for in chapter 5.

[38] As it stands, the CNPC in (65) is a local derivational constraint on movement operations. It can also be reinterpreted as a local representational constraint, as in (i).

(i)   *Complex NP Constraint* (representational version):
      *... α_1 ... [$_{DP}$ ... [$_{CP}$ ... t_1 ...]] ...

As before, such a representational version must involve a concept like that of trace, which encodes the base position from which movement takes place.

[39] The convergence of the Subjacency Condition and the CNPC does not really come as a surprise in this respect because the Subjacency Condition was developed as a (somewhat more general) replacement for the CNPC.

see this, we can consider constructions involving PP movement to specifiers of DP in German. Relevant examples are given in (68).[40]

(68) a. [$_{DP_2}$ [$_{PP_1}$ Von   Peter] [$_{D'}$ das Gerücht t_1]] habe ich t_2 gehört
                        by/of Peter   the rumour   have I   heard

       b. Fritz hat [$_{DP}$ das Gerücht [$_{CP}$ dass Maria [$_{PP_1}$ von Peter] gestern   ein
           Fritz has   the rumour   that Maria   by Peter   yesterday a
           Buch gekriegt hat] gehört
           book given   was heard

       c. *[$_{PP_1}$ Von Peter] hat Fritz [$_{DP}$ das Gerücht [$_{CP}$ dass Maria t_1 gestern   ein
                 by Peter has Fritz   the rumour   that Maria   yesterday a
           Buch gekriegt hat] gehört
           book given   was heard

       d. *Fritz hat [$_{DP}$ [$_{PP_1}$ von Peter] [$_{D'}$ das Gerücht [$_{CP}$ dass Maria t_1 gestern
           Fritz has   by Peter   the rumour   that Maria   yesterday
           ein Buch gekriegt hat]] gehört
           a   book given   was heard

(68-a) exhibits the basic construction: A PP (*von Peter*, 'by/of Peter') can undergo DP-internal movement to a position in front of the determiner (*das* ('that')). (The PP may be interpreted as the source or, preferably, as the theme.) The available tests for constituency show unequivocally that the fronted PP has not left the DP.[41] In (68-b), the same kind of PP *von Peter* is merged outside of a DP, as an optional argument of a CP embedded in a DP. The CNPC now predicts that such a PP cannot undergo movement to a position outside of the complex DP; this prediction is borne out (see (68-c), which is ungrammatical under the relevant reading where PP$_1$ originates in the embedded clause). However, the CNPC does not rule out the possibility that PP movement in this context ends up in the specifier of the complex DP, as in (68-d). Still, this sentence is ungrammatical: There is no way that the PP *von Peter* can be interpreted as belonging to the embedded CP (with an interpretation as the source of the book).[42] Given that the PP *von Peter*, the determiner *das* ('das') and the noun

---

[40] See Lindauer (1995) for extensive discussion. The construction belongs to the realm of colloquial (in fact, slightly substandard) speech, and it typically requires certain intonation patterns. As Lindauer shows, its distribution is also restricted in a number of further ways. The grammaticality contrasts in (68) are crystal-clear, though.

[41] For instance, in (68-a), the string *von Peter das Gerücht* occupies the position in front of the finite verb in a verb-second clause. With very few potential exceptions (that do not resemble the construction in (68-a)), only one constituent can occupy this position. See Müller, St. (2005).

[42] An interpretation as the source of the rumour (i.e., with base-generation of the PP outside of the CP) is possible in principle (though the sentence sounds clumsy, presumably not least of all because an expected

Gerücht ('rumour') can co-occur as such, and in this order (as shown by the well-formedness of (68-a)), it seems clear that the illformedness of (68-d) is due to the fact that illegitimate extraction from a CP has taken place: The correct generalization is that extraction from CPs in DPs is blocked in general, even if movement does not go beyond DP. Thus, some locality constraint on extraction from CP is evidently violated in (68-d), and it cannot possibly be the CNPC. Clearly, the same constraint will also suffice to cover the standard CNPC effects, as in (66) and (67). The existence of this other constraint (whatever it ultimately amounts to; see below) invariably makes the CNPC redundant.[43]

3.4 The Sentential Subject Constraint

Another constraint proposed by Ross (1967) is the Sentential Subject Constraint; see (69).

(69) *Sentential Subject Constraint*:
No element dominated by a CP may be moved out of that CP if that CP is a subject.

Data that motivate the constraint are given in (70). In (70-a), extraction takes place from an object CP, and the result is well formed. In contrast, in (70-bc), extraction takes place from a subject CP, and the results are ill formed.[44]

(70) a. $[_{DP_1}$ Who] did the reporters expect $[_{CP}$ that the principal would fire t_1] ?
     b. *$[_{DP_1}$ Who] was $[_{CP}$ that the principal would fire t_1] expected by the reporters ?
     c. *$[_{DP_1}$ Who] did $[_{CP}$ that Mary was going out with t_1] bother you ?

3.5 The Subject Condition

The Sentential Subject Constraint can be generalized: DP subjects are also islands, even if they do not qualify as complex in the sense of the Complex NP Constraint. This is illustrated by the data in (71).

---

argument in the embedded CP must then be absent, a garden path effect).

[43] Note that the illegitimacy of movement from a CP in a DP to a position outside CP, but inside DP, has the same kind of (fatal) repercussions for the Subjacency Condition (Chomsky (1977); see above) as it has for the CNPC.

[44] The Sentential Subject Constraint in (69) is a derivational constraint, but it can be reformulated as a representational constraint; see (i).

(i) *Sentential Subject Constraint* (representational version):
*... α_1 ... $[_{CP}$ ... t_1 ...] ... if CP is a subject.

(71)   a.   *[DP$_2$ Who(m)] has [DP$_1$ a comment about t$_2$] annoyed you ?
    b.   *[PP$_3$ About whom] has [DP$_1$ a comment t$_3$] annoyed you ?

The more general version of the constraint that covers both Sentential Subject Condition effects as in (70) and the data in (71) has become known as the Subject Condition (see Chomsky (1973; 1986a), Huang (1982), and Freidin (1992), among many others).[45]

(72)   *Subject Condition*:
    No element may be moved out of a subject.

At this point, recall that the original (rigid) A-over-A Principle could derive data like those in (44) (which involve extraction from a subject, just like (71) does), but only at the price of wrongly also excluding examples like those in (48) (with extraction from an object). Bresnan's (1976) version of the A-over-A Principle and the F-over-F Principle are both compatible with extraction from an object as in (48), but they fail to exclude extraction from a subject as in (44). This gap can be closed by the Subject Condition.

The Subject Condition (like the Sentential Subject Constraint in the formulation in (69)) remains unspecific about exactly what property an item must have to count as a "subject" in the sense of (72). Options include certain designated positions (e.g., being in Specv, or in SpecT), or certain argument-structural properties (e.g., being the last-merged item of a subcategorization list, or being introduced into the structure by v). I will not try to clarify the subtleties that arise here at this point. However, it would seem to be clear that the fact that the constraint mentions "subject" (in whatever form) implies that it does not qualify as a good constraint (given (39)): It is construction-specific, and thus not general enough.

3.6   The Wh-Island Condition

The *Wh*-Island Condition is a derivational constraint introduced in Chomsky (1973). Simplifying a bit (and adapting it to current assumptions about clausal structure), it can be formulated as in (73).[46]

---

[45] The Subject Condition in (72) is a derivational constraint. An equivalent representational version is given in (i).

(i)   *Subject Condition* (representational version):
    *... α_1 ... [$_\beta$ ... t$_1$ ...] ... if β is a subject.

[46] Here is a representational version:

(i)   *Wh-Island Condition*:
    *... α_1 ... [CP β_2 ... t$_1$ ...] ..., where β is a *wh*-element in SpecC or C.

(73)   *Wh-Island Condition*:
       Movement must not cross a CP with a *wh*-element in SpecC or C.

A consequence of the *wh*-island condition is that movement from an embedded *wh*-clause results in ungrammaticality. It has often been observed that the interaction of various factors determines the degree of illformedness. Extraction of a non-argument XP from a *wh*-island typically gives rise to strong deviance (see (74-a) vs. (74-b)); extraction of an argument (in particular: an object) XP from a *wh*-island often yields somewhat better results (see (74-c) vs. (74-d), and contrast (74-b) with (74-d)).

(74)   a. How_1 do you think [$_{CP}$ that Mary solved the problem t_1] ?
       b.   *How_1 do you wonder [$_{CP}$ whether Mary solved the problem t_1] ?
       c.   [$_{DP_1}$ Which book] do you think [$_{CP}$ that John read t_1] ?
       d. ?*[$_{DP_1}$ Which book] do you wonder [$_{CP}$ [$_{PP_2}$ to whom] John gave t_1 t_2] ?

Furthermore, *wh*-island effects are generally ameliorated if the *wh*-clause is an infinitive and the moved item is a complement DP (see Chomsky (1986a) and Frampton (1990), among others). Thus, compare (75) with (74-d).

(75)   ??[$_{DP_1}$ Which book] don't you know [$_{CP}$ whether to read t_1] ?

An effect that is similar to the *wh*-island effect arises with topicalization. A fronted topic in an embedded clause blocks extraction from this clause (of any type, i.e., there is no difference between, say, argument and adjunct extraction, or between topicalization and *wh*-movement from a clause in which topicalization has occurred). This is shown by the data in (76), for topicalization in English and for *wh*-movement in German.

(76)   a.   [$_{DP_1}$ This book] Mary thinks that Bill gave t_1 [$_{PP_2}$ to John]
       b.   *[$_{DP_1}$ This book] Mary thinks that [$_{PP_2}$ to John] Bill gave t_1 t_2
       c.(?)[$_{DP_1}$ Wen] denkst du   [$_{CP}$ dass Maria t_1 mag] ?
            whom think   you   that Maria   likes
       d. *[$_{DP_1}$ Wen] denkst du   [$_{CP}$ $Maria_2$ mag_3 t_2 t_1 t_3] ?
            whom think   you   Maria likes

There is some disagreement as to whether topicalization in languages like English or German targets a specifier of C, or a specifier of some other functional category (perhaps a functional catgory Top, as suggested in Müller & Sternefeld (1993), Rizzi (1997), Bayer & Salzmann (2009), and much recent work), or, as has also been pro-

posed, a specific adjunction position in the clausal domain.[47] For concreteness, suppose the first option is correct: Topicalization is movement to SpecC (in languages like English, Danish, and Icelandic, embedded topicalization then will have to involve CP recursion, given the presence of a complementizer above the topic). If so, the *topic island* effect visible in (76-bd) can be given basically the same analysis as the *wh*-island phenomenon. In line with this, a *Topic Island Condition* has variously been suggested; and further generalization that subsumes the *Wh*-Island Condition and the Topic Island Condition under a single constraint seems possible. Such a generalized A-bar intervener (or "operator island") constraint might then also contribute to improving the constraint with respect to the requirement (39) – as they stand, both *Wh*-Island Condition and the Topic Island Condition are construction-specific (hence, they do not qualify as general constraints).

3.7   The Superiority Condition

The Superiority Condition has first been suggested in Chomsky (1973, 246). The original formulation of this derivational constraint is as follows.

(77)   *Superiority Condition*:
        No rule can involve, X, Y in the structure
        ... X ... [$_\alpha$ ... Z ... -WYZ ...] ...
        where the rule applies ambiguously to Z and Y and Z is superior to Y.

For present purposes, we can equate "α is superior to β" with "α asymmetrically c-commands β.[48] Focussing just on displacement, X stands for the landing site of movement in (77); and Z and Y are two items that can in principle be affected by the operation (since, by assumption, they satisfy the structural condition of the move-

---

[47] Topicalization goes hand in hand with verb second movement in German, which would seem to exclude the adjunction analysis for this language. See, however, Frey (2000; 2004a) for arguments that German might also have a kind of embedded topicalization to a position following the complementizer, which does not trigger verb second movement.

[48] This is how the concept of superiority has come to be understood. Chomsky's original notion of superiority is in fact more general; see (i) from Chomsky (1973, 246):

(i)   A catgory A is superior to a category B in the phrase marker if every major category dominating A dominates B as well but not conversely.

Taken for itself, this would also subsume typical A-over-A configurations (in Bresnan's sense, which, as noted above, is already envisaged as a possibility in Chomsky (1973)), where a category α dominates a category β: In this case, α should be superior to β, and this might then also be relevant for the Superiority Condition. On this view, the Superiority Condition in (77) might in fact emerge as extremely modern, and might very much look like the (G)MLC that will be addressed below. However, the specific structural placement of Z and Y in (77) excludes such a view (Z and Y cannot stand in a domination relation here), which justifies the interpretation of superiority adopted in the main text.

ment transformation). In this case, (77) states that only the higher, asymmetrically c-commanding item can undergo the operation.

Some consequences of the Superiority Condition are illustrated in (78). If there are two *wh*-phrases wh_1 and wh_2, the *wh*-movement transformation "applies ambiguously" to wh_1 and wh_2 (as long as no other constraints, like the A-over-A Principle, are violated). In this case, the structurally higher *wh*-phrase (i.e., the one that asymmetrically c-commands the other one) must undergo the movement. Thus, movement must apply to the *wh*-subject rather than to the *wh*-object in (78-a) vs. (78-b), and in (78-c) vs. (78-d).

(78)   a.   Who$_1$ t$_1$ saw what$_2$?
        b.   *What$_2$ did who$_1$ see t$_2$?
        c.   I wonder [$_{CP}$ who$_1$ t$_1$ bought what$_2$]
        d.   *I wonder [$_{CP}$ what$_2$ who$_1$ bought t$_2$]

Similarly, if two objects are *wh*-phrases, as in the three pairs of examples in (79) (with a control infinitive, an exceptional case marking infinitive, and a ditransitive construction), only the structurally higher object can undergo movement (see Pesetsky (1982, 602ff) and references cited there).[49]

(79)   a.   Who$_2$ did you [$_{VP}$ persuade t$_2$ [$_{CP}$ PRO to read what$_1$]] ?
        b.   ?*What$_1$ did you [$_{VP}$ persuade who$_2$ [$_{CP}$ t$_1'$ PRO to read t$_1$]] ?
        c.   Who$_2$ did you [$_{VP}$ expect [$_{IP}$ t$_2$ to buy what$_1$]] ?
        d.   ?*What$_1$ did you [$_{VP}$ expect [$_{IP}$ who$_2$ to buy t$_1$]] ?
        e.   I know what$_2$ I should give t$_2$ [$_{PP}$ to whom$_1$]
        f.   ?*I know whom$_1$ I should give what$_2$ [$_{PP}$ to t$_1$]

Like the A-over-A Principle (in its various versions), the Superiority Condition qualifies as a good constraint in the sense of the requirements (39) and (41): It is simple and general (i.e., not construction-specific – even though the only type of dependency it has so far been exemplified by is *wh*-movement, it makes predictions for all transformations); and it is not complex (i.e., it does not require the postulation of reference sets, or correlating non-adjacent steps of derivations). What is more, the parallel behaviour of the A-over-A Principle and the Superiority Condition does not stop here.

---

[49] The existence of superiority effects with two objects provides a good argument against dispensing with the Superiority Condition in favour of the ECP (see (5)), as it had been suggested by Aoun, Hornstein & Sportiche (1981), Chomsky (1981), Lasnik & Saito (1984), and many others: On this view, the ECP as an LF filter can exclude sentences such as(78-b) and (78-d) because the traces of LF-moved subject *wh*-phrases emerge as not properly governed in the presence of another *wh*-phrase in SpecC. This does not help with the ungrammatical examples in (79), where only *wh*-objects are involved (whose traces are always properly governed by virtue of being lexically governed).

The Superiority Condition is also a constraint that looks quite modern and can thus be straightforwardly rephrased in a more current framework, as requiring movement of the closest item that bears a feature matching the attracting feature on the head of the landing site (thus, a rule "applies ambiguously to Z and Y" iff both Z and Y could satisfy the movement-inducing feature associated with the target head X); "closeness" is understood in terms of minimal c-command. On this view, (77) demands that if there is a head X with a movement-inducing feature $\mathcal{F}$, of all the items $Y_1,...,Y_n$ that are in the c-command domain of X and bear a matching feature $\mathcal{F}$, only the highest Y_i (i.e., the Y_i that c-commands all other Y_j's) can be moved to the specifier of X. Thus, just like the A-over-A Principle can be reformulated as a feature-based constraint (the F-over-F Principle; see (60)), the Superiority Condition can be reformulated as a feature-based constraint, as in (80). This faithful updating of (77) has become known as the Minimal Link Condition (MLC); see Chomsky (1995; 2000; 2001b; 2008).[50]

(80) *Minimal Link Condition* (MLC):
      In a structure $\alpha_{[\bullet F \bullet]}$... [... $\beta_{[F]}$ ... [... $\gamma_{[F]}$ ...] ...] ..., movement to [●F●] can only affect the category bearing the [F] feature that is closer to [●F●].

The only difference to the (revised) A-over-A Principle (i.e., the F-over-F Principle) is that β c-commands γ in the Superiority Condition, whereas β dominates γ in the F-over-F Principle; compare (80) with (81) (= the F-over-F Principle in (60)).

(81) *F-over-F Principle*:
      In a structure $\alpha_{[\bullet F \bullet]}$... [$_{\beta_{[F]}}$ ... [$_{\gamma_{[F]}}$ ...] ...] ..., movement to [●F●] can only affect the category bearing the [F] feature that is closer to [●F●].

3.8   The (Generalized) Minimal Link Condition

It seems obvious that further generalization is possible. The F-over-F Principle and the MLC can be combined into a constraint that I will refer to in what follows as the *(Generalized) Minimal Link Condition* (or (G)MLC); see Takano (1994), Kitahara (1997), Müller (1998), Fitzpatrick (2002), Rackowski & Richards (2005), and Medeiros (2008), among many others. The (G)MLC is given in (82).

---

[50] Basically the same constraint has been proposed by Fanselow (1990a; 1991) (under the label "Relativized Minimality Condition" – see below for the similarly named but different constraint in Rizzi (1990)), and by Ferguson (1993) and Ferguson & Groat (1994). The constraint can be given a representational reformulation; see (i).

(i)   *Minimal Link Condition* (representational version):
      *... $\gamma_{[F]}$... [... $\beta_{[F]}$ ... [... t_γ ...] ...] ... if the head of which γ is the specifier bears a [●F●] feature.

(82)   *Generalized Minimal Link Condition*:
       In a structure $\alpha_{[\bullet F\bullet]}$... [... $\beta_{[F]}$ ... $\gamma_{[F]}$ ...] ..., movement to $[\bullet F\bullet]$ can only
       affect the category bearing the $[F]$ feature that is closer to $[\bullet F\bullet]$.

The notion "closer" is simply understood in terms of path length (see (25) and (i) of
footnote 9). Thus, β is closer to α than γ if β dominates or c-commands γ – in both
cases, the path from β to α is shorter than the path from γ to α.

3.9   The Clause Non-final Incomplete Constituent Constraint

The Clause Non-final Incomplete Constituent Constraint is proposed in Kuno (1973)
as a more general version of the Sentential Subject Constraint, which it is therefore
supposed to replace. It is given in (83).[51]

(83)   *Clause Non-final Incomplete Constituent Constraint*:
       It is not possible to move any element of a category α (α = DP or CP) in
       a clause non-final position out of α if what is left over in α constitutes an
       incomplete α.

The notion of incompleteness can be understood as in (84).

(84)   *Incompleteness*:
       An XP α is incomplete if an obligatory element is missing.

An obligatory element may, as a first approximation, be understood as an element that
is obligatorily subcategorized. Thus, the complement of a preposition is obligatory
in this sense, and so is (typically, but perhaps not always) the object of a verb; in
contrast, the complement of a noun is not obligatory. Hence, if DP is missing in $[_{PP}$
P –], the PP is incomplete; if DP is missing in $[_{VP}$ V –], the VP is incomplete; but
if PP is missing in $[_{NP}$ N –], the NP is not incomplete. Of course, incompleteness
of a daughter is inherited by the mother. Thus, a DP is incomplete if it contains a PP
where the DP complement of P is missing (but not if it contains an NP where the PP
complement is missing); and so forth.

   The Clause Non-final Incomplete Constituent Constraint indirectly distinguishes
between object DPs and subject DPs in English because the former are typically

---

[51] The Clause Non-final Incomplete Constituent Constraint is a local derivational constraint, but it can be
reformulated as a representational constraint on outputs.

(i)   *Clause Non-final Incomplete Constituent Constraint* (representational version):
       *... α_1 ... [$_\beta$ ... t$_1$ ...] ... if (a)–(c) hold:
       a. β = DP or CP.
       b. β is in a clause non-final position.
       c. β is incomplete.

clause-final whereas the latter are typically not clause-final. Turning to objects first, extraction from an object DP is predicted to be possible even if what remains is an incomplete DP as long as the object is clause-final, and predicted to be impossible if an incomplete DP results in a position that is not clause-final (e.g., if extraction targets the first object in a double object construction). These predictions are confirmed by the data in (85).

(85)   a.   [$_{DP_1}$ Which man] did you buy [$_{DP}$ a picture of t_1] ?

            b.   [$_{PP_2}$ Of which man] did John give [$_{DP}$ a picture t_2] to Bill ?

            c.   ?*[$_{DP_1}$ Which man] did John give [$_{DP}$ a picture of t_1] to Bill ?

In (85-a), the DP is clause-final (also see (48-b)); in (85-b), the DP counts as complete (recall that arguments of N are optional). Only in (85-c) are both requirements violated: The DP from which movement takes place is in a non-final position, and if movement occurs, it counts as incomplete.

    Examples that Kuno offers to show the predictions for subject DPs are given in (86). In (86-a), extraction takes place from an object DP, leaving an incomplete constituent in clause-final position; in (86-b), passivization has taken place, and the same DP as in (86-a) is now a subject. Still, this sentence is judged as grammatical by Kuno, which is in accordance with his assumptions because the DP does not count as incomplete.[52] Finally, in (86-c) preposition stranding takes place, and the subject DP is both non-final and incomplete, which results in ungrammaticality.[53]

(86)   a.   [$_{DP_1}$ Which cars] did the explosion damage [$_{DP}$ the hoods of t_1] ?

            b.   [$_{PP_2}$ Of which cars] were [$_{DP}$ the hoods t_2] damaged by the explosion ?

            c.   *[$_{DP_1}$ Which cars] were [$_{DP}$ the hoods of t_1] damaged by the explosion ?

Furthermore, Kuno points out that the Clause Non-final Incomplete Constituent Constraint makes predictions for CPs that make it possible to dispense with the Sentential Subject Condition (recall the data in (70)), and that actually turn out to be empirically superior. Extraction from a CP that becomes incomplete as a result of the movement operation (in the technical sense of (84)) is possible if CP is a clause-final object (see (87-a)), and impossible if CP is a clause non-final subject (see (87-c)), but possible

---

[52] Also see Chomsky (2008) and section 4 of chapter 4 below for these kinds of examples.

[53] (86-b) is expected to be ungrammatical under the Subject Condition if the structure is as indicated. However, it has been suggested that these kinds of PPs (also cf. (85-b)) may in fact be base-generated outside the subject DP (see Koster (1987, 196f.), Cinque (1990, 47), Sternefeld (1991b, 121), Müller (1995, 397f.), Barbiers (2002, 54), and Gallego (2007, 349)), in which case the Subject Condition would be compatible with (86-b) (and the Clause Non-final Incomplete Constituent Constraint would be vacuously fulfilled here). I will come back to this construction in chapter 4, in the context of discussing Chomsky's (2008) approach to extraction from subjects.

again if CP is a clause-final subject, as in passive constructions (see (87-b)). This latter prediction is not made by the Sentential Subject Constraint (or the Subject Condition).

(87) a. [$_{DP_1}$ Who] did the reporters expect [$_{CP}$ that the principal would fire t_1] ?
      b. [$_{DP_1}$ Who] was it expected by the reporters [$_{CP}$ that the principal would fire t_1] ?
      c. *[$_{DP_1}$ Who] was [$_{CP}$ that the principal would fire t_1] expected by the reporters ?

As a further example showing how the Clause Non-final Incomplete Constituent Constraint works, consider next a sentence like (88), where successive-cyclic movement has taken place.

(88) [$_{DP_1}$ Who] does John think [$_{CP_2}$ Mary has persuaded t_1 [$_{CP_3}$ that Bill is a spy]] ?

At first sight, one might think that (88) violates the Clause Non-final Incomplete Constituent Constraint because extraction takes place from the t_1 position, which looks like it is clause non-final. So, does the wellformedness of (88) pose a problem? Closer inspection reveals that (88) does not violate the Clause Non-final Incomplete Constituent Constraint after all because the only CP from which movement takes place is CP_2; and CP_2 is incomplete after the movement operation, but it is in a clause-final position.

The case is slightly more involved with the examples in (89-ab).

(89) a. [$_{DP_1}$ Which man] did you buy [$_{DP}$ a picture of t_1] from Mary ?
      b. [$_{DP_1}$ Which tree] did you see [$_{DP}$ the leaves of t_1] in the yard ?

Kuno assumes that (89-ab) are well-formed, and he takes this to follow from the Clause Non-final Incomplete Constituent Constraint. The idea is that what is problematic about the starred data is "the fact that the incomplete ... phrases are followed by nonoptional elements [...] In [(89-ab)], ... incomplete ... phrases appear either clause-finally or, if not, are followed only by optional elements in the sentences." However, it can easily be verified that, whatever the merits of this intuition may be, the wellformedness of the data in (89-ab) does in fact not follow from the formulation of the constraint in (83). Therefore, we can conclude that the sentences in (89) pose a genuine problem for Clause Non-final Incomplete Constituent Constraint.

Considerations like these notwithstanding, we may ask, as before, whether the Clause Non-final Incomplete Constituent Constraint qualifies as a good constraint in the sense of (39), (41). This does not seem to be the case: The constraint is neither simple (since it juxtaposes a number of logically independent concepts, which becomes particularly evident in the representational formulation of the constraint in footnote

51) nor general (it mentions DPs and CPs, which makes it construction-specific); it is local, though (hence, optimal from the point of view of complexity).

The Clause Non-final Incomplete Constituent Constraint has been regularly discussed (and sometimes even adopted) in syntactic work of the last two decades (see, e.g., Lasnik & Saito (1992, 91 & 194), Kayne (1998, 179), Chomsky (2001b, 46), and Nevins (2004, sect. 6)). This does not seem to hold for the Post-Sentential Subject Extraction Constraint, a brief discussion of which may nevertheless be revealing.

3.10   The Post-Sentential Subject Extraction Constraint

Zaenen & Pinkham (1976) introduce the Post-Sentential Subject Extraction Constraint in (90).

(90)   *Post-Sentential Subject Extraction Constraint*:
        It is impossible to move a DP across a sentential subject.

Unlike the Clause Non-final Incomplete Constituent Constraint, this constraint is supposed to complement (rather than replace) the Sentential Subject Constraint. The example pair in (91) highlights a consequence of (90) for *wh*-movement: A non-sentential subject DP permits *wh*-movement across it; a sentential subject CP does not.[54]

(91)   a. $[_{DP_1}$ Who] do you think $[_{CP_1}$ that $[_{DP_2}$ Bill's resignation] would surprise t_1] ?
        b.   *$[_{DP_1}$ Who] do you think $[_{CP_1}$ that $[_{CP_2}$ for Bill to resign] would surprise t_1] ?

The same asymmetry is shown for topicalization across a sentential subject in (92).

(92)   a. $[_{DP_1}$ John] $[_{DP_2}$ Bill's resignation] would not surprise t_1
        b.   *$[_{DP_1}$ John] $[_{CP_2}$ for Bill to resign] would not surprise t_1

As before, there is a representational alternative to the derivational formulation of the constraint in (90).[55]

Since (90) is designed to complement the Sentential Subject Constraint, one might think that a more general constraint should be possible that subsumes the two individ-

---

[54] An infinitival clause is chosen here because a finite non-extraposed subject clause is highly marked to begin with.

[55] Here is a representational version that employs the concept of trace.

(i)   *Post-Sentential Subject Extraction Constraint* (representational version):
        *...α_1 ... [... β ... [... t_1 ...] ...] if β is a sentential subject c-commanding t_1, and α is a DP.

ual constraints. According to the Sentential Subject Constraint, all sentential subjects are islands. According to the Post-Sentential Subject Extraction Constraint. the domain to the right of a sentential subject is an island. A more general constraint might therefore state that all sentences with sentential subjects are islands, period. Unfortunately, this is not possible, for a simple reason: Sentential subjects themselves can be moved; see (93).

(93)   That John would be late, Mary didn't think was very likely.

Still, these considerations may at least suffice to show that the Post-Sentential Subject Extraction Constraint is not as simple as it arguably should be. Furthermore, it is not maximally general because it imposes restrictions both on the category label of the moved item (DP) and on the grammatical function (or syntactic position) of the island-inducing item (sentential subject).

    The constraint that is discussed in the next subsection differs greatly from the one just discussed, both with respect to its empirical scope and with respect to its relevance for further developments in syntactic theory.

3.11   The Condition on Extraction Domain

Phrases which do not enter a derivation as a result of being subcategorized by the item that they merge with (modifiers, adjuncts) are always islands. This can be formulated in a preliminary way as the Adjunct Condition in (94) (see Cattell (1976), Kayne (1984), Longobardi (1985), and Chomsky (1986a), among others).[56]

(94)   *Adjunct Condition*:
         Movement must not take place from an XP that has been merged without sub-
         categorization being involved.

The Adjunct Condition straightforwardly excludes Complex NP Constraint construc-
tions in which a relative clause is crossed by movement (see (66)). Furthermore, the
sentences in (95) are correctly predicted to be ill formed.

---

[56] There are some putative exceptions. Extraction from some kinds of infinitival adjunct clauses in German, e.g., may yield half-way acceptable results. Consider (i) (from Müller (1995, 87)).

(i)   ?Was$_1$ bist du   gekommen [$_{CP}$ (?*um)   anzuschauen t$_1$] ?
        what are you come   in order to watch

As a rule of the thumb, the fewer items the infinitival CP has, the more extraction is improved. Best results are obtained when the CP which from extraction takes place consists of just a bare verb form; and this might ultimately suggest a structural analysis in terms of complex verb formation (or a serial verb construction), without a CP adjunct being present.

Putative exceptions to the Adjunct Condition in English are also typically confined to infinitival contexts; see Chomsky & Lasnik (1977) and Chomsky (1986a), among others. I presuppose that they can all be accounted for systematically, without calling (94) into question.

(95)   a. *[$_{DP_1}$ Who] did you get jealous [$_{CP}$ because I talked to t_1] ?
        b. *[$_{PP_1}$ To whom] did they leave [$_{CP}$ before speaking t_1] ?
        c. *[$_{DP_1}$ Who] did they leave [$_{CP}$ before speaking to t_1] ?

The Subject Condition (cf. (72)) and the Adjunct Condition can be unified as the Condition on Extraction Domain (CED). The original version of the CED is due to Huang (1982).[57]

(96)   *Condition on Extraction Domain* (CED, Huang's original version):
        A phrase A may be extracted out of a domain B only if B is properly governed.

Based on Huang (1982) and Lasnik & Saito (1984), Chomsky (1986a) develops a comprehensive approach to CED effects; it centers around the notion of *barrier*. Cinque (1990) and Manzini (1992) further extend this research; these studies have useful simplifications and modifications. The following version of the CED freely draws on all the concepts developed in these approaches. It maintains Huang's original insight but rephrases the constraint by invoking only simpler concepts – in particular, the complex notion of *proper government* is replaced with the primitive notion of *complement* (i.e., sister of a lexical head).[58]

(97)   *Condition on Extraction Domain* (CED):
        a. Movement must not cross a barrier.
        b. An XP is a barrier iff it is not a complement.

Conceptually, the CED is a step in the right direction because we move from an intrinsic definition to a contextual definition of locality domains: Whether some XP is a bounding node or not (in the sense of the Subjacency Condition) is simply listed; however, whether some XP is a barrier or not can be determined by looking at the syntactic context in which it occurs.

A barriers-based approach to locality in terms of the Condition on Extraction Domain can account for Subject Condition and Adjunct Condition effects. It also derives the relative clause case of the Complex NP Constraint. If argument clauses selected by N are in fact not merged in complement position (as suggested by Stowell (1981) and Kiss (1986), among others), Complex NP Constraint phenomena can be explained in toto. A further constraint that can (ideally; see below) be dispensed with if the

---

[57] The basic insight was arguably first formulated by Cattell (1976), via his notion of a "syntactic configuration" (see Cattell (1976, 23 & 33)). Kayne (1984) and Longobardi (1985) employ a similar concept (cf. the theory of "g-projections"), with many interesting extensions ("Connectedness effects"); also cf. ściteKoster:87 concept of Global Harmony.

[58] The notion of *barrier* in (97) shows up here only for reasons of perspicuity. The CED might just as well state that movement must not take place from a non-complement.

Condition on Extraction Domain is adopted is the Freezing Principle. The version of the Freezing Principle in (98) combines aspects of Ross's (1967) Frozen Structure Constraint and Wexler & Culicover's (1980) Freezing Principle.[59]

(98)   *Freezing Principle*:
        Movement cannot take place from a moved XP.

Some consequences of the Freezing Principle in (98) are illustrated in (99) with English and German data.

(99)   a. *Who$_1$ do you think [$_{CP}$ t$_1'$ that [$_{DP_2}$ pictures of t$_1$] John would like t$_2$] ?
        b. *Who$_1$ do you think [$_{CP}$ [$_{PP_2}$ to t$_1$] he will talk t$_2$] ?
        c. *John$_1$, I know [$_{CP}$ [$_{DP_2}$ which picture of t$_1$] Mary bought t$_2$]
        d. *[$_{PP_1}$ Über Fritz]$_1$ glaube ich [$_{CP}$ [$_{DP_2}$ ein Buch t$_1$] hat Maria t$_2$
                 about Fritz   believe I   a   book   has Maria
               geschrieben]
               written
        e. *Who$_1$ do you think [$_{CP}$ t$_1'$ that [$_{DP_2}$ pictures of t$_1$] were painted t$_2$] ?

In (99-a), *wh*-extraction takes place from an object DP that has undergone (embedded) topicalization, and the result is ill formed. Similarly, in (99-b), a PP has undergone topicalization, and P-stranding via *wh*-movement in this context leads to ungrammaticality. In (99-c), an object *wh*-phrase has undergone movement to the SpecC position of an embedded question, and this makes extraction (topicalization, in the case at hand) from the object *wh*-phrase impossible. (99-d) shows the same effect with topicalization to a SpecC position of an embedded verb-second clause in German: Extraction from a topicalized item is prohibited. ((99-d) involves topicalization from a topicalized object DP, but the freezing effect arises with all kinds of extraction from a topicalized item, including, e.g., *wh*-movement.) Finally, in (99-e), it is shown that extraction from an A-moved subject DP in a passive clause gives rise to a freezing effect, too.[60]

   The Freezing Principle in (98) can be derived from the CED if the standard assumption is made that all movement is to a non-complement position. If so, movement will invariably turn an element (that may have been transparent in situ, in a

---

[59] Note, however, that according to Wexler & Culicover's (and also partially according to Ross's) assumptions, there are more freezing contexts than just the ones identified by (98); I will stick to the more narrowly defined notion of freezing in (98), though.

[60] In English, where T has an EPP property, subject DPs that are base-generated in Specv (typically) end up in SpecT. Thus, the opacity of subject DPs of this type follows from both the Subject Condition and the Freezing Principle.

complement position) into a barrier in the sense of (97).[61]

3.12   Relativized Minimality

Unlike most of the constraints that have been discussed in the above subsections, Relativized Minimality has been proposed as a representational constraint. It is introduced in Rizzi (1990); (100) is the updated version of the constraint suggested in Rizzi (2001).[62]

(100)   *Relativized Minimality*:
  a.   Chain links require a Minimal Configuration (MC).
  b.   Y is in a Minimal Configuration (MC) with X iff there is no Z such that (i) and (ii) hold.
      (i)   Z is of the same structural type as X.
      (ii)   Z intervenes between X and Y.

----

[61] There is a proviso. As it stands, the Freezing Principle in (98) is a derivational constraint. To exclude the examples in (99) by either (97) or (98), it must therefore be ensured that a derivation in which extraction of some item β takes place from some item α when α is still in its base position is blocked. In such a derivation, β is first moved out of α, and only then does α undergo movement itself. A derivation of this type can have one of two possible forms (see Collins (1994) and Chomsky (1995), among others): First, movement of α may go to a lower position than the preceding movement of β; this would violate the Strict Cycle Condition. Second, movement of α may go to a higher position than the preceding movement of β, with β subsequently moving across α again, as an instance what Collins calls "chain interleaving". Collins (1994) argues that chain interleaving is ruled out by the transderivational Fewest Steps Condition; alternatively, Müller & Sternefeld (1996) propose that derivations of this type can be excluded by an independently motivated theory of improper movement (based on the Principle of Unambiguous Binding, PUB). I will come back to this issue. For the time being, it may be noted that this kind of problem does not show up in an approach in which the CED (or the Freezing Principle) is reformulated as a representational constraint; see Browning (1991). Browning suggests to simply look at the S-structure representation and determine whether or not a category is a barrier in its S-structure position. Thus, on Browning's view, in a representation like (i), β is a barrier (because it is not a complement): Having undergone movement, β can never be a complement to some head (it has to be in a specifier position or in an adjoined position).

(i)   *... α_1 ... [$_\beta$ ... t_1 ...]$_2$ ... t_2

[62] The original version from Rizzi (1990) is given in (i) (with a slight simplification concerning the requirement that movement chains demand antecedent-government).

(i)   *Relativized Minimality* (original version):
  a.   (Certain) chain links require antecedent-government.

Here, intervention is understood in terms of c-command: Z intervenes between X and Y if X c-commands Z, Z c-commands Y, and Y does not c-command X.[63]

Relativized Minimality is similar, but far from identical to the Minimal Link Condition (MLC) (hence, also from the (G)MLC; but I focus on the former to simplify comparison). To see the similarities and differences, the MLC in (80) is repeated here, in a form that is slightly revised so as to make a comparison with Relativized Minimality easier, as (101); and the standard, representational Relativized Minimality condition is transferred to a derivational constraint in (102). In the course of doing so, a number of secondary issues are glossed over; the focus is just on the gist of the constraint. (The minimal domain of α comprises specifiers and head-adjunction sites of α.)

(101)   *Minimal Link Condition* (MLC):
        In a structure α ... [... β ... [... γ ...] ...] ..., movement to a position δ in the minimal domain of α cannot affect γ if β c-commands γ and β *and* γ *bear the same feature required by* α.

(102)   *Relativized Minimality* (derivational version):
        In a structure α ... [... β ... [... γ ...] ...] ..., movement to a position δ in the minimal domain of α cannot affect γ if β c-commands γ and β *and* δ *are of the same structural type*.

Thus, according to the MLC, an intervener blocks movement if it bears the same *feature* that is required by the movement-inducing head, as the item that is supposed to undergo the movement; in contrast, according to Relativized Minimality, an intervener blocks movement if it occupies the same kind of *position* as the intended landing site for the item that is supposed to undergo the movement. Therefore, the two constraints

---

        b.   X antecedent-governs Y only if there is no Z such that (i) and (ii) hold.
                        Z is a typical potential antecedent-governor for Y.
                        Z c-commands Y and does not c-command X.

The notion of "typical potential antecedent governor" is defined as follows.

(ii)   *Typical potential antecedent governor*:
        a.   Z is a typical potential antecedent governor for Y, Y in an A-chain = Z is an A specifier c-commanding Y.
        b.   Z is a typical potential antecedent governor for Y, Y in an A′-chain = Z is an A′ specifier c-commanding Y.
        c.   Z is a typical potential antecedent governor for Y, Y in an X^0-chain = Z is a head c-commanding Y.

[63] This last requirement may be obsolete assuming binary branching, a restrictive notion of c-command, and an approach to phrase structure in which two sisters cannot possibly be of "the same structural type".

make different empirical predictions. For instance, Relativized Minimality cannot derive superiority effects as in (78) and (79). On the other hand, the MLC does not account for all the cases that motivate Relativized Minimality in Rizzi (1990). Let us go through them one by one. First, Rizzi notes that Relativized Minimality excludes cases of so-called "super-raising" as in (103). Here, A-movement to SpecT takes place across a finite clause with a SpecT (i.e., A-) position filled by an expletive *it*.

(103)   *$John_1$ seems that *it* is likely [$_{TP}$ t_1 to win]

(103) can also be ruled out by the MLC, assuming that the feature that drives movement to SpecT in the root clause is one that is borne by both *it* and *John*.

Next, consider the sentences in (104). Relativized Minimality blocks (104-abcd), assuming that all italicized items occupy A-bar positions. This seems uncontroversial with the *wh*-island effect in (104-b) (= (74-d)), where both items occupy SpecC positions; and also with (104-c), where topicalization has applied across a *wh*-island. To exclude (104-a), it must be assumed that *beaucoup* occupies an A-bar position. Similarly, some extra assumption is called for to derive the *wh*-island effect in (104-d) (= (74-b)) from Relativized Minimality (at least at first sight, *whether* looks like a C head, not like a specifier).

(104)   a. *$Combien_1$ a-t-il   *beaucoup* consulté [t_1 de livres] ?
             how many did he a lot   consult   of books
         b. ?*[$_{DP_1}$ *Which book*] do you wonder [$_{CP}$ [$_{PP_2}$ *to whom*] John gave t_1 t_2] ?
         c. ?*[$_{DP_1}$ *This book*] I wonder [$_{CP}$ [$_{PP_2}$ *to whom*] John gave t_1 t_2]
         d. *How_1 do you wonder [$_{CP}$ *whether* Mary solved the problem t_1] ?

Does the MLC suffice to rule out the examples in (104)? This may be so in the case of (104-b) and (104-d), assuming that a *wh*-moved *wh*-phrase still has a feature like [wh] that is looked for by some higher C head, and assuming that *whether* also qualifies as an item that has a [wh] feature accessible to a higher C head – however, this assumption is by no means uncontroversial.[64] (105-c) can be excluded by the MLC only if it is assumed that topicalization and *wh*-movement in English are triggered by the same kind of feature (see Chomsky (1977)); but given the different properties of the two movement types, this assumption should not be viewed as uncontroversial either (see Müller & Sternefeld (1993)). Finally, it is far from obvious how the MLC could block (104-a) – *beaucoup* does not seem to be a *wh*-item in any relevant sense, and it certainly cannot be attracted to the specifier position of an interrogative C in French.

---

[64] Also see chapter 5 for a qualification.

A third consquence of Relativized Minimality is that it accounts for effects that had originally been assumed to require a separate *Head Movement Constraint* (HMC; see Travis (1984), Baker (1988)). As shown in (105) (from Chomsky (1986a), Rizzi (1990)), movement of a head across an intervening head is impossible.

(105)   a.   *Could*$_1$ they t$_1$ *have* left ?
         b.   *Have$_2$ they *could* t$_2$ left ?

The contrast between (105-a) and (105-b) can also be accounted for via the MLC if the two verb forms competing for movement have the same kind of feature that is sought for in the target position (an adjunction site of C, under standard assumptions about head movement). That said, it is not clear that (105-b) needs to blocked by a locality constraint on movement in the first place. Arguably, a non-finite verb form (like *have* in (105-ab)) cannot show up attached to the head of a finite clause in English, for independent reasons. (*Could* in (105-ab) is finite since it is the highest head of an extended verbal dependency in the pre-movement structure.)

Subsequent research on the Relativized Minimality constraint has focussed on the possibility that a finer-grained differentiation among "structural types" of positions (in the sense of (100), (102)) is called for – e.g., one that distinguishes between different types of head movement positions, or between different types of A-bar positions; see Roberts (1991; 2001) and Rizzi (2001; 2004). Irrespective of these further developments, we may ask whether Relativized Minimality qualifies as a good constraint, one that respects the meta-requirements (39) and (41). This would seem to be the case: The constraint is not complex; and it is simple and general (thus, it is not construction-specific and does not mention particular categories).[65] Still, in view of (i) the redundancies that arise from the significant amount of overlap between Relativized Minimality and MLC in excluding ill-formed sentences, and (ii) the conceptual similarity of the two constraints, it seems uncontroversial that only one of them (if any; see below) should be adopted in grammatical theory.

3.13 Conclusion

This concludes the overview of locality constraints as they are relevant in the context of the (G)MLC and the CED.[66]

---

[65] This is particularly obvious if Rizzi's (2001) later version in (100) or the derivational reconstruction in (102) is adopted, perhaps less so under the original formulation in (i) of footnote 62.

[66] Needless to say, many important locality constraints have been left out – for instance, the *Left Branch Condition* (LBC) and the *Coordinate Structure Constraint* (CSC), to name just two (see Ross (1967)). The reason for this omission is that these constraints are not directly related to the (G)MLC and the CED, and no attempt will be made in later parts at deriving their effects by invoking the PIC (even though both the LBC and the CSC will show up again briefly in chapter 7 below). In addition, there are some influential

From a more general point of view, we can further distinguish between two types of (local derivational or local representational) locality constraints. On the one hand, there are *rigid* locality constraints that define local domains for movement irrespective of what kind of position the moved item ends up in, and what kind of inherent properties it has. On the other hand, there are *relativized* locality constraints where the nature of the ultimate landing site and/or properties of the moved item are relevant to the question of whether a given syntactic configuration acts as local domain. Consider the following two lists in (106-a) and (106-b) (with classic island constraints all belonging to the first group).

(106)   a.   *Rigid Locality*:
                (i)   Complex NP Constraint
                (ii)   Sentential Subject Constraint
                (iii)   Subject Condition
                (iv)   *Wh*-Island Condition (and Topic Island Condition)
                (v)   Clause Non-final Incomplete Constituent Constraint
                (vi)   Post-Sentential Subject Extraction Constraint
                (vii)   Adjunct Condition
                (viii) Condition on Extraction Domain
         b.   *Relativized Locality*:
                (i)   A-over-A Principle
                (ii)   F-over-F Principle
                (iii)   Superiority Condition
                (iv)   Minimal Link Condition
                (v)   Relativized Minimality

At least at first sight, it looks as though both types of constraints are needed, but it is far from clear which phenomena should be accounted for by which constraint type.

Against this background, let me come back to the question of what a good locality constraint in syntax is. As noted before, the requirement that constraints should not be complex (see (41)) directly excludes global constraints like the *Projection Principle*, translocal constraints like *Avoid Pronoun*, and transderivational constraints like *Fewest Steps*, *Merge before Move*, and *Shortest Paths*.[67]

---

locality constraints that I more or less ignore in what follows even though they *are* related to the (G)MLC and/or the CED, among them the Path Containment Condition developed in Pesetsky (1982) and some constraints mentioned in footnote 57 above, viz., the Connectedness condition introduced in Kayne (1984), and developed further in Longobardi (1985), as well as Koster's (1987) Global Harmony constraint. Here, the reason for omission is that the constraints in question seem inherently tied to a representational approach, and thus do not significantly interact with the enterprise to be carried out in the present book.

[67] As regards *Shortest Paths* (see (24)), it can be noted that it is similar in its effects to a combination

Next, the requirement that good constraints are simple and general (and contribute to overall elegance of the theory) excludes construction-specific constraints like the *Complex NP Constraint*, the *Sentential Subject Constraint*, the *Subject Condition*, the *Adjunct Condition*, the *Subject Condition*, the *Wh-Island Condition* (as well as the *Topic Island Condition*), the *Clause Non-Final Incomplete Constituent Constraint*, and the *Post-Sentential Subject Extraction Constraint*.

Of the locality constraints that have been discussed in the present chapter, this leaves the A-over-A Principle, the F-over-F Principle, the Superiority Condition, the MLC, the (G)MLC, Relativized Minimality, and the CED. The (original) A-over-A Principle has been shown to be problematic; the A-over-A Principle revised according to Bresnan (1976) and the closely related F-over-F Principle are now part of the (G)MLC, as is the Superiority Condition/MLC. As noted above, maintaining both Relativized Minimality and the (G)MLC leads to redundancies; since it also covers F-over-F Principle effects, the latter constraint is arguably the more general one. Given these considerations, one would expect two constraints to remain of all the constraints that have been addressed so far: the (G)MLC and the CED. Indeed, these two local constraints are widely adopted in contemporary investigations of movement.[68]

At this point, let me introduce a third requirement on what a good locality constraint looks like. This requirement is deeply rooted in recent minimalist work (see, e.g., Chomsky (2005)), and it can be expressed as in (107).

(107)   Constraints are of type (i) or (ii).
         (i)   principles of efficient computation
         (ii)   interface conditions

Principles of efficient computation in the sense of (107-i) are basically local derivational economy constraints. In contrast, interface conditions in the sense of (107-ii) are representational constraints on outputs which are not motivated syntax-internally, but result from demands imposed on syntactic representations by the phonological and semantic interfaces. Thus, constraints that rely on concepts without independent motivation in terms of either economy or interface conditions are excluded.

How do the (G)MLC and the CED fare with respect to the requirements in (107)?

---

of the F-over-F Principle (see (60) or (81)) and the MLC (see (80)). However, given that transderivational constraints should ideally be dispensed with, *Shortest Paths* does not qualify as a viable alternative to stating the F-over-F Principle and the MLC separately; but see below.

[68] That said, there are certain Relativized Minimality effects that do not follow straightforwardly from the (G)MLC or the CED, or from the set of assumptions that I introduce in chapters 3 and 4 to derive versions of the (G)MLC and the CED; in particular, this holds for certain kinds of *wh*-islands and topic islands (and legitimate *wh*-island violations), and for the HMC. I address the former issue in chapter 5, and I argue that the HMC is not a valid generalization in chapter 6.

The general consensus seems to be that the (G)MLC can be viewed as a principle of efficient computation, whereas the CED cannot straightforwardly be construed in such a way. Furthermore, neither constraint looks like an interface condition imposed by properties of the phonological or semantic component of grammar. In view of this state of affairs, the (G)MLC is typically adopted as a primitive in current miminalist approaches whereas the CED is often abandoned, accompanied by attempts to derive it from more basic assumptions that adhere to (107). In chapter 2, I argue (i) that there is good reason to give up the (G)MLC after all in a derivational approach to grammar that embraces the assumptions in (107); and (ii) that none of the existing attempts at deriving CED effects in accordance with (107) is successful. Given this outcome, the task will be to derive (G)MLC and CED effects from more basic assumptions that are compatible with (107); this is the subject of chapters 3 and 4.

Chapter 2

(G)MLC and CED in Minimalist Syntax

1. Introduction

The goal of this chapter is to argue that both the (G)MLC and the CED are incompatible with core minimalist principles in a strictly derivational approach to syntax, and that, consequently, their main predictions should be derived from more basic assumptions. This conclusion would seem to be unusual for the (G)MLC; in line with this, I am not aware of minimalist reconstructions of (G)MLC effects in the literature. In contrast, it seems to be a standard assumption that the CED is incompatible with minimalist core tenets; accordingly, several attempts have been made to derive its effects. However, I show that existing proposals of how to derive CED effects turn out to be either incompatible with minimalist principles upon closer inspection or problematic for other reasons (or both). The conclusion will be that, to the extent that the predictions of the (G)MLC and the CED are empirically correct, these locality constraints should be derived from independently motivated assumptions that are compatible with basic minimalist principles.

2. The (Generalized) Minimal Link Condition: State of the Art

2.1 Overview

The (G)MLC is repeated in (1).

(1) *Generalized Minimal Link Condition*:
In a structure $\alpha_{[\bullet F\bullet]}$... [... $\beta_{[F]}$ ... $\gamma_{[F]}$ ...] ..., movement to $[\bullet F\bullet]$ can only affect the category bearing the [F] feature that is closer to $[\bullet F\bullet]$.

The (G)MLC in (1) arguably qualifies as an economy constraint that contributes to efficient computation.[1] However, I would like to contend that something is wrong with

---

[1] Gärtner & Michaelis (2007, 175) show for a certain type of minimalist grammar (as developed by Stabler (1996; 1998)) that, other things being equal, adding (a version of) the (G)MLC reduces generative capacity.

the (G)MLC in a derivational grammar. Brody (2002) argues that a derivational approach to syntax should minimize search space, its representational residue; thus, the amount of structure that is visible and accessible to syntactic operations at any given step should be as small as possible. Given this tenet, it follows that constraints that minimize search space should be strengthened in a derivational grammar; in contrast, constraints that presuppose search space should be abandoned. A constraint that minimizes search space is the Phase Impenetrability Condition (PIC; see Chomsky (2000; 2001b)); in contrast, the (G)MLC is a constraint that presupposes search space. Furthermore, redundancies can be shown to arise if both the PIC and the (G)MLC are adopted. In addition to conceptual problems with the (G)MLC, I will also argue that empirical problems arise because the (G)MLC is both too strong and too weak: It is too strong because it excludes certain argument crossings as they arise with A-movement operations; and it is too weak because it only captures a subclass of the intervention effects that it arguably should derive: It fails to account for intervention effects that involve neither c-command nor dominance.

I will proceed as follows. Subsection 2.2 discusses problematic (but ubiquitous) instances of argument crossing and presents a superiority-like intervention effect that is unexpected under the (G)MLC, and that can be taken to suggest that some other, more general constraint (or system of constraints) might underlie intervention effects in general. Subsection 2.3 provides some background assumptions and lays out conceptual arguments against the MLC.

2.2   Empirical Arguments against the (G)MLC

2.2.1   Argument Crossing

There are a number of well-known problems with a simple version of the (G)MLC. An obvious problem is that subject raising from a vP-internal position to SpecT is wrongly expected to be blocked by the (G)MLC if object movement to Specv has occurred.[2] Consider, for instance, the derivation in (2), where a *wh*-object undergoes movement across the non-*wh*-subject, which in turn has to end up in SpecT.

(2)   (I wonder) what John read

    a.   [$_{VP}$ read$_3$ what$_1$]

    b.   [$_{vP}$ what$_1$ John$_2$ read$_3$ [$_{VP}$ t$_3$ t$_1$]]

Interestingly, they also show that a grammar that incorporates (a version of) the CED without simultaneously adopting the (G)MLC is less computationally efficient than a grammar that lacks both the CED and the (G)MLC; the most restrictive grammar results if both (G)MLC and CED are adopted.

[2] At least, this holds as long as we assume that object movement must end up in a position in vP that is higher than the base position of the subject; but see Richards (2001) for a different option (cf. the countercyclic operation of 'tucking-in').

    c.   [$_{TP}$ T [$_{vP}$ what$_1$ [$_{V'}$ John$_2$ read$_3$ [$_{VP}$ t$_3$ t$_1$]]]]

    d.   [$_{TP}$ John$_2$ T [$_{vP}$ what$_1$ t$_2$ read$_3$ [$_{VP}$ t$_3$ t$_1$]]]

    e.   [$_{CP}$ C [$_{TP}$ John$_2$ T [$_{vP}$ what$_1$ t$_2$ read [$_{VP}$ t$_3$ t$_1$]]]]

    f.   [$_{CP}$ what$_1$ C [$_{TP}$ John$_2$ T [$_{vP}$ t$_1'$ t$_2$ read [$_{VP}$ t$_3$ t$_1$]]]]

The problem is that *what*$_1$ is closer to T in (2-c) than *John*$_2$, and should therefore preclude movement of *John*$_2$ to SpecT in (2-d).

Note that the problem at hand is not confined to subject/object interaction. It also shows up with local object movements (object shift, scrambling) in double object constructions; see Bobaljik (1995) and Collins & Thráinsson (1996), among others – at least, this holds as long as it looks as though both objects are equipped with the same kinds of features that are involved in the movement, and that the (G)MLC is sensitive to. As a case in point, consider multiple object shift of pronouns in Danish and multiple object shift of non-pronominal DPs in Icelandic, as discussed by Vikner (1990) and Collins & Thráinsson (1996), respectively; see (i-a) (Danish) and (i-b) (Icelandic).

(3)   a.   Peter viste   hende$_1$ den$_2$ jo   t$_1$ t$_2$
               Peter showed her   it   indeed

    b.   Ég lána Maríu$_1$ bækurnar$_2$ ekki t$_1$ t$_2$
               I   lend Maria   the books not

Here, it seems that there has to be a derivational step where the direct object (DP$_2$) first undergoes movement across the adverbial (*jo*, *ekki*), thereby crossing the indirect object (DP$_1$) in violation of the (G)MLC; crucially, the indirect object has the same kind of feature that motivates the movement since it subsequently moves to a position in front of the direct object.[3] Several solutions to this problem have been proposed. A radical solution would be to assume that the landing site of the moved item is in fact a a specifier position in a phrasal domain that is lower than the phrasal domain in which the other item is base-merged (see, e.g., Bobaljik's (1995) discussion of "leap-frogging" vs. "stacking"); but this option is not available with PIC-driven movement as in (2).

---

[3] The problem is made worse by the fact that in those cases where the intervening item does not bear the relevant feature, and consequently stays in situ throughout the derivation, ungrammaticality results even though there should not be any (G)MLC violation involved; see (ii-a) (Danish) and (ii-b) (Icelandic).

(i)   a.   *Peter viste   den$_2$ jo   Marie$_1$ t$_2$
               Peter showed it   indeed Marie

    b.   *Ég lána bækurnar$_2$ ekki Maríu$_1$ t$_2$
               I   lend the books not Maria

Thus, it seems that there are contexts in which the (G)MLC can be freely violated by derivational steps. Chomsky (1995) envisages a way out in terms of the concept of *equidistance*, which is assumed to play a role in the formulation of the (G)MLC. On this view, there is no (G)MLC-based intervention if two items are sufficiently close to one another (e.g., if they are both specifiers of the same head).

The equidistance approach is abandoned in Chomsky (2000; 2001b) in favour of the stricter formulation of the (G)MLC in (1). The problem that the (G)MLC poses for subject raising in (2) is then addressed by observing that after *wh*-movement of *what*$_1$ to SpecC, the subject DP is the closest goal for T after all (the intervening object having left its position). At first sight, it seems that an execution of this idea implies giving up the Strict Cycle Condition (SCC): Movement in TP would have to follow movement in CP, in violation of strict cyclicity. Still, Chomsky suggests that there is a way out of this dilemma that respects both the SCC and the (G)MLC in strict versions: The idea is that the (G)MLC is not evaluated at each step of the derivation; rather, it is only evaluated at the phase level. Thus, subject raising in (2-d) would indeed violate the (G)MLC; but TPs are not phases, and the (G)MLC is therefore not operative at this stage. The (G)MLC (more specifically, an appropriately revised representational version of the constraint) does apply to the output in (2-f) because CP is a phase. However, at this point, there is no overt DP in Specv left that would separate the subject trace and T, and, given some obvious adjustments, it follows that the (G)MLC is respected. Assuming vP to be the relevant phase for (G)MLC checking with object movement across another object, a similar analysis can be given for the cases in (3).

Of course, there is now a change of perspective that is non-trivial: The (G)MLC cannot be conceived of as a derivational constraint on operations anymore; it acts as a representational constraint on certain kinds of structures (viz., trees with phases at the root). I will argue below that this amplifies a conceptual problem incurred by the (G)MLC.

For the time being, it can be concluded that the (G)MLC in (1) is too strong as it stands; this undergeneration problem can only be solved at the cost of introducing additional assumptions – about the base positions of arguments, about the notion of "closeness" (defined in terms of equidistance), or about the nature of the (G)MLC itself (a reinterpretation as a representational constraint). The next subsection will reveal that the (G)MLC is also too weak.

2.2.2 *Intervention without C-Command or Dominance*

The (G)MLC excludes derivations where an item is moved across a c-commanding or dominating intervener that would also qualify as an item that can be affected by the movement operation (because it has the same kind of movement-inducing feature). This way, superiority effects as in (4-a) in English (where the intervener c-commands

the trace of the moved item; recall (78) of chapter 1) and F-over-F effects as in (4-b) in German (where the intervener dominates the trace of the moved item; recall (62) of chapter 1) can be derived.

(4)   a.   *I wonder [$_{CP}$ what$_2$ who$_1$ bought t$_2$]
         b.   *dass [$_\alpha$ t$_1$ zu lesen]$_3$ [$_{DP}$ das Buch   ]$_1$ keiner t$_3$ versucht hat
             that   to read   the book$_{acc}$ no-one   tried   has

With this in mind, consider the German data in (5) (see Heck & Müller (2000)). In all three cases, we are dealing with a multiple *wh*-question where the moved item is a *wh*-object, and another *wh*-object is included in an adverbial clause. In (5-a), *wh*-movement is local (it does not cross a clause boundary) and unproblematic. The sentence may sound slightly clumsy, and it may perhaps be somewhat difficult to parse, but it is certainly well formed. In (5-c), the *wh*-item included in the adverbial clause has undergone movement. This results in ungrammaticality, which does not come as a surprise given the Adjunct Condition (or the CED, or whatever derives the adjunct island effect).

(5)   a.   Wen$_1$ hat Fritz [$_{CP}$ nachdem er was$_2$ gemacht hat] t$_1$ getroffen ?
            whom has Fritz   after   he what done   has   met
         b.   *Wen$_1$ hat Fritz [$_{CP}$ nachdem er was$_2$ gemacht hat] gesagt [$_{CP}$ dass
            whom has Fritz   after   he what done   has   said   that
            Maria t$_1$ liebt] ?
            Maria   loves
         c.   *Was$_2$ hat Fritz [$_{CP}$ nachdem er t$_2$ gemacht hat] gesagt [$_{CP}$ dass Maria
            what has Fritz   after   he   done   has said   that Maria
            wen$_1$ liebt] ?
            whom loves

The interesting piece of data is (5-b). Here, a *wh*-object undergoes long-distance movement from a declarative clause embedded by a bridge verb, and the result is ungrammatical. The source of the ungrammaticality must be in the adverbial clause; as shown in (6), a minimally different example in which it does not show up is grammatical.[4]

(6)   Wen$_1$ hat Fritz gesagt [$_{CP}$ dass Maria t$_1$ liebt] ?
      whom has Fritz said   that Maria   loves

Furthermore, it cannot be the adverbial clause as such that produces ungrammaticality in (5-b). If the *wh*-item in the adverbial clause is replaced with a non-*wh*-item, *wh*-

---

[4] At least, this holds for those speakers of German who tolerate *wh*-movement from *dass*-clauses to begin with; there are other speakers for whom only *wh*-scope marking and extraction from verb-second clauses result in well-formed long-distance *wh*-dependencies.

extraction becomes possible again; see (7), which suffers from clumsiness and parsing difficulties but is a vast improvement over (5-b).

(7)   Wen$_1$ hat Fritz [$_{CP}$ nachdem er das$_2$ gemacht hat] gesagt [$_{CP}$ dass Maria t$_1$
       whom has Fritz   after   he that done   has said   that Maria
       liebt] ?
       loves

Thus, the obvious conclusion has to be that (5-b) instantiates yet another intervention effect: The presence of the *wh*-item in the adverbial clause blocks long-distance movement of the other *wh*-item. However, since *was$_2$* in the adverbial clause neither c-commands nor dominates the base position of *wen$_1$* in (5-b), the (G)MLC has nothing to say about the illformedness of the example. Unless it can be shown that the intervention effect in (5-b) is of a significantly different nature than the intervention effect in (4-a) (and I will argue in chapter 3 that there is no reason for such an assumption, with (5-b) and (4-a) amenable to exactly the same analysis that also predicts (5-a) to be well formed), we may therefore conclude that the (G)MLC is not only too strong; it is also too weak.

2.3   Conceptual Arguments against the (G)MLC

2.3.1   The Standard Approach

In an incremental-derivational approach to movement as developed in Chomsky (2000; 2001b; 2005; 2008), two constraints prove particularly relevant; they reduce derivational search space by imposing strong restrictions on what counts as an active, accessible part of the derivation. First, the Strict Cycle Condition (SCC), arguably indispensable in any derivational approach to syntax, restricts possible positions for the operation-inducing head (for instance, for a head that triggers Agree via a probe feature, or a head that drives movement operations and creates the target for movement). Second, the PIC significantly reduces the positions in which the derivation can look for an item that is affected by an operation (for instance, an item that is to be moved, or a goal of an Agree operation). For present purposes, the SCC can be formulated in a classical way, as in (8) (see Chomsky (1973), Perlmutter & Soames (1979), and (23) in chapter 1).[5]

---

[5] Things may be slightly different if we adopt the concept of cyclic spell-out, according to which domains that have been rendered inaccessible via the PIC are immediately sent off to the phonological and semantic interfaces; see Chomsky (2001a, 4). Under the assumption that the domain of a phase head undergoes cyclic spell-out (in a true, non-metaphorical sense, i.e., via literal pruning of the tree created so far) after the phase is complete, a somewhat weaker version of the SCC can be derived without postulating a separate constraint (the version is weaker given that whereas the SCC is about all phrases, cyclic spell-out only affects the c-command domains of phases). However, throughout this book (with the possible exception of chapter 7), I will leave open the issue of whether the idea of cyclic spell-out is to be taken literally; consequently, I stick

(8)   *Strict Cycle Condition* (SCC):
      Within the current XP α, a syntactic operation may not target a position that is included within another XP β that is dominated by α.

A version of the PIC is given in (9) (see Chomsky (2000, 108), Chomsky (2001b, 13)); (9) is identical to (17) of chapter 1.

(9)   *Phase Impenetrability Condition* (PIC):
      The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

The notions of (i) "edge" and (ii) "phase" have been clarified in chapter 1. Essentially, (i) the edge of a head X is the left-peripheral minimal residue outside of X′; it includes specifiers of X, of which there can in principle be arbitrarily many (irrelevantly for the purposes of this chapter, it also comprises adjuncts to XP); see Chomsky (2001b, 13). (ii) The propositional categories CP and vP are phases; other XPs (except perhaps for DP) are not. With this in mind, let us look abstractly at syntactic derivations, and determine the search space available to the derivation at any given point. Thus, suppose that ZP, XP and UP are phases in (10) (this is illustrated by underlining). Then, in (10-a), an operation can have a probe (more generally, an operation-inducing feature) only in YP (because of the SCC), and an operation can look for a goal only in YP or in the residue or head of XP (because of the PIC). In the subsequent step (10-b), the probe must be in ZP, and the search space for a goal grows as indicated.

(10) *Search space under PIC*:

   a. $\underbrace{[_{YP}...Y\ \underbrace{[_{XP}...[_{X'}X}}\ [_{WP}...W\ [_{UP}...U...\ ]]]]]$
        $\quad\quad\quad\quad$ SCC $\quad\quad$ PIC

   b. $\underbrace{[_{ZP}...Z\ [_{YP}...Y\ \underbrace{[_{XP}...[_{X'}X}}\ [_{WP}...W\ [_{UP}...U...\ ]]]]]}$
        $\quad\quad\quad$ SCC $\quad\quad\quad\quad$ PIC

Crucially, the PIC does not allow an operation involving Y and an item in WP.[6]

_____

to the SCC as a separate constraint throughout.

[6] Chomsky (2001b) argues that such operations are in fact attested, and he gives the following example: Suppose that YP = TP, XP = vP, and WP = VP. The PIC then precludes an operation involving T and DP in VP; but such an operation must arguably be legitimate for instances of long-distance agreement with VP-internal nominative DPs, attested in a number of languages. Chomsky's solution is to weaken the phase impenetrability requirement in such a way that a phase is evaluated with respect to the PIC at the next phase level (see Chomsky (2001b, 14)). For the moment, I abstract away from these complications; I will come back to this issue in chapter 3.

It is an attractive feature of incremental-derivational approaches to syntax that complexity can be reduced, compared to representational approaches. Such a reduction of complexity becomes manifest in three different domains. First, the system does not permit look-ahead: At any given stage of the derivation, operations in later cycles and their effects cannot be considered. Second, the system relies on cyclicity: At any given stage of the derivation, the SCC makes it impossible to target a position by a syntactic operation that is not in the minimal domain of the current XP. And third, the system incorporates a phase impenetrability requirement (PIC) that significantly reduces the search space for the goal of an operation. In effect, all syntactic material in the domain that the PIC renders opaque can (and must) be ignored for the remainder of the derivation. So far, so good. However, closer inspection reveals conceptual problems with assuming both the (G)MLC and the PIC, as it is standardly done in minimalist syntax: The (G)MLC inherently depends on a certain amount of search space to work on, but the PIC reduces search space.[7]

2.3.2 Weak and Strong Representationality

In his comparison of derivational and representational approaches to syntax, Brody (2002) observes that a representational approach can be strictly non-derivational. In contrast, a derivational approach is usually representational to some extent, by adhering to the very concept of syntactic structure. Brody calls a derivational approach *weakly representational* if "derivational stages are transparent (i.e., representations), in the sense that material already assembled can be accessed;" and he calls it *strongly representational* if it "is weakly representational and there are constraints on the representations." On this view, the approach sketched in the previous subsection is strongly representational: This is not the fault of the SCC or the PIC; these are derivational constraints on operations. In the formulation given in (1), the (G)MLC is also a derivational constraint; however, this is not the case anymore if we re-interpret the (G)MLC in the way suggested at the end of the previous section to account for the existence of subject raising in examples like (2). Here, the (G)MLC is a representational constraint that is evaluated at the phase level; it checks the legitimacy of structures rather than operations. Brody concludes from this (and from related observations) that a representational approach has an inherent advantage over a derivational approach in this domain. Let us assume that the argument is correct. Then, given a derivational approach, the task will be to reduce its representational residue – ideally, a derivational theory should not even be weakly representational. This implies abandoning all constraints that presuppose too much structure (in a sense to be made precise); a good

---

[7] In fact, one might argue that the PIC could reduce search space even more, by relying on phrases rather than phases as the relevant locality unit. See chapter 3.

candidate for exclusion then is the (G)MLC.[8]

2.3.3 A Redundancy

Interestingly, a simultaneous adoption of the (G)MLC and the PIC leads to redundancies: As noted by Chomsky (2001b, 47, fn. 52), "the effect on the MLC is limited under the PIC, which bars 'deep search' by the probe." Thus, the (G)MLC can only become relevant in the relatively small portions of structure permitted by the PIC; it thus loses much of its original empirical coverage. Against the background of Brody's argument involving (weak or strong) representationality of derivational approaches, this can be viewed as further evidence that derivational approaches should dispense with the (G)MLC in toto. I would like to contend that, in a derivational approach, minimality effects should not be covered by a constraint that accesses a significant amount of syntactic structure, i.e., a representation, and then chooses between two items that may in principle participate in a given operation (as it is done by the (G)MLC). Rather, minimality effects should emerge as epiphenomena of constraints that reduce the space in which the derivation can look for items that may participate in an operation (as it is done by the PIC); ideally, all competition among items (that a priori qualify for some operation) that must be resolved is in fact independently resolved if the search space is sufficiently small.

Summing up so far, the (G)MLC turns out to be both empirically and conceptually problematic. It gives rise to empirical problems because it is both too strong (wrongly excluding certain legitimate cases of argument crossing) and too weak (wrongly permitting certain illegitimate cases of intervention effects). And it gives rise to conceptual problems in a derivational approach based on the PIC because it requires a substantial amount of search space (depending on rich representations), and creates redundancies (ruling out sentences that are also ruled out by the PIC). These conceptual considerations can now in fact be added to the list of meta-requirements on good constraints of chapter 1 as in (11-de) ((11-abc) are repeated from (39), (41), and (107) in chapter 1).

(11)   a.   Constraints should be as simple and general as possible.
        b.   Constraints should not be complex.
        c.   Constraints are of type (i) or (ii).
                (i)   principles of efficient computation

---

[8] Note that this problem with the (G)MLC persists even if one assumes a reinterpretation of this constraint suggested by an anonymous reviewer: "Once a probe [or a structure-building feature in the case at hand; GM] encounters a feature of a certain structural type, it cannot look further for another feature of the same structural type." Now $\gamma_{[F]}$ in (1) can be ignored for the evaluation of the (G)MLC (thereby dispensing with the original conceptual motivation for the (G)MLC); but search for $\beta_{[F]}$ by $\alpha_{[\bullet F \bullet]}$ is still in principle unbounded, and inherently dependent on massive search space.

       (ii)   interface conditions

    d.   Constraints do not require massive search space.

    e.   Constraints do not give rise to redundancies.

The (G)MLC satisfies (11-abc) but fails (11-de). In view of this, I would like to conclude that the (G)MLC should be discarded, and its effects derived from more basic assumptions that do not create conceptual problems like the ones just discussed. I will argue in chapter 3 that its effects can indeed be derived from a strengthened version of the PIC, which satisfies all of (11-a-e) (at least, it does so in the version that I will eventually adopt in chapter 3). For now, I will leave it at that, and turn to the fate of the CED in minimalist syntax.

3.   The Condition on Extraction Domain: State of the Art

3.1   Overview

The version of the CED assumed so far is repeated in (12) (see (97) of chapter 1).

(12)   *Condition on Extraction Domain* (CED):

    a.   Movement must not cross a barrier.

    b.   An XP is a barrier iff it is not a complement.

There is a general consensus that the CED cannot be maintained as such in minimalist syntax; in subsection 3.2, I argue that it is not compatible with the meta-requirements on good constraints listed in (11). After that, I discuss three kinds of minimalist (in an extended sense) approaches that strive to derive CED effects from independently motivated assumptions (subsections 3.3–3.5); all of these analyses can be shown to be problematic (from the current perspective at least). The end result will then be that the situation with the CED is exactly the same as with the (G)MLC: A proper minimalist reduction is still outstanding.

3.2   Problems with the CED

The CED is a simple, general constraint, and it is local, i.e., not complex (in the sense of (11-a), (11-b)). However, the CED does not seem to qualify as a principle of efficient computation, and it does not seem to be interpretable as an interface condition in any obvious way. Thus, it is not compatible with requirement (11-c). Furthermore, in the formulation in (12), it looks as though the CED relies on concepts that lack independent motivation, viz., the notion of barrier. However, as noted above, closer scrutiny of (12) reveals that this is not actually the case: The notion of barrier here is just another word for the notion of non-complement XP, which is fairly basic (if not completely primitive). Finally, as far as (11-d) and (11-e) are concerned, we can note that the CED does not presuppose massive search space, and that it does not automatically lead to redundancies with any other well-established constraint.

Still, in view of the incompatibility of the CED with basic minimalist tenets, attempts have been made to derive (some version of) this constraint (or its effects). For concreteness, three kinds of analyses can be distinguished. In a first type of approach, CED effects are derived by invoking assumptions about *elementary operations like Merge and Agree*. Analyses of this kind are Sabel (2002) and Rackowski & Richards (2005). In a second type of approach, specific assumptions about *cyclic spell-out* are invoked; see Uriagereka (1999), Nunes & Uriagereka (2000), and Nunes (2004) for one basic line of research, and Johnson (2003) for another one. A third type of approach strives to derive CED effects as instances of freezing effects; see Kitahara (1994), Takahashi (1994), Boeckx (2003), Gallego & Uriagereka (2006), and Stepanov (2007) (and Rizzi (2006; 2007) for a somewhat narrower concept of freezing that does not derive CED effects). I will argue that the first kind of analysis relies on special assumptions that mimic assumptions in Chomsky's (1986a) theory of barriers; that the second kind of analysis is incompatible with the assumption that only the complement of a phase head is affected by spell-out, whereas the specifier domain and the head itself remain available for further operations on subsequent cycles; and that the third kind of analysis is incompatible with the existence of CED effects where an XP is a barrier in its in situ position. Perhaps most importantly, all these approaches have nothing to say about *melting* effects, a class of data that I will introduce in subsection 3.6, and that I will argue to corroborate the analysis to be developed in chapter 4.

I will now address the three kinds of analyses in turn, beginning with those that rely on special assumptions about elementary operations.

3.3   Elementary Operations

3.3.1   Merge in Sabel (2002)

In Sabel's (2002) approach, extraction from subjects and adjuncts is argued to be impossible because these items are not merged with a lexical head, and a required S-projection cannot be formed. Sabel starts out with the version of the CED in (13) (which he refers to as "Barrier", and which is for all intents and purposes identical to the CED adopted here so far). (13) is assumed to follow as theorem: "I will argue [...] that [(13)] is motivated by θ-theoretic considerations" (Sabel (2002, 292)).

(13)   *Condition on Extraction Domain* (CED; Sabel's version):
A category A may not be extracted from a subtree T_2 (X^{max}) of T_1 if T_2 was merged at some stage of the derivation with a complex category (i.e., with a non-head).

He states (p. 295): "I assume that transparency and barrierhood in the case of CED-islands is a consequence of θ- (or [...] 'selection'-) theory." The crucial assumptions about the elementary operation Merge that prove necessary are given in (14).

(14)   a.   Head/complement Merge results in co-indexing; it establishes a selec-
              tional (superscript) index on a head and its complement (head/specifier
              Merge does not, and adjunction does not create co-indexing either).
        b.   Selectional indices are projected from a head to its XP.

Thus, a complement shares a superscript with the minimal XP dominating it; a spec-
ifier (or adjunct) does not. This becomes relevant for the definition of an additional
concept called *S(election)-projection*; see (15).[9]

(15)   *Selection-Projection*:
        X heads the smallest projection containing α_n. Then Y is an S-projection of
        X iff
        a.   Y is a projection of X, or
        b.   Y is a projection of Z, where Z bears the same index as X.

It follows from (15) that S-projections stop at specifiers (and adjuncts) but continue
from complements upwards (since the complement bears the same index as the head
that it is a complement of). The representational locality constraint *Uniform Domain*
that replaces the CED capitalizes on this difference; see (16).

(16)   *Uniform Domain* (UD):
        Given a nontrivial chain CH $= <\alpha_i,...,\alpha_n>$ with n>1, there must be an X such
        that every α is included in an S(election)-projection of X.

Since UD is representational, it cannot be a constraint on movement as such; it must
be a constraint on the resulting movement chain. Consider the effects of UD on such
a movement chain. First, as soon as the path from a base position to a (final) landing
site of movement includes a specifier or adjunct, propagation of the original selec-
tional index stops. Second, if selectional index transmission stops, the S-projection
ends, because of (15-b). Third, in such a case, there can therefore not be some node X
anymore such that every member of the movement chain is included in an S-projection
of X. Members of the movement chain will invariably belong to the dominance do-
mains of more than one S-projection. Fourth, UD is thus violated, and CED effects
are derived for all kinds of non-complements.

    Consider now how this approach fares with respect to basic minimalist principles,
and against the background of the meta-requirements listed in (11). The first thing to
note is that Uniform Domain does not seem to qualify as either an economy constraint
(it does not per se contribute to efficient computation any more than the original CED
does; see Gärtner & Michaelis (2007)) or an interface condition (there is no theory

---

[9] S-projections bear some resemblance to *g-projections* as defined in Kayne (1984, 167); cf. footnote 57 of
chapter 1.

of semantic interpretation that requires all members of a complex chain to show up in a single S-projection). Secondly, the analysis is incompatible with the concept of a phase (Chomsky (2000; 2001b; 2008)) because checking whether UD is violated or not requires scanning portions of syntactic structure that are much larger than what is permitted under the PIC. Third, the analysis requires special concepts that do not seem to be needed otherwise (S-projection, selectional indices), and that are explicitly prohibited as violations of the Inclusiveness Condition in recent minimalist work (see, e.g., Chomsky (2001b; 2005; 2008)). Finally, it is worth noting that the whole analysis ultimately rests on one central assumption, viz., (14-a), which is not independently motivated. (14-a) introduces a difference between head/complement and head/specifier (or head/adjunct) relations. Thereby, it basically mimicks the concept of L-marking, which is used to define barriers in Chomsky (1986a) (the analogon of clause (b) of (12)).[10]

We may thus conclude that there is good reason to abandon Sabel's (2002) analysis from the perspective of a minimalist, phase-based approach that embraces the meta-requirements on constraints in (11).

3.3.2   Agree in Rackowski & Richards (2005)

Whereas Sabel's analysis relies on specific assumptions about the elementary operation Merge, Rackowski & Richards (2005) derive (a version of) the CED by invoking special assumptions about Agree. Taking the framework of Chomsky (2000; 2001b) as a point of departure, Rackowski & Richards (2005) make the following asumptions about movement.

(17)   a.   A probe must Agree with the closest goal that can move.
       b.   A goal α can move if it is a phase (CP, vP, and DP are phases).
       c.   A goal α is the closest one to a probe if there is no distinct goal β such that for some X (X a head or a maximal projection[11]), X c-commands α but not β.
       d.   Once a probe P is related by Agree with a goal G, P can ignore G for the rest of the derivation.
       e.   v has a Case feature that is checked via Agree. It can also bear EPP-

---

[10] In Chomsky (1986a, 24), L-marking is defined as follows:

(i)   *L-marking*:
      Where α is a lexical category, α L-marks β iff β agrees with the head of γ that is θ-governed by α.

Simplifying a bit, non-L-marked XPs count as blocking categories, and most blocking categories then also qualify as barriers.

[11] Note that X$'$ categories must not count.

> features that move active phrases to its edge.
> f.   [+wh] C has a [+wh] feature that is checked via Agree (and sometimes Move).

The version of the CED that can be derived under these assumptions is given in (18).

(18)   *Condition on Extraction Doman* (CED; Rackowski & Richards' version):
Only those CPs and DPs that Agree with a phase head on independent grounds (e.g., direct objects and complement clauses) are transparent for *wh*-extraction.

To see how the system based on (17) works (and, more specifically, how (18) is accounted for), consider the simple instance of successive-cyclic long-distance *wh*-movement from an object CP in (19).

(19)   [$_{CP}$ Who do you [$_{vP}$ think [$_{CP}$ that we should [$_{vP}$ hire –]]]] ?

Given the assumptions in (17), the derivation of (19) must proceed as sketched in (20).[12]

(20)   a.   [$C_{[+wh]}$ [v [C [v *who*]]]]
       b.   [$C_{[+wh]}$ [v [C [v_y *who$_y$*]]]]   (Agree v-DP)
       c.   [$C_{[+wh]}$ [v [C [*who$_y$* v_y ~~*who$_y$*~~]]]]   (Move DP)
       d.   [$C_{[+wh]}$ [v [C [*who* v ~~*who*~~]]]]   (Forget Agree)
       e.   [$C_{[+wh]}$ [v_z [C_z [*who* v ~~*who*~~]]]]   (Agree matrix v-CP)
       f.   [$C_{[+wh]}$ [v [C [*who* v ~~*who*~~]]]]   (Forget Agree)
       g.   [$C_{[+wh]}$ [v_x [C [*who$_x$* v ~~*who*~~]]]]   (Agree matrix v-DP)
       h.   [$C_{[+wh]}$ [*who$_x$* v_x [C [~~*who$_x$*~~ v ~~*who*~~]]]]   (Move DP)
       i.   [$C_{[+wh]}$ [*who* v [C [~~*who*~~ v ~~*who*~~]]]]   (Forget Agree)
       j.   [$C_{[+wh]w}$ [*who$_w$* v [C [~~*who*~~ v ~~*who*~~]]]]   (Agree C-DP)
       k.   [*who$_w$* $C_{[+wh]w}$ [~~*who$_w$*~~ v [C [~~*who*~~ v ~~*who*~~]]]] (Move DP)
       l.   [*who* $C_{[+wh]}$ [~~*who*~~ v [C [~~*who*~~ v ~~*who*~~]]]]   (Forget Agree)

The first (relevant) operation (in (20-b)) is Agree of v and DP in (what will become) the embedded clause; the Agree operation is indicated here by co-indexing (a mere mnemotechnical device without theoretical significance). Based on this Agree relation, DP can next move to a specifier of v (see (20-c)). After this, the Agree relation is forgotten about, in accordance with (17-d) (see (20-d)). Subsequently, in (20-e), matrix v undergoes Agree with the head of the embedded CP. This is the crucial step,

---

[12] Two remarks. First, Rackowski & Richards (2005, 283) "sketch the derivation as though movement begins once the tree has been completed", and I follow them here. Second, whereas (20) is basically copied from Rackowski & Richards (2005), I have inserted 'Forget' operations (cf. (17-d)) that are not explicitly marked in the original paper but may serve to simplify exposition.

and it warrants a bit of further discussion.

For one thing, it is clear that given the minimality requirement in (17-c) (a version of the (G)MLC, including specifically its F-over-F part), the *wh*-phrase can only leave its CP and move on, to a specifier of v in the matrix clause, if v has first undergone Agree with this CP: Otherwise CP would qualify as an illegitimate intervener. For another, given the precise formulation of (17-c), vP does not count as intervener for movement of the *wh*-phrase even though it dominates it in the pre-movement structure. The reason for this is that there is no X^0 or XP category that c-commands the *wh*-phrase in (20-e) but fails to c-command vP: The only item that does c-command the *wh*-phrase without c-commanding vP is v′, which is neither an X^0 nor an XP category. Thus, we can conclude that extraction from a CP is possible only if matrix v can independently undergo Agree with C (Rackowski and Richards argue that this can be motivated by the assumption that clauses generally need case); in contrast, there does not have to be an Agree relation involving the embedded vP to make extraction possible. (As we will see momentarily, the analysis ultimately depends on a stronger assumption: There must not be an Agree relation involving the embedded vP.)

As shown in (20-f), Agree between matrix v and C can then be forgotten about; after that, matrix v can undergo Agree with the *wh*-phrase in the embedded Specv, and attracts it to its specifier, following which Agree is again forgotten (see (20-ghi)). Now Agree of matrix $C_{[+wh]}$ is possible with matrix vP, or with *who* (again, by stipulation – see (17-c) –, v′ does not count, and vP thus does not intervene). Choosing the latter option (see (20-j)), $C_{[+wh]}$ attracts the *wh*-phrase, creating SpecC, and the Agree relation can be ignored again (see (20-kl)). Note that the analysis assumes (contrary to most work on successive-cyclic movement) that *wh*-movement does not use the specifier of an embedded $C_{[-wh]}$ as an escape hatch; only the final $C_{[+wh]}$ head triggers movement of the *wh*-phrase to its specifier. Thus, on this approach there are no intermediate traces in SpecC positions.

Given this account of how movement from object CPs can apply in accordance with all constraints of grammar, the question arises of what goes wrong with subject CPs (or adjunct CPs), i.e, how CED effects are derived. The answer given by Rackowski & Richards is this: Subjects, like adjuncts, do not enter into an Agree relation with v: v cannot probe into its own specifier, given the c-command requirement on Agree. (Of course, to derive the Adjunct Island effect, it must then be assumed that adjuncts are located in positions outside the c-command domain of v, too.) Thus, if the (embedded) CP in (20-e) occupies a specifier (rather than complement) position of v, Agree(v-CP) will not be an option, and the CP will continue to block any Agree (and, therefore, Move) relation involving the *wh*-phrase throughout the remainder of

the derivation, because of the minimality requirement in (17-c).[13]

This concludes the sketch of Rackowski & Richards's (2005) approach to CED effects.[14] There are various potential problems with this analysis, both empirical and conceptual. An empirical problem is that the analysis does not permit successive-cyclic movement to take place via embedded SpecC positions. The counter-evidence from partial *wh*-movement in languages like German (see (21-a)) may be explained away by assuming an indirect dependency approach (cf. Dayal (1994)) according to which there is no movement(-like) relation between the proper *wh*-phrase and a scope marker in a higher SpecC position. However, such a way out is not available for the closely related *wh*-copy construction (see (21-b)), which is substandard but possible for a number of German speakers. Here it seems difficult to deny that there is a movement relation involved; and the lower copy most clearly shows up in a position where Rackowski & Richards' analysis does not permit it to show up, viz., the embedded SpecC position.

(21) a. Was meinst du [$_{CP}$ wen$_1$ wir t$_1$ einladen sollen] ?
          what think you   whom we   invite   should

      b. Wen$_1$ meinst du [$_{CP}$ wen$_1$ wir t$_1$ einladen sollen] ?
          whom think you   whom we   invite   should

One might think that this problem can simply be solved in Rackowski & Richards' approach by assuming intermediate steps to an embedded SpecC position to be an option in the course of long-distance movement. However, if this were possible, the account of CED effects with subject CPs would be undermined. This is shown in the abstract derivation in (22). Here, the embedded CP is base-merged in a specifier of the matrix v. In the embedded vP, things are exactly as before; see (22-abcd). However, in the next steps (see (22-ef)), declarative C undergoes Agree with the *wh*-phrase and attracts it to its specifier position (exactly as it is standardly assumed). Given (17-c), this implies that CP does not count as an intervener anymore, even if there is no Agree relation targeting it; thus, there is nothing that could stop further movement of the *wh*-phrase from the subject clause to the matrix SpecC position in (22).

(22) a. [C$_{[+wh]}$ [[C [v *who*]] v]]

---

[13] Rackowski & Richards (2005, 585) contemplate the possibility that, whereas Agree is standardly assumed to be confined to c-command contexts (see Chomsky (2001b; 2005; 2008)), some languages might permit Agree under *m-command* (i.e., including the option of affecting a specifier and its head) after all. If so, subjects are predicted to be transparent for extraction, an option that according to Rackowski & Richards might account for the empirical evidence in Japanese. I will return to this issue of (seemingly) legitimate violations of the Subject Condition in various languages below, and in chapter 4.

[14] The reasoning has been carried out on the basis of extraction from CP, but the analysis immediately carries over to DPs.

b. $[C_{[+wh]} [[C [v_y \; who_y]] v]]$ (Agree v-DP)

c. $[C_{[+wh]} [[C [who_y \; v_y \; \sout{who_y}]] v]]$ (Move DP)

d. $[C_{[+wh]} [[C [who \; v \; \sout{who}]] v]]$ (Forget Agree)

e. $[C_{[+wh]} [[C_z [who_z \; v \; \sout{who}]] v]]$ (Agree C-DP)

f. $[C_{[+wh]} [[who_z \; C_z [\sout{who_z} v \; \sout{who}]] v]]$ (Move DP)

g.   ...

Thus, it can be concluded that the stipulation that an intermediate C does not undergo Agree with (and therefore cannot attract via an EPP-feature) a *wh*-phrase is indispensable in Rackowski & Richards (2005) analysis.

Incidentally, there is another potential loophole that must also be closed, and that can only be closed in a stipulative way: Given the c-command requirement on Agree, v cannot undergo Agree with a CP that is its specifier. However, what about other functional categories of the matrix clause, like C or T? Suppose that C (or T) would be able to undergo Agree with an embedded subject CP. C (or T) would then first carry out Agree with a subject CP; this would not violate the minimality requirement in (17-c) because vP could not act as an intervener (CP is in v's specifier). After this, an item in the subject CP could undergo Agree with matrix C (or T). Then, extraction from a subject would be possible after all even without prior movement of the *wh*-phrase to the embedded SpecC position: The barrier/intervention status of the subject CP has now been removed by an Agree relation with the same item that attracts the *wh*-phrase. To close this second loophole, again, an additional stipulation seems necessary that explicitly excludes Agree relations between C or T and a subject CP.[15]

From a more general point of view, I take these considerations to indicate a general shortcoming of the approach: The most important assumption needed to derive CED effects is that extraction from XP requires an Agree operation involving v and XP: Since v can Agree with (something in) its complement but not with its specifier or an adjunct, the latter two types of categories are derived as barriers. However, it seems clear that the explicit "c-command by v" *requirement* does little more than mimicking the L-marking requirement of Chomsky (1986a). In addition, the restriction to v (vs. T, C) as a possible probe for Agree relations with CP is also very similar to the restriction to lexical categories that is part of the definition of L-marking. Finally, it can be observed that the approach currently under consideration gives rise to a curious asymmetry: For the purposes of minimality, the vP-v′ distinction must be ignored;

---

[15] This restriction is all the more peculiar since T *does* regularly undergo Agree with subjects. Exempting subject clauses from Agree with T (in favor of default agreement) seems a priori superfluous and stipulative. – That said, the mechanism in (17) may plausibly be interpreted in such a way that it is not an Agree relation per se that makes an XP transparent; rather, Agree of XP with Y makes items in an XP accessible for Agree with Y. On this interpretation, only Agree of C and CP must be excluded.

however, for the purposes of deriving the CED, this distinction must be maintained.

To sum up so far, I have argued that Sabel's (2002) and Rackowski & Richards's (2005) approaches to CED effects in terms of specific assumptions about elementary operations (Merge, Agree) are both problematic, each in a number of ways. Arguably, though, there is a single most pressing conceptual problem for both analyses that arises under the present perspective, and that is the fact that it is unclear in what sense one can say that the central assumptions of the two approaches go beyond what is encoded in Chomsky's (1986a) concept of L-marking, which does not meet minimalist requirements. This objection does not apply to approaches to the CED that rely on special assumptions about the nature of spell-out. I turn to these approaches in the next subsection.

3.4   Spell-Out

Uriagereka (1999), Nunes & Uriagereka (2000), Nunes (2004), Johnson (2003), and, to some extent, Stepanov (2007) all present accounts of CED effects that rely on the idea that the classification of an XP as opaque or transparent for extraction can be correlated with an independently assumed difference with respect to spell-out. What is more, the assumption is that spell-out properties associated with certain XP types directly account for the barrier status of the XP in question. Apart from that, the three kinds of approach that can be distinguished in this general group can be shown to differ substantially. I begin with the cyclic spell-out analysis developed by Uriagereka and Nunes, turn to Johnson's theory after that, and finally address Stepanov's approach.

3.4.1   Cyclic Spell-Out: Uriagereka (1999), Nunes & Uriagereka (2000), Nunes (2004)

In a number of articles, Juan Uriagereka and Jairo Nunes have developed the hypothesis that (some version of) the CED can be derived by invoking specific (though independently motivated) assumptions about cyclic spell-out. Throughout this subsection, I focus on the original version of the approach laid out in Uriagereka (1999).

Uriagereka's central goal is to derive a version of Kayne's (1994) Linear Correspondence Axiom (LCA) from minimalist assumptions; and the main theoretical innovation that he argues for in the course of this attempt is the concept of multiple (i.e., cyclic) spell-out in a derivation, which has since (in a slightly different form; see below) become standard in minimalist approaches. The original LCA given in Kayne (1994) has the following consequences for the positioning of heads and specifiers (as well as for the nature of specifiers):[16]

---

[16] The original LCA from which (23) follows reads as follows.

(i)   *Linear Correspondence Axiom* (LCA; Kayne (1994)):

(23)   *Consequences of the LCA*:
     a.   A head precedes its complement (β).
     b.   A specifier (α) must formally qualify as an adjunct. It is unique and precedes its head.

As a result, all phrases must take the form depicted in (24) under the LCA.

(24)   [$_{XP}$ α [$_{XP}$ X β]]

Whereas Kayne's (1994) LCA restricts possible phrase markers, the reconstruction of the LCA in Chomsky (1995) restricts possible linearizations of a priori unordered phrase markers at PF; these phrase markers consist of heads, complements, and any number of (genuine, i.e., non-adjunct) specifiers of a given head (plus, irrelevantly for present purposes, adjuncts). Thus, on Chomsky's view, the LCA ensures linearization of trees in a bare phrase structure model. Chomsky's (1995) version of the LCA looks as in (25); it is a recursive definition composed of a base step and an induction step.

(25)   a.   Base step: If α c-commands β, then α precedes β.
     b.   Induction step: If γ precedes β and γ dominates α, then α precedes β.

It can be noted that (25-b) is essentially the Nontangling Condition (see Partee et al. (1993, 437)); a standard version of this condition is given in (26).

(26)   *Nontangling Condition*:
     In any well-formed constituent structure tree, for any nodes x and y, if x pre-

---

     d(A) is a linear ordering of T.

A linear ordering of terminal symbols is defined as in (ii); d(X) (the image of X under d) is understood as in (iii); and the set A of pairs of nodes for which asymmetric c-command holds is defined as in (iv).

(ii)   *Linear ordering* of terminal symbols (L):
     a.   transitive: $\forall x,y$: $<x,y> \in L \wedge <y,z> \in L \rightarrow <x,z> \in L$
     b.   total: $\forall x,y$: $<x,y> \in L \vee <y,x> \in L$
     c.   antisymmetric: $\forall x,y$: $\neg(<x,y> \in L \wedge <y,x> \in L)$

(iii)   a.   D = dominance relation between non-terminal symbols
     b.   d = dominance relation between non-terminal and terminal symbols
     c.   d(X) = set of terminal symbols that are dominated by a non-terminal X (the 'image' of X under d)
     d.   d$<X,Y>$ (image of non-terminal $<X,Y>$ under d) =
         $\{<a,b>\}$: $a \in d(X) \wedge b \in d(Y)$
     e.   Let S be a set of ordered pairs $<X_i,Y_i>$ ($0<i<n$). Then:
         d(S) = $\bigcup$ for all i ($0<i<n$) of d($<X_i,Y_i>$)

(iv)   a.   A = $\{<X_j,Y_j>\}$, such that for each j: X_j c-commands Y_j asymmetrically
     b.   T = set of terminal symbols of a phrase structure tree P

cedes y, then all nodes dominated by x precede all nodes dominated by y.

Adopting Chomsky's (rather than Kayne's) perspective on the LCA, Uriagereka (1999) sets out to derive both the base step and the induction step in (25) – i.e., the LCA in toto.

Deducing the Base Step of the LCA   Assuming a partition of phrases into head, complement, and specifier, there are a priori *n*! ways to "lay the mobile on the ground" (as Uriagereka puts it). The question then is why, out of the six possible orders that result if there is a unique specifier, it is the Spec–Head–Comp order that is chosen (assuming the validity of the LCA). Two orders are immediately excluded: If the specifier intervenes between head and complement, as in (27-ef), the Nontangling Condition (or whatever derives it) is violated. But what about the remaining orders in (27-abcd), out of which only (27-d) can, by assumption, be chosen?

(27)   a.   Comp Head Spec
        b.   Head Comp Spec
        c.   Spec Comp Head
        d.   Spec Head Comp   (optimal order, given LCA)
        e.   Comp Spec Head   (*Nontangling)
        f.   Head Spec Comp   (*Nontangling)

Uriagereka ventures the hypothesis that there is an order of Merge operations (a "Merge-wave of terminals"), and that the most economical way to map phrase structure onto linear order is to "harmonize (in the same local direction) the various wave states, thus essentially mapping the merge order into the PF linear order in a homomorphic way." The claim is that it follows from these assumptions that the orders in (27-b) and (27-c) are excluded. This reasoning may or may not be viewed as convincing (Uriagereka acknowledges that the approach relies on "hand-waving until one establishes what such a Merge-wave is", which he does not set out to do), but let us assume for the sake of the argument that it is. Then, the question remains why the command relation collapses into precedence, and not the opposite ((27-d) vs. (27-a)). The answer that Uriagereka gives is that this may not really be a problem on closer inspection because what is needed is just *some* optimal solution; there may not be *the* optimal solution. Against this background, let us now consider how the induction step of the LCA in (25) can be derived.

Deducing the Induction Step of the LCA   The central innovation that Uriagereka (1999) proposes is the concept of a command unit (CU), which can be defined as in (28).

(28)   *Command Unit* (CU):
        A command unit emerges in a derivation through the continuous application of Merge to the same object.

A command unit arises if one of the two elements involved in a Merge operation is a lexical item taken directly from the numeration; as soon as both elements are not lexical items (but have internal structure resulting from previous applications of Merge), there is no way to assume the whole structure to go back to a contiuous application of Merge to one and the same object. This distinguishes between complements (i.e., first-merged items) and specifiers (i.e., non-first merged items). A complement YP of some head X may in principle have rich internal structure arising from successive application of Merge (as long as it does not have a complex internal specifier at any point), and Merge applying to X and YP then enlarges the command unit but does not have to create a new one. In contrast, if some item ZP becomes a specifier of some head X after Merge of X and ZP, then the two items involved will invariably be two separate command units as long as the specifier has (a non-trivial) internal structure.

On this basis, Uriagereka (1999) sets out to derive the Nontangling Condition. Assuming that the induction step of the LCA in (25) is not explicitly stipulated, the question arises of how linearization works in derivations with more than one command unit (i.e., derivations with complex specifiers). Uriagereka's answer is that there are various steps of linearization, each of which involves only command units; this is the idea of multiple spell-out. To implement the idea, he devises two approaches, which can be dubbed the "conservative approach" and the "radical approach". According to the conservative approach, a collapsed Merge structure is no longer phrasal after spell-out; it is more like a giant lexical compound, or a word. According to the more radical approach, a spelled-out command unit does not merge with the rest of the structure; interphrasal association is accomplished in the performative components. In what follows I concentrate on the conservative account so as to simplify matters and keep the focus on what we are really concerned with in the present context (viz., CED effects). Thus, the core idea is that spell-out must apply as soon as a maximal command unit is reached. What spell-out does is flatten the hierarchical structure of a phrase structure tree created by previous Merge operations. As a result of this, the command unit is not a syntactic object (in the technical sense made precise in Chomsky (1995)) anymore. Its internal structure cannot be affected by syntactic operations anymore because, after spell-out, there *is* no internal structure left.

CED Effects   It should be obvious how this set of assumptions accounts for the barrier status of specifiers, and thus for (the main bulk of) CED effects. Suppose that we want to derive the version of the CED in (12): Movement must not take place from an XP that is not a complement. As noted by Uriagereka (1999), the multiple spell-out model based on command units predicts that "if a non-complement is spelled out independently from its head, any extraction from a non-complement will involve material from something that is not even a syntactic object; thus, it should be as hard as extracting part of a compound." Given that complements are not separate spell-out

domains, such a restriction is not active: At the point where movement takes place, the internal structure of the complement is still acessible because spell-out did not have to apply earlier. Note that this account leaves one case of specifiers that are predicted not to create separate command units (hence, no separate spell-out domains): If the specifier is a single lexical item (e.g., a D pronoun like *she*), it should simply extend the command unit generated thus far. However, since there is no extraction from a non-complex specifier almost by definition, this consequence is unproblematic.

As observed by Uriagereka (1999), a potential problem does arise in the form of sentences involving subject extraction, as in (29).

(29)   Which woman$_1$ did you say [t$_1$ left] ?

Here it looks as though at least the relevant feature information of the *wh*-subject that makes it visible to the operation of *wh*-movement must still be accessible after spell-out; Uriagereka confines himself to stating that "the answer to this puzzle relates to the pending question of *wh*-feature accessibility in spelled-out phrases."[17]

This may suffice as a sketch of how CED effects are derived from assumptions about cyclic spell-out in Uriagereka (1999).[18] The analysis is elegant, but there are problems. Most importantly, it can be noted that Uriagereka's (& Nunes') approach is fundamentally incompatible with the notion of a *phase* as the relevant domain for cyclic spell-out (Chomsky (2000; 2001b; 2008)). Spell-out domains that correspond to command units in the sense of (28) are variable in size. One the one hand, they may be larger than the spell-out domain of a phase – in fact, extremely large (possibly, they may span a whole sentence): As long as no complex specifier is merged (either no specifier, or specifiers consisting only of lexical items), a new spell-out domain will not be created. On the other hand, spell-out domains can be smaller than the spell-out domain of a phase: E.g., a complex specifier (belonging to any category) is always a spell-out domain. Given that phases are mainly motivated by economy considerations in Chomsky (2000; 2001b; 2008), as derivational units that serve to minimize search space, such an enormous variability in the definition of spell-out domains makes the approach currently under consideration radically incompatible with the phase model that I adopt throughout. What is more, Uriagereka's and Nunes' model is incompatible with yet another basic tenet of the phase-based approach to syntax developed in

---

[17] Interestingly, a similar problem shows up in the approach to displacement developed in Gazdar et al. (1985); see subsection 4.2.1 below for remarks on this approach.

[18] The analyses of CED effects in Nunes & Uriagereka (2000) and Nunes (2004) differ in minor respects and may have slightly different foci (e.g., in Nunes & Uriagereka (2000) the analysis is extended to parasitic gaps and their ability to circumvent CED violations, by adopting a sideward movement approach). However, they share the same basic logic.

Chomsky (2000; 2001b; 2008): Only the complement of a phase head is affected by spell-out whereas the specifier domain and the head itself remain available for further operations on subsequent cycles; the general flattening of command units in the Uriagereka/Nunes model makes it impossible to maintain such an option (hence, ultimately, the PIC).[19]

3.4.2 Renumeration: Johnson (2003)

An approach to CED effects that is in some respects similar to the one developed in Uriagereka (1999) in that it relies on assumptions about cyclic spell-out (but that differs in others, as we will see) is developed in Johnson (2003). The first important assumption that Johnson makes is that subjects must be adjuncts, exactly as in Kayne (1994) (i.e., there are no specifiers; see the previous subsection).[20] The second assumption that plays a fundamental role is that incremental phrase structure generation via Merge proceeds somewhat differently than standardly envisaged: Rather than being able to take two items from the numeration, Merge can only take one item from the numeration at a time (the "host"), and attaches this item to the tree created so far (if any), with the host projecting. The third important assumption is that there is an additional operation, called "Renumerate", which places a tree constructed by Merge back into the numeration. Why should such an operation be necessary? The answer is that there is an additional constraint ("If an X^0 merges with a YP, then YP must be its argument") that makes it impossible to combine, via Merge, a verb taken from the numeration directly with a complex adjunct created so far. Thus, the following representation cannot be generated by taking V from the numeration and attaching it, as a host, to the adjunct PP.

(30)   *[$_{VP}$ [$_V$ left] [$_{PP}$ after this talk]]

The only legitimate derivation is one where an adjunct PP like *after this talk*, after having been created by successive applications of Merge, undergoes the operation Renumerate, which transfers it back to the numeration. V (*left*, in the case at hand), may then be targetted by Merge with v, with v (as the host) projecting a vP, and Merge finally applying again to [$_{PP}$ *after this talk*], taking it from the numeration and attaching it to vP, yielding (31). In this derivation, the constraint that a minimal, bare

---

[19] For other potential problems with the approach, see Johnson (2003, 190).

[20] Accordingly, the version of the CED that Johnson sets out to derive is called the Adjunct Condition. His version of this constraint is given in (i); also compare (94) in chapter 1.

(i)   *Adjunct Condition* (Johnson's (2003) version):
      When a phrase's underlying position in a phrase marker is such that it is a sister to another phrase but doesn't project, it is an island for extraction.

X^0 category cannot be combined with an adjunct is respected.

(31)   [$_{vP}$ v [$_{VP}$ left]] [$_{PP}$ after this talk]]

More generally, it follows from Johnson's assumptions that all adjuncts (which, as noted, includes subjects) must be renumerated before they are merged. Evidently, this restriction does not hold for complements. It is this difference between adjuncts and complements that Johnson (2003) then exploits in his account of CED effects: Phrases that must renumerate are barriers. There is one final assumption that is needed to derive CED effects on the basis of the difference between complements and adjuncts with respect to renumeration. Johnson (p. 204) calls this assumption "Numerphology", and it reads: "Elements in the numeration get their syntax to phonology mapping values fixed."[21] This, then, accounts for the barrier status of adjuncts: Once an adjunct has undergone Renumerate, "Numerphology will force all of the terms within that adjunct to have their linear position fixed. As a consequence, every term within that adjunct must surface adjacent to some other term within that adjunct. Under the reasonable assumption that movement out of the adjunct would require the moved term to no longer be adjacent to material within the adjunct, this will preclude movement from the adjunct," as Johnson (2003, 209) puts it.

The first thing to note is that the basic idea is quite similar to that of Uriagereka (1999): Some XP is a barrier because independently assumed constraints force it to undergo some procedure (Renumerate in one case, establishing a complete command unit in the other), which then applies spell-out of the XP. A spelled-out XP is assumed to resist internal modification, which implies that movement operations cannot affect parts of it. However, a crucial difference between the two analyses is that in Johnson's case, it is the constraint against adjuncts as sisters of lexical items that is ultimately responsible for early (in fact, premature) spell-out, whereas in Uriagereka's case, it is the very existence of a complex non-complement.

Like Uriagereka's (1999) analysis, Johnson's (2003) proposal seems hardly compatible with a phase-based approach to locality along the lines of Chomsky (2000; 2001b; 2008). The postulation of an operation "Renumerate" is perhaps not an optimal concept from a conceptual point of view; it qualifies as inherently counter-cyclic. Furthermore, it is unclear whether the enormous amount of technical machinery that is needed to derive CED effects in this approach can be said to be justified, and compatible with minimalist principles as they were laid out above. Irrespective of this, the core assumption that only arguments can merge with a lexical item looks empirically problematic in view of languages like German, where certain kinds of undisputable non-

---

[21] Based on this set of assumptions, Numerphology can arguably be supported by evidence from focus projection (see Johnson (2003, sect. 3)). For reasons of space and coherence, I will not go into this here.

arguments regularly show up between arguments and the verb, and a base-generation of these items as sisters of the verb has often been proposed as by far the simplest solution (see, e.g., Frey & Tappe (1991); and also Larson (1988) for relevant discussion of non-arguments in English double object constructions). Similarly, the assumption that there is no phrase-structural difference between specifiers and adjuncts seems dubious (however, that said, the gist of Johnson's analysis could presumably be transferred without much ado to a framework in which specifiers and adjuncts are different kinds of items).

A further potential problem is that Johnson's (2003) analysis (in contrast to Uriagereka's analysis) actually permits a loophole for extraction from adjuncts: Since it only requires that items in an adjunct have their linear position with respect to each other fixed, by showing up "adjacent to some other term within that adjunct", the prediction is made that string-vacuous extraction of the left-most item from an adjunct should be unproblematic. Thus, for instance, string-vacuous *wh*-extraction from a subject in German in (32) should be an option; compare (32-a) (string-vacuous *wh*-movement to SpecC via *was für*-split from a subject in an embedded question) with (32-b) (non-string vacuous movement of the same type, this time from a subject in a matrix clause with a C that is overtly filled by verb-second movement). It is difficult to test whether the prediction that string-vacuous extraction from an adjunct (in Johnson's terminology) should be fine is empirically corroborated or not because examples of the type in (32-a) can always have an alternative representation with the *wh*-item in situ, within a pied-piped subject DP, due to the very fact that string-vacuous movement is involved. However, it seems fair to conclude that such an amelioration effect would be a priori unexpected: If it could be shown to exist, this would be a spectacular finding.[22]

(32)   a. (*)(Ich weiß nicht) [$_{CP}$ was$_1$ [$_C$ Ø] [$_{DP}$ t$_1$ für Leute] dem Peter Bücher
            I   know not   what   for people the   Peter books
        schenkten
        gave

---

[22] Note however that Heck (2008, 116 & 242) suggests that such an effect might indeed obtain with certain pied piping constructions. – One might hope that choice of different phonological (e.g., intonational) patterns might distinguish the *was*-in-situ/full DP movement version of (32-a) from the one employing string-vacuous movement of *was*, which would make it possible to test the different predictions after all. However, it is not clear that there is a phonological pattern that might signal that extraction has taken place (among other things, pauses are of course also compatible with a non-movement derivation). Furthermore, from Johnson's point of view, one might argue that intonational changes would fatally disrupt the integrity of the renumerated subject in much the same sense that lack of adjacency does.

  b.   *[CP Was$_1$ [C schenkten$_2$ [C Ø]] [DP t$_1$ für Leute] dem Peter Bücher
         what   gave   for people the Peter books
       t$_2$] ?

Similarly, it seems that empty operator movement from subjects (and adjuncts) is
predicted to be unproblematic (since it cannot alter adjacency relations among items
reducible to spell-out), contrary to fact (cf. Chomsky (1977), Browning (1987)); see,
e.g., the minimal pairs in (33).[23]

(33)   a.   This is a man [CP Op$_1$ I saw [DP a picture of t$_1$]]
       b.   *This is a man [CP Op$_1$ [DP a picture of t$_1$] was published]
       c.   John is easy for us [CP Op$_1$ to convince Bill to do business with t$_1$]
       d.   *John is easy for us [CP Op$_1$ to convince Bill to leave [when he meets t$_1$]]

To sum up so far, Johnson's (2003) analysis, like Uriagereka's (1999), is sufficiently
problematic (at least under the basic set of assumptions that I adopt in this monograph)
to look for an alternative account of CED effects.

3.4.3   Late Adjunct Insertion: Stepanov (2007)

Stepanov (2007) argues for a heterogeneous approach to CED effects that distin-
guishes between the Subject Condition and the Adjunct Condition (also see Frank
(2002)). For the former, he adopts a version of the freezing approach; I will address
freezing approaches in some detail below. For the latter, he suggests that *late, counter-
cyclic insertion* of adjuncts (as it has sometimes been proposed to derive certain anti-
reconstruction effects with principle C; see Lebeaux (1988), Freidin (1994), Chomsky
(1995), Epstein et al. (1998), Fox (2000)) ultimately provides an explanation: At the
point where extraction takes place, the adjunct is not yet part of the structure. The
effect is similar to the one occurring with subjects and adjuncts in Uriagereka's and
Johnson's approaches: XPs are islands because they are not present as transparent,
modifiable syntactic objects at the relevant stage of the derivation where movement
takes place. The only difference is that they are not accessible *anymore* when move-
ment applies in the first kind of analysis, and not accessible *yet* when movement ap-
plies in the second kind of analysis. Evidently, this approach to the Adjunct Condition
depends on adopting the late insertion hypothesis for adjuncts. Given that there is
both strong empirical and conceptual evidence against late insertion of adjuncts, and
given that alternative theories that account for the core data are readily available (see
Chomsky (2001a), Fischer (2004, ch. 3 & 5), and references cited in the latter), I think

---

[23] Also compare Heck (2008, 70) (and Heck (2009, 99)) on the consequences that allowing phonologi-
cally empty items to violate island constraints would have for analyses of pied piping that rely on feature
percolation.

that such a radically acyclic operation should be dispensed with. If so, the Adjunct Condition part of the CED cannot be derived in this way.

3.5   Freezing

Freezing approaches have been developed by Kitahara (1994), Takahashi (1994), Boeckx (2003), and Stepanov (2007). All these analyses presuppose that (relevant subcases of) CED effects can be traced back to freezing effects; i.e., an item becomes a barrier after movement has taken place. They differ in what is taken to be responsible for the occurrence of freezing. I will now go through these approaches one by one, beginning with Kitahara (1994).

3.5.1   Freezing and Head Movement: Kitahara (1994)

The first freezing approach to be discussed here is actually not typical: The freezing effect does not arise from movement of the XP from which extraction is to take place (as in the analyses to be discussed below), but rather from movement of a head to the domain in which XP is located.[24]

Kitahara's (1994) goal is fairly modest: He strives to provide a minimalist *reformulation* of the CED that makes the following predictions: First, a complement is never a barrier. Second, an adjunct is always a barrier. And third, a specifier is a barrier only if its head has been the target of head movement. To this end, the CED is replaced with the Inner Minimal Domain Requirement (IMDR) in (34).

(34)   *Inner Minimal Domain Requirement* (IMDR):
       Extraction out of a category K is possible only if for every X^0-chain H such that K $\in$ the minimal domain of H, K $\in$ the inner minimal domain of H.

The concept of minimal domain that is relevant in (34) is understood as in Chomsky (1993), on the basis of the concept of domain; see (35-ab). Essentially, a minimal domain is a flat version of the richer notion of a domain; the central requirement ensuring this is the reflexive domination condition in (35-b).

(35)   *Domain, minimal domain*:
       For any X^0-chain CH $<\alpha_1,...,\alpha_n>$:
       a.   the domain of CH = the set of nodes (i.e., categories) contained in the least full-category maximal projection dominating α_1 that are distinct from and do not contain any α_i.
       b.   the minimal domain of CH = the smallest subset K of the domain of CH such that for any $\Gamma \in$ the domain of CH, some $\beta \in$ K reflexively domi-

---

[24] This is in fact a bit more in the spirit of the proposal in Wexler & Culicover (1980); recall footnote 59 in chapter 1.

nates Γ.

The sensitivity to head movement that is part of the intended consequences of the IMDR is captured by the definition of inner minimal domains in (36).

(36)   *Inner minimal domain*:
        For any X^0-chain CH $<\alpha_1,...,\alpha_n>$:
        the inner minimal domain of CH = the (maximal) subset S of the minimal domain of CH such that each member of S is dominated by every maximal projection dominating α_i.

Let us see what predictions the IMDR in (34) makes for adjuncts, complements, and subjects. Consider the abstract configuration in (37) first.

(37)   [$_{XP}$ [$_{XP}$ Spec [$_{X'}$ X Comp]] Adj]

Adjuncts (Adj) are derived as barriers as follows: Given that (a possibly multi-segmental category) A *dominates* B if every segment of A dominates B whereas (a possibly multi-segmental category) A *contains* B if some segment of A dominates B (see Chomsky (1986a)), and given that minimal domains are defined in terms of containment whereas inner minimal domains are defined in terms of dominance, the adjunct in (37) is in the minimal domain of the (trivial) head chain X but not in the *inner* minimal domain of X; therefore, extraction from an adjunct is blocked by the IMDR. Complements (Comp) are different: The complement of X in (37) is both part of the minimal domain of X and of the inner minimal domain of X. The same goes for the specifier (Spec) in (37): Spec is part of the minimal domain of X since it is contained in XP; and since Spec is also dominated by XP, it is also part of the inner minimal domain of X. Hence, extraction from both complements and specifiers in (37) is predicted to be possible.

Consider next the abstract configuration in (38), where head movement has applied.

(38)   [$_{HP}$ Spec$_1$ [$_{H'}$ [$_H$ X$_i$] [$_{XP}$ Spec$_2$ [$_{X'}$ t$_i$ Comp]]]]]

The minimal domain of the non-trivial head chain $<X_i, t_i>$ includes Spec$_1$, Spec$_2$, and Comp (because all these items are contained in HP, the next full-category maximal projection that dominates X$_i$). Spec$_2$ and Comp are also part of this head chain's inner minimal domain but Spec$_1$ is not because Spec$_1$ is not dominated by every maximal projection dominating some member of the head chain: XP dominates the final member of the head chain but not Spec$_1$. More generally, it follows from (34) that a

specifier becomes opaque if head movement takes place to its head.[25]

The predictions that the analysis makes are thus clear enough; however, the empirical evidence brought forward by Kitahara (1994) in support of it is arguably not quite as straightforward. Kitahara postulates that extraction from subjects in English gives rise to results that are much less acceptable than comparable cases of extraction from subjects in Icelandic; see (39-a) vs. (39-b).

(39)   a.   *Who$_1$ do you think [$_{CP}$ that [$_{DP}$ pictures of t$_1$] are on sale] ?
      b.   ?Hverjum$_1$ heldur þú   [$_{CP}$ að   [$_{DP}$ myndir af t$_1$] séu til sölu] ?
          who   think you   that   pictures of   are on sale

The logic of the IMDR-based account would then lead us to conclude that a subject in English acts as a specifier that is targeted by head movement before extraction out of the subject can take place, wheras a subject in Icelandic permits extraction because its head is not targetted by head movement (at the relevant step of the derivation where extraction takes place). As a matter of fact, Kitahara (1994) assumes that in both cases, the subject DP moves to SpecAgr/S, which he takes to dominate TP (following Chomsky (1993) and much related work).[26] By assumption, in English, subject raising must *follow* T-to-Agr/S movement (for reasons having to do with case-checking). In contrast, in Icelandic, subject raising can *precede* T-to-Agr/S movement, and there is thus a legitimate order that respects the IMDR in Icelandic: First, subject raising to SpecAgr/S takes place; second, extraction from subject occurs; and third, T-to-Agr/S head movement applies, which turns the subject into a barrier by removing it from the inner minimal domain of the T chain. However, this final step comes too late in the derivation to block extraction, which has already taken place, legitimately.

This analysis may account for the difference in (39), but the assumptions about head movement that are required do not appear totally convincing. By standard assumption, English is a language without syntactic verb movement to T, whereas Icelandic has verb movement to T (or Agr); see Vikner (1995), among many others. This is (close to) the opposite of what is claimed by Kitahara (it is not a full contradiction since Kitahara talks about T movement, not V movement). Still, it is unclear whether strong independent evidence in favour of the postulated asymmetry in T-to-Agr/S movement can be brought forward.

---

[25] Note in passing that this is not per se incompatible with the view that head movement may in fact *remove* barriers (rather than just create them). The evidence that head movement opens up barriers is originally only concerned with *complements* (see Baker (1988), Sternefeld (1991b), Müller & Sternefeld (1993), Müller (1995); also see Bobaljik & Wurmbrand (2003), Gallego & Uriagereka (2006), and den Dikken (2007; 2008) for recent discussion, and below).

[26] Note that this movement step, by itself, does not create a barrier, under Kitahara's assumptions. It is irrelevant for determining barrier status.

Independently of this problem, as well as of various potential empirical problems, it can be noted that Kitahara's (1994) analysis does not seem to meet minimalist requirements: The IMDR neither contributes to efficient computation in an obvious sense, nor is it an interface requirement; it is no more and no less than an elegant reformulation of the CED, as Kitahara acknowledges. In addition, it employs concepts that do not seem independently motivated (dominance vs. containment, minimal domain, inner minimal domain, etc.).

Interesting and highly original though it is, Kitahara's approach does not seem to have been widely adopted. The case is different with the freezing approach that is developed by Takahashi (1994), and that is also argued for by Stepanov (2007).[27] This approach centers around the notion of chain uniformity.

3.5.2   Freezing and Chain Uniformity: Takahashi (1994), Stepanov (2007)

The central principle underlying Takahashi's (1994) account of CED effects is Chain Uniformity.

(40)   *Chain Uniformity*:
        Chains must be uniform.

For present purposes, we need not worry what exactly it means for a chain to be "uniform" in the sense of (40). The only thing that is important at this point is that the Uniformity Corollary on Adjunction (UCA) in (41) is supposed to be a by-product of (40) (see Takahashi (1994, 25)).

(41)   *Uniformity Corollary on Adjunction* (UCA):
        Adjunction is impossible to a proper subpart of a uniform group, where a uniform group is a non-trivial chain or a coordination.

Focussing on the chain part of (41) for now, Takahashi assumes that chain members are full *copies*. Therefore, after movement of some XP, no adjunction to either XP copy created by movement is permitted.

Another assumption made by Takahashi (1994) is that some version of the transderivational Shortest Paths condition (see (24) from chapter 1) holds; the constraint is informally stated by Takahashi as in (42).

(42)   *Shortest Move*:
        Make the shortest move.

---

[27] The exposition here follows Takahashi (1994); Stepanov (2007) adopts a version of Takahashi's approach, but only for the Subject Condition part of the CED (as noted above, the Adjunct Condition part is treated differently).

By assumption, every possible intermediate landing site (for a given movement type: A, A-bar, head) must be used in the course of movement, by adjunction to XP.[28]

Takahashi's observation then is that assuming both the UCA and Shortest Move will produce a dilemma in cases of extraction from a subject that has undergone movement from Specv to SpecT: It is impossible to satisfy both the UCA and Shortest Move simultaneously in this context. Either one of the two copies must be targetted by adjunction, in violation of the UCA (as in (43-a)), or a non-local movement step must be carried out, in violation of Shortest Move (skipping the DP adjunction site and moving to TP-Adj directly, as in (43-b)).

(43)   a.   *Who$_1$ has [$_{TP}$ [$_{DP_2}$ *who$_1$* [$_{DP}$ a comment about who$_1$]] T [$_{vP}$ [$_{DP_2}$ a comment about who$_1$] [$_{v'}$ v-annoyed [$_{VP}$ t$_V$ you]]]] ?
        b.   *Who$_1$ has [$_{TP}$ *who$_1$* [$_{DP_2}$ a comment about who$_1$] T [$_{vP}$ [$_{DP_2}$ a comment about who$_1$] [$_{v'}$ v-annoyed [$_{VP}$ t$_V$ you]]]] ?

The prediction then is that CED effects with subjects only show up if the subject is moved from its base position. External arguments in situ (in Specv) and derived subjects (as in passive clauses) should be transparent as long as they can stay in situ (which they can in various languages).[29]

As for the Adjunct Condition part of the CED, Takahashi assumes that clauses with adjuncts are coordination-like structures; hence, adjunction to adjuncts is blocked by the UCA in (41), and the Shortest Move condition rules out any instance of movement out of an adjunct. This part of the analysis raises various questions which I will not address here; see Stepanov (2007, 99-100) for critical discussion.

In my view, the hypothesis that an appropriate theory of freezing can derive all CED effects is empirically problematic. However, I will defer discussion of potential empirical counter-evidence to the end of subsection 3.5. For the moment, I will confine myself to pointing out that some of the assumptions that the analysis relies on

---

[28] In a phase-based approach that incorporates the PIC (see (9)), this requirement would be very similar to assuming that every phrase is a phase, with "adjunction to XP" replaced with "movement to the edge domain (a specifier) of X". I will come back to this issue in chapter 3.

[29] As noted by Takahashi (1994, 31-32), there are actually a few more loopholes that need to be closed in order to successfully account for CED effects as freezing effects derivable from the UCA and Shortest Move. Perhaps most importantly, a derivation needs to be excluded in which adjunction to DP applies when DP is still in situ (this is not prohibited by the UCA because, at this point, the subject DP is still a trivial chain); next, DP (with the *wh*-item adjoined to it) undergoes movement to SpecT; finally, *wh*-extraction takes place from the DP in SpecT. Such a derivation violates neither the UCA nor Shortest Move. Takahashi suggests that it is excluded by the Fewest Steps condition discussed in chapter 1, as an instance of illict chain interleaving (see Collins (1994)). Another derivation that needs to be excluded is one in which *wh*-movement to SpecC precedes subject raising from Specv to SpecT; such a derivation violates the Strict Cycle Condition.

are conceptually problematic (at least from the point of view adopted in this mono-
graph). First, in the guise of Shortest Move, the analysis involves a trans-derivational
constraint. This is incompatible with the meta-requirement formulated in (11-b) (and
adopted in much recent minimalist work) according to which constraints should not
be complex. As noted above, in some cases it is straightforwardly possible to transfer
an account in terms of the transderivational Shortest Paths condition (which, under at
least some reading of (42), may be viewed as a no more than a somewhat more explicit
version of Takahashi's Shortest Move) to an account in terms of the local derivational
(G)MLC (e.g., in the case of superiority effects).[30]

The problem is that it is by no means clear whether a version of the (G)MLC
could take over the role of Shortest Move in Takahashi's account of CED effects. In
the form that it takes in (1) at least (as well as in most recent minimalist work), this
is certainly not the case. The main difference is that the (G)MLC, in all its versions,
is a constraint minimizing the distance between the attracting head that provides the
landing site for movement and some item that can undergo the movement. In contrast,
Shortest Move does not care about the (head providing the) landing site as such; it is
a minimality requirement imposed on the moved item. This reflects the fundamental
difference between "attract"-type theories of minimality and "greed"-type theories of
minimality; and only the latter is of use in Takahashi's approach to CED effects. A
similar problem arises with the (apparaent) necessity to invoke the transderivational
constraint Fewest Steps in order to rule out derivations of ungrammatical exmaples
involving freezing that would respect both the UCA and Shortest Move (see footnote
29).

A second type of conceptual problem concerns the Chain Uniformity constraint. It
is doubtful whether this constraint could be motivated either as a constraint imposed by
requirements of the interfaces or as an economy condition (as would be demanded by
(11-c)).[31] Furthermore, this requirement makes it necessary to scan large domains of
syntactic structure; at least in the form that it takes in (40), it does not seem compatible
with a phase-based approach, where the active part of the derivation is very small at

---

[30] For the sake of the argument, I abstract away for now from the result of the previous section – viz.,
that the (G)MLC should itself be abandoned since it fails the meta-requirements (11-d) (no massive search
space) and (11-e) (no redundancies).

[31] This would seem to be obvious for the economy part. As for the potential status of Chain Uniformity as
an interface constraint, is is unclear why the semantic interface should worry about what basically amounts
to an identity requirement of chain members when principles of semantic interpretation usually require
chain-members to be *non-identical* (see, e.g., Heim & Kratzer (1998)).

any given stage.[32] Finally, it is worth pointing out that the analysis crucially relies on the copy theory of movement. If one abandons the copy theory of movement (in favour of either trace theory, remerger/multidominance theory, or the hypothesis that movement is proper displacement that leaves no reflex in the original position; see chapter 3), the analysis cannot be maintained.

3.5.3 Phi-Completeness: Boeckx (2003)

Another freezing approach to CED effects is developed by Boeckx (2003). Boeckx (2003) (like Stepanov (2007)) postulates that the CED is not homogeneous, and that the Adjunct Condition and the Subject Condition are to be treated differently. As for the Adjunct Condition, Boeckx pursues an approach that relies on the concept of Φ-inertness: First, the empirical evidence tells us that probes cannot undergo Agree with anything inside an adjunct. Second, suppose that Move always involves Agree. It follows from these two assumptions that the barrier status of adjuncts is predicted.

As for the Subject Condition, Boeckx (2003) suggests that freezing is involved. In fact, the approach is designed to be a minimal variation of Takahashi's (1994) approach that captures his basic insight that "the ban on extraction out of displaced constituents results from what one might call a *chain conflict*" (see Boeckx (2003, 104)). Takahashi's combination of UCA and Shortest Move is replaced with the constraint in (44): a version of the *Freezing Principle* that I call *Constraint on Φ-complete Domains* (Boeckx does not give it a name).

(44)   *Constraint on Φ-complete Domains*:
         Agree cannot penetrate a domain that is already Φ-complete.

On this basis, the analysis works as follows. When an XP has undergone movement and reached its final landing site, it freezes – it is Φ-complete. Next, movement out of XP requires an Agree relation into XP. Therefore, moved (Φ-complete) XPs are barriers.

This approach to CED effects raises several questions. The first concerns the status of the constraint in (44). Given the Phase Impenetrability Condition (Chomsky (2000; 2001b; 2008)), (44) is a second constraint that imposes a locality requirement on syntactic operations. Neither constraint is reducible to the other one (not all phases are Φ-complete, not all Φ-complete items are phases, and phases provide a domain that is accessible from outside, which (44) must not do), but many redundancies arise. Furthermore, the Constraint on Φ-complete Domains does not seem to qualify as either a requirement imposed by the phonological or semantic interface, or an economy

---

[32] There might be a more local way of construing Chain Uniformity, though, where chains do not have to be scanned for their properties as a whole, but only highly local chain links are considered, in line with the PIC.

constraint; it is therefore excluded by the meta-requirement on constraints in (11-c).

A second problem concerns the derivation of the Adjunct Condition. It does not suffice to state the empirical observation that there is no Agree into adjuncts (as it is done by Boeckx (2003)); this must eventually be made to follow from some theoretical assumption(s). The problem now is that the stipulation that seems to be needed would be something like (45), which parallels (44).

(45)   *Constraint on Adjuncts*:
        Agree cannot penetrate an adjunct.

(45) is subject to the same criticism as (44), given the meta-requirement in (11-c). Both constraints look a lot like CED-type constraints, with Move replaced with Agree. However, by reducing the Adjunct Condition for Move to an Adjunct Condition for Agree, and (a more liberal version of) the Subject Condition for Move to a Constraint on Φ-complete Domains for Agree, the problem of deriving these constraints is not actually solved; it is merely transferred to some other domain.[33]

Third, there are questions concerning the scope of the account. Given the Constraint on Φ-complete Domains and the assumption that Move presupposes Agree, it must be ensured that extraction from (what looks like a) Φ-complement object DP or CP in complement position does not violate the constraint; thus, there is a danger that the approach is too strong. On the other hand, it may also be that the approach is too weak. Irrespective of the issue of whether there are CED effects with subjects that are not confined to freezing (see below), it can be noted that the approach is supposed to cover *all* kinds of freezing effects, i.e., not only those that arise with subject raising to SpecT. However, there are freezing effects with categories where it does not seem to make sense to attribute them the property of being or not being Φ-complete; cf. the examples with illicit extraction from a topicalized VP in German in (46) and from a topicalized PP in English in (47) (also see (99) in chapter 1).

(46)   a.   Ich denke [$_{CP}$ [$_{VP}$ das Buch gelesen]$_2$ hat keiner t_2]
              I   think   the book read   has no-one

        b.   [$_{DP}$ Was]$_1$ denkst du [$_{CP}$ t'_1 hat keiner [$_{VP}$ t_1 gelesen]$_2$] ?
              what   think you   has no-one   read

        c.   *[$_{DP}$ Was]$_1$ denkst du [$_{CP}$ [$_{VP}$ t_1 gelesen]$_2$ hat keiner t_2] ?
              what   think you   gelesen   has no-one

(47)   a.   Who$_1$ do you think that he will talk [$_{PP_2}$ to t_1] ?
        b.   *Who$_1$ do you think that [$_{PP_2}$ to t_1] he will talk t_2 ?

---

[33] Incidentally, the strategy that can be observed here is reminiscent of the attempt in Chomsky (1986a) to unify locality domains for government and for movement by postulating a common concept of barrier.

Thus, I take it that the approach in Boeckx (2003) is not yet the last word on the issue of deriving CED effects.[34]

3.5.4 *Criterial Freezing: Rizzi (2006; 2007)*

The discussion in the present subsection of the approach in terms of Criterial Freezing developed by Rizzi (2006; 2007) functions more like an interlude than a proper continuation of the topic of deriving CED effects: The approach developed in Rizzi (2006; 2007) and related work is not primarily concerned with CED effects, but it is based on a principle that is very similar to a number of freezing constraints that yield CED effects as a consequence, and this is why I mention it here. Rizzi understands the constraint of Criterial Freezing as in (48); at least at first sight, there are strong similarities with Boeckx's (2003) constraint in (44).

(48)   *Criterial Freezing*:
         In a criterial configuration, the criterial goal is frozen in place.

However, Rizzi does not propose that (48) can derive CED effects. Quite on the contrary: To account for the contrast between (49-a) and (49-b) in French, Rizzi (2007) actually assumes that a criterially frozen subject is *not* a barrier for extraction. On Rizzi's view, the construction in (49-b) is legitimate *because* the subject DP is endowed with "nominal and Φ-features", and these are maintained in the criterial subject position if only *combien* is extracted (in contrast to (49-a), where Criterial Freezing is assumed to be violated).[35]

(49)   a. *$[_{DP_2}$ Combien$_1$ de personnes] veux-tu $[_{CP}$ que $[_{TP}$ t$_2$ viennent à
                how many of people   do you want   that   come   to
            ton   anniversaire]] ?
            your birthday

        b. ?Combien$_1$ veux-tu $[_{CP}$ que $[_{TP}$ $[_{DP_2}$ t$_1$ de personnes] viennent à
            how many do you want   that   of people   come   to
            ton   anniversaire]] ?
            your birthday

Having clarified that Criterial Freezing does not provide an account of CED effects, let me now return to approaches that do. Gallego & Uriagereka's (2006) theory of phase sliding is a case in point.

---

[34] Note in passing that there is an interesting potential tension between Boeckx's (2003) approach on the one hand, and Rackowski & Richards's (2005) on the other: In the latter approach, Agree with XP makes XP transparent for extraction out of it, in the former, Agree with XP (producing Φ-completeness) renders XP opaque.

[35] The slight deviance of (49-b) is attributed to a Left Branch Condition violation, and thus independent of the CED, on Rizzi's view.

3.5.5   Phase Sliding: Gallego & Uriagereka (2006)

The basic assumption here is that there is a specific freezing constraint; the constraint is similar to Boeckx's Constraint on Φ-complete Domains and Rizzi's Criterial Freezing; but it also incorporates Uriagereka's idea of 'flattening' complex (non-complement) constituents (and also, to some extent, Johnson's concept of renumeration). The constraint in question is called Edge Condition; see (50).

(50)   *Edge Condition*:
        Syntactic objects in phase edges become internally frozen.

(50) does not impose a ban on extraction from moved items per se; it only blocks extraction from items that have undergone movement to phase edges. (Of course, the two options may be identical, if all movement is movement to phase edge positions.)

As for the basic idea, (50) is not radically different from what can be found in other approaches. Gallego & Uriagereka's (2006) main new contribution is that they assume that v-to-T movement may result in TP (rather than vP) becoming the relevant phase; i.e., movement of v carries the phase property along. This accounts for a curious asymmetry with extraction from subjects in Spanish: Preverbal subjects are barriers, postverbal subjects are not; cf. (51-a) vs. (51-b).

(51)   a.   De qué conferenciantes$_1$ te   parece   que me
             of what speakers   to you seem-3.SG that to me
             van a impresionar   [$_{DP}$ las propuestas t$_1$]
             go-3.SG to impress   the proposals

        b.   *De qué conferenciantes$_1$ te   parece   que [$_{DP}$ las propuestas t$_1$]
             of what speakers   to you seem-3.SG that   the proposals
             me   van a impresionar
             to me go-3.SG to impress
             'Which speakers does it seem to you that the proposals by – will impress
             me?'

In (51-a), the subject DP has remained in its in situ position Specv; v-(+V)-to-T movement makes the subject DP postverbal, removes phase status from vP and imposes it on TP; Specv is thus not in a phase edge position anymore when *wh*-extraction from DP applies; and consequently, there is no CED effect, given the Edge Condition in (50): The subject DP is transparent. In contrast, in (51-b), the subject has undergone movement to SpecT; v-(+V)-to-T movement targets a lower position to the right, making the subject DP preverbal. Again, head movement turns TP into a phase; since the subject DP is now in the edge domain of a phase, it becomes internally frozen, and *wh*-extraction from DP becomes illegitimate. As in other derivational approaches to freezing, it must additionally be ensured that a derivation that employs a reverse order of operations – *wh*-extraction following head movement (so that Specv is not a

barrier anymore), but preceding subject raising to DP – is excluded (the Strict Cycle Condition being an obvious candidate).

Gallego & Uriagereka's (2006) proposal has a number of far-reaching empirical consequences, not all of which seem to be confirmed. Consider verb-second structures in German or Dutch as a case in point. Assuming (as is standard) that verb-second in these languages is derived via V-to-v-to-T-to-C movement, we should expect that all specifiers (of the main projection line, i.e., the clausal spine; see Sells (2001)) in a verb-second clause except for SpecC are transparent for extraction. This is certainly not the case, as is shown by an example such as (32-b) that was discussed above: A postverbal subject in verb-second clauses is still a barrier for extraction in German, as predicted under the standard CED. On the other hand, SpecV is a barrier for extraction in German even though it is not a phase edge. As shown in (52), dative DPs in German are islands for extraction in German (see Müller (1995) and Fanselow (2001a), among others).

(52) *[$_{PP_1}$ Über wen] hat der Verleger   [$_{DP}$ einem Buch   t$_1$] keine
        about whom has the publisher$_{nom}$ a   book$_{dat}$ no
     Chance   gegeben ?
     chance$_{acc}$ given

Assuming dative objects to occupy a specifier of V in German (with the accusative object merged as a complement of V), this state of affairs follows directly from the classical CED but must remain a mystery under Gallego & Uriagereka's reconstruction in terms of the Edge Condition.

A second problem with the approach is conceptual in nature. As noted before (in the context of the discussion of Uriagereka (1999)), phases are first and foremost motivated by complexity considerations in Chomsky (2001b; 2008) and related work; allowing phases to have flexible size is not compatible with this view. This fundamental incompatibility also arises, to different degrees, in Grohmann (2000), Marušič (2005), den Dikken (2007), and Gallego (2007). The problem is particularly worrisome in Gallego & Uriagereka's (2006) approach because it would seem that iterated instances of head movement can postpone the creation of a phase again and again – in the extreme case, if all heads of a clause are related to one another by head movement, there is only one phase left.[36]

---

[36] A side remark may be due on the role of head movement in Gallego & Uriagereka's (2006) approach in comparison with Kitahara's (1994) approach discussed earlier. One might think at first sight that Gallego & Uriagereka's (2006) approach predicts the opposite of Kitahara's (1994) approach discussed above – in one approach, head movement creates transparency of a specifier (Gallego & Uriagereka); in the other approach, head movement creates opacity of a specifier (Kitahara). However, this is not the case. In both approaches, head movement turns a specifier of the landing site into a barrier; and in both approaches, a specifier associated with the head in situ is predicted to be transparent for extraction.

3.5.6   Freezing Analyses: Conclusion

To sum up the discussion of freezing, I would like to contend that, independently of individual questions raised by the analyses, two general problems with freezing accounts of the CED can be identified that show up in all analyses. First, on the conceptual side, it turns out that all freezing analyses rely on additional, otherwise unmotivated constraints – Chain Uniformity (and the UCA) in Takahashi (1994), the Constraint on Φ-Complete Domains (and a separate Constraint on Adjuncts) in Boeckx (2003), and the Edge Condition in Gallego & Uriagereka (2006). It is far from clear whether these constraints can be taken to comply with basic minimalist tenets, as formulated above (see (11-c)). Second, on the empirical side, freezing analyses have nothing to say about CED effects that arise in contexts where the barrier has not undergone movement. Here is an example: Assuming that particles like *denn*, *wohl*, *ja* demarcate the vP edge in German, it seems clear that the subject DP (DP$_3$) is in situ in the German examples in (53). Nevertheless, a CED effect occurs with extraction out of the subject.[37]

(53) a. *Was$_1$ haben denn [$_{DP_3}$ t$_1$ für Bücher] [$_{DP_2}$ den Fritz]   beeindruckt ?
          what have PRT   for books$_{nom}$ the Fritz$_{acc}$ impressed

     b. *[$_{PP_1}$ Über wen] hat wohl [$_{DP_3}$ ein Buch t$_1$] [$_{DP_2}$ den Fritz]
          about whom has PRT   a   book$_{nom}$ the Fritz$_{acc}$
          beeindruckt ?
          impressed

---

[37] A few remarks are in order here. Subject raising to SpecT is optional in German; see, e.g., Grewendorf (1989). If the subject DP$_3$ in (53) shows up to the left of the particle, this arguably implies that it has undergone movement to SpecT. The resulting sentences are also ungrammatical; but this fact would be expected under a freezing approach. In this context, it might also be worth pointing out that the experimental study conducted by Jurka (2008; 2010) (on the basis of questionnaires handed out to non-linguist native speakers) suggests that, indeed, extraction from in-situ subjects in German is significantly degraded compared with extraction from objects; this can be interpreted to mean that there are CED effects with in situ subjects in Specv in German. (However, in my view, one should be very careful in attributing experimental studies of this type too much importance, for the simple reason that there are always very many confounding factors which cannot possibly be controled for, and which might influence the subjects' judgements; in my view, this problem is particularly severe (pace Jurka (2010, 25-26)) if the subjects are non-linguists rather than linguists because only the latter are trained in abstracting away from intervening factors.) Finally, some apparent counter-examples to the generalization that in-situ subjects are barriers in German have been brought forward in the literature (e.g., by Haider (1983; 1993a) and Diesing (1992)). I address these examples in chapter 4 (and argue that closer inspection reveals that they are not counter-examples at all, with some core cases instantiating the *melting* effect addressed in the next subsection, others involving unaccusative constructions, yet others non-movement phenomena, and so on).

3.6 *General Remarks*

More generally, from the discussion of existing attempts to derive the CED from more basic assumptions in a minimalist setting, the following conclusions emerge. First, analyses that are centered around the working of elementary operations like Move or Agree rely on special assumptions that mimic assumptions in Chomsky's (1986a) theory of barriers. Second, analyses that are based on specific concepts of cyclic spell-out are incompatible with basic assumptions about the timing of spell-out and the size of spell-out domains in standard phase theory, and ultimately with the notion of phase in general. Third, analyses that rely on freezing are all incompatible with the existence of CED effects where an XP is a barrier in its in situ position. Furthermore, it has turned out that all of the approaches discussed so far make it necessary to stipulate separate constraints and/or concepts that are not independently motivated, and that may not always fall under either economy or interface requirements.

Finally, and perhaps most interestingly, all these analyses have nothing to say about what I call *melting* effects, a class of data that I will discuss in some detail in chapter 4. In melting constructions, it looks as though an XP may qualify as a barrier in one case and as transparent in another even though it has exactly the same structural relationship with the surrounding heads in the two contexts. A melting effect with local scrambling of an object DP_2 in front of the subject DP_3 in German is illustrated in (54). (54-a) is ungrammatical, as expected under many theories (and under the classical CED). However, (54-b) is drastically improved and, in fact, far from ungrammatical, even though the subject DP_3 has not changed position (only the object DP_2 has – it has undergone local scrambling to a position in front of the subject). This is unexpected under the CED and under all the approaches discussed in this subsection that attempt to derive it.

(54) a. *Was_1 haben [$_{DP_3}$ t_1 für Bücher] [$_{DP_2}$ den Fritz] beeindruckt ?
           what have   for books$_{nom}$ the Fritz$_{acc}$ impressed

      b. Was_1 haben [$_{DP_2}$ den Fritz] [$_{DP_3}$ t_1 für Bücher] t_2 beeindruckt ?
          what have   the Fritz$_{acc}$ for books$_{nom}$ impressed

The phenomenon is not confined to German. As shown in (55), virtually the same melting effect shows up with local object scrambling in front of an in-situ subject in Czech.

(55) a. *[$_{DP_1}$ Holka] neuděřila [$_{DP_3}$ žádná t_1] Petra$_2$
           girl$_{nom}$ hit   no$_{nom}$ Petr$_{acc}$
          'No girl hit Petr.'

      b. [$_{DP_1}$ Holka] neuděřila Petra$_2$ [$_{DP_3}$ žádná t_1] t_2
          girl$_{nom}$ hit   Petr$_{acc}$ no$_{nom}$
          'No girl hit Petr.'

4.   Locality Constraints: State of the Art

4.1   Conclusion

Let me sum up the findings of this chapter. I have argued that neither the (G)MLC nor
the CED qualify as legitimate constraints in a strictly derivational, phase-based ap-
proach to syntax that meets minimalist demands. As far as the (G)MLC is concerned,
there seem to be few (if any) serious competitors in recent minimalist work. As for
the CED, I went through a number of proposals for deriving its effects, concluding
that none of the existing attemps can be regarded as fully successful (from the present
perspective of a phase-based approach that adopts the meta-requirements in (11) at
least). This leaves us with the situation that a principled, minimalist account of both
the (G)MLC and the CED is still outstanding.

     One option that one might pursue at this point is to assume that (at least certain)
restrictions on movement might in fact not reveal the influence of principles of gram-
mar, but might be due to parsing difficulties. On this view, grammar would freely
permit sentences with extraction from, say, subject DPs or adjunct CPs, and it would
freely permit cases where a lower *wh*-phrase is moved across a higher *wh*-phase. The
only problem with these sentences would be that they are more difficult to process for
the human brain. In a nutshell: Island effects would belong to the domain of *perfo-
mance*, not to the domain of *competence*. Such a view has been entertained by various
scholars over the last decades, with varying degrees of formal explicitness and vary-
ing degrees of experimental backing by behavioural and neurophysiological studies;
see, e.g., Hawkins (1999), Kluender (2004), and Culicover (2008) (and literature cited
there). Interestingly, it seems that within the framework of Head-Driven Phrase Struc-
ture Grammar (HPSG; see Pollard & Sag (1994)), performance-based accounts have
more and more come to replace competence-based accounts of island effects. Since
I take it to be instructive to consider this drastic step in theory construction, I now
address the issue in some more detail in an excursus.

4.2   Interlude: Islands in HPSG

4.2.1   Slash Features in GPSG

In Generalized Phrase Structure Grammar (GPSG; see Gazdar (1981; 1982), Mal-
ing & Zaenen (1982), and Gazdar, Klein, Pullum & Sag (1985)), the displacement
property of natural languages is not modelled by movement transformations. Rather,
displacement is accounted for by postulating SLASH features – category-valued fea-
tures that encode the properties of the displaced item. SLASH features are passed on
from daughter to mother (or vice versa, given that the approach is inherently non-
derivational), thereby connecting the original base-position of the displaced item with
the position in which it eventually shows up (its ultimate "landing site"). Any theory
of SLASH feature percolation consists of three parts: bottom, middle, and top. First,

the SLASH feature is introduced in (or close to) the base position (bottom); second, the SLASH feature is passed on in syntactic trees (middle); and third, SLASH feature percolation terminates once the filler (the displaced item) is reached (top); see (56).

(56) $\underbrace{[\text{What...}}_{\text{top}} \underbrace{[\text{do you think that Mary bought}}_{\text{middle}} \underbrace{[\text{t}]]]}_{\text{bottom}}$

For each part, a certain rule or principle is required. For instance, in Gazdar et al. (1985), the bottom of displacement constructions is handled by a SLASH Termination Metarule (see (57)) which simply states that for every context-free phrase structure rule that introduces a lexical item and an XP as its sister, the XP can be null – i.e., a trace.[38]

(57)   SLASH *Termination Metarule*:
         X → *W*, XP ⇒ X → *W*, XP[+NULL]

Many cases of overgeneration that would be induced by (57) can be avoided by assuming a *Feature Specification Default* (FSD) (viz., FSD no. 3) according to which [NULL] can be instantiated on a given category only if this is forced by some rule (like (57)). Crucially, Gazdar et al. (1985) then also postulate a *Feature Co-occurrence Restriction* (FCR) (FCR no. 19, to be precise) according to which the presence of [+NULL] implies the simultaneous presence of [SLASH] on a category, and this is the starting point of SLASH feature percolation.

Turning next to Gazdar et al.'s (1985) treatment of the middle of displacement constructions, it suffices to assume that [SLASH] is (both a head and) a foot feature, which implies that it is shared between daughter and mother not only along the projection line of the head, but also between a non-head daughter and its mother. This is ensured by an indendently motivated constraint, the so-called *Foot Feature Principle*.

Finally, as concerns the top part of displacement constructions, Gazdar et al. (1985) simply assume that there are phrase structure rules (more precisely, immediate dominance rules; see the last footnote) that expand a clausal category (like S) by introducing the displaced item and its sister: the head of the clausal category that bears the SLASH feature corresponding to the displaced item; see (58) (where S could just as well be conceived of as standing for CP, and H, for C′).

---

[38] A few remarks. First, Gazdar et al. (1985) assume that the structure-building rules of grammar do not yet encode linearization; they are pure immediate dominance rules (hence the comma notation). Second, as it stands, (57) only permits traces as sisters of lexical heads. This implies that specifiers either cannot move, or move by virtue of some other rule; arguing for the second possibility, Gazdar et al. (1985) take this to be a welcome result. – Note (again) that the theory-internal problem generated by (what looks like) specifier movement that shows up here parallels the problem that Uriagereka (1999) encounters with this construction; see the discussion of (29) above.

(58)   S → XP, H/XP

4.2.2 Displacement in HPSG

Work on displacement in Head-Driven Phrase Structure Grammar (HPSG) heavily relies on the SLASH theory of movement developed in GPSG (see Pollard & Sag (1994), Sag & Wasow (1999)). There are a few changes, though. First, the feature SLASH does not simply take a category as its value anymore, but rather a list of categories. This makes it possible to assume that there can be more than one locally unbound trace in a given category – an option that would in any event be required in a minimalist (more generally, Principles and Parameters) approach that permits a combination of, e.g., (i) A-movement of a subject from Specv to SpecT; (ii) A-bar movement of a *wh*-object from CompV to SpecC; and (iii) v(+V)-to-T movement. In such a case, the SLASH feature on, say, vP would have to encode three missing constituents. Less theory-dependent arguments for abandoning the one-hole property of displacement as it follows under Gazdar et al.'s (1985) conception of SLASH features are already provided in Maling & Zaenen (1982) on the basis of English movement constructions instantiating double extraction (also see Pesetsky (1982) for extensive discussion), and Scandinavian movement constructions like those instantiating triple extraction in Swedish; cf. (59-ab). To account for the data, Maling & Zaenen (1982) essentially anticipate Pollard & Sag's (1994) move to lists of categories as feature values for SLASH.

(59)   a.   Problems this involved$_1$ my friends on the East Coast$_2$ are difficult to talk
             to t$_2$ about t$_1$

        b.   [Sadana här känsliga politiska fragor]$_1$ har   ja flera   studenter som$_2$
               such touchy   political questions have I   many students that
             det   inte finns   nagon som$_3$ jag tror   [t$_2$ skulle vaga prata med
             there not is-found anyone that I   believe   should dare talk with
             t$_3$ om   t$_1$]
               about

The second major change from the treatment of displacement in GPSG to the treatment of displacement in HPSG concerns the role of traces. In HPSG, it is sometimes argued that traces can (and should) be dispensed with: The information about what is missing does not have to be encoded by the trace because it is already present on the lexical item that the trace is an argument of. However, it is not the case that all HPSG-based approaches to movement do without traces; Sag & Wasow (1999) abandon the concept of a trace, but Pollard & Sag (1994), Müller, St. (2007), and Levine & Sag (2003) do not.

Abstracting away from the value of SLASH features and from the possibility to dispense with traces, the displacement theories of HPSG and GPSG are very similar. Taking, e.g., the (trace-less) version of HPSG in Sag & Wasow (1999) as a point of

departure, the bottom, middle, and top are handled as follows.

Bottom   Simplifying a bit, lexical items are associated with hierarchically ordered lists of subcategorization features (that have to be discharged from the head in syntactic trees by merging with arguments – I will come back to this in chapter 3), and with an argument structure list that corresponds to it. There is a feature SLASH (or GAP) that, if present on a lexical item, signals that something that shows up on the argument structure list does not have a correspondent in the subcategorization feature list: Something that is not on the subcategorization feature list even though it is part of the argument structure list will have to be in the SLASH feature list. Technically, this is accomplished by postulating a subtraction operation for lists: If A and B are lists, then $A \ominus B$ is a list that results from removing the items of B from A. A SLASH feature is optionally introduced on lexical items for one of its arguments: If SLASH shows up, a modification of the Argument Realization Principle that accounts for standard linking effects ensures that each argument that might show up on the subcategorization list of a lexical item may alternatively show up on the SLASH list. This derives the effects of the SLASH *Termination Metarule*.[39] Here is an example for optional localization of some subcategorization feature value (which would normally belong in COMPS) in the SLASH feature list.[40]

---

[39] Of course, the prediction then is that only subcategorized items can undergo movement. Assuming that languages may decide whether subjects are subcategorized or not, differences with respect to the mobility of subjects across languages can be accounted for. As for adjunct movement, extra assumptions are required. For instance, it has been suggested that adjuncts may also enter a (generalized version of) subcategorization lists. Thus, Bouma, Malouf & Sag (2001) propose that, in addition to the standard argument structure list, there is a more comprehensive DEPS list that makes adjuncts subcategorizable that do not belong to the argument structure of a lexical item. On the basis of an entry in the DEPS list, a subcategorization feature can be generated for an adjunct, which can then either be locally discharged (via the subcategorization list), or transferred (via the SLASH list); in the latter case, adjunct movement takes place.

[40] The notation is taken directly from Sag & Wasow (1999), who differentiate the subcategorization list further into a specifier list and a complements list. In addition, they notate the SLASH feature as GAP; I have taken the liberty to replace their GAP feature with a SLASH feature here. All the other abbreviations (e.g., SEM) should be self-explanatory.

(60)   *The lexical item 'loves' with a* SLASH *feature*:

$$
\left[
\begin{array}{ll}
\text{SYN} &
\left[
\begin{array}{ll}
\text{HEAD} & \left[\, verb\,\right] \\
\text{SPR} & \left\langle \boxed{3} \right\rangle \\
\text{COMPS} & \left\langle \ \right\rangle \\
\text{SLASH} & \left\langle \boxed{4} \right\rangle
\end{array}
\right] \\[6em]
\text{ARG-ST} &
\left\langle \boxed{3}
\begin{array}{l}
\text{NP}_i \\
\left[\text{AGR} \quad 3sing\right]
\end{array}
,\ \boxed{4}\text{NP}_j \right\rangle \\[4em]
\text{SEM} &
\left[
\begin{array}{ll}
\text{MODE} & \text{prop} \\
\text{INDEX} & \text{s} \\
\text{RESTR} & \left\langle
\left[
\begin{array}{ll}
\text{RELN} & \text{love} \\
\text{LOVER} & \text{i} \\
\text{LOVED} & \text{j}
\end{array}
\right]
\right\rangle
\end{array}
\right]
\end{array}
\right]
$$

loves
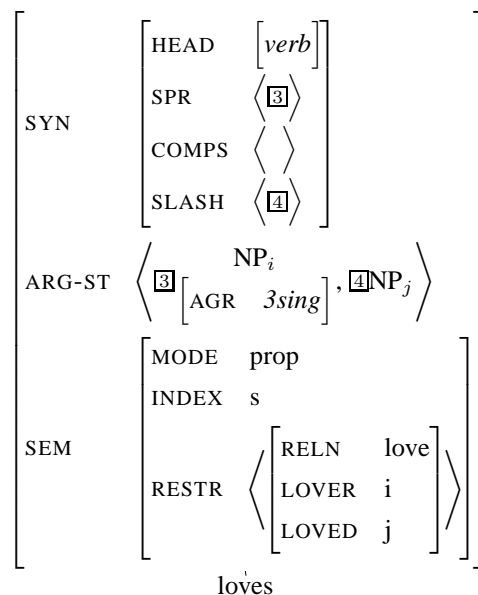
Middle   Turning next to the middle, the role of the Foot Feature Principle in GPSG
is taken over by the *Gap Principle* in Sag & Wasow (1999), see (61).

(61)   *Gap Principle*:
        A well-formed phrase structure licensed by a head rule (except the head-filler
        rule; see below) must satisfy the following structural description:

$$\left[\text{SLASH} \quad \boxed{1} \oplus ... \oplus \boxed{n}\right]$$

$$\left[\text{SLASH} \quad \boxed{1}\right] \qquad ... \qquad \left[\text{SLASH} \quad \boxed{n}\right]$$
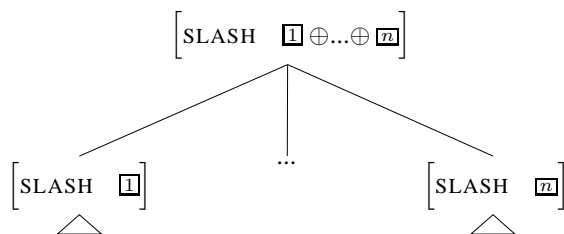
Here is an example of how the Gap Principle works.

(62)

$$
\begin{array}{c}
\text{S} \\
\left[\text{SLASH} \;\; \langle\text{NP}\rangle\right]
\end{array}
$$

- NP — we
- VP $\left[\text{SLASH} \;\; \langle\text{NP}\rangle\right]$
  - V — know
  - S $\left[\text{SLASH} \;\; \langle\text{NP}\rangle\right]$
    - NP — Dana
    - VP $\left[\text{SLASH} \;\; \langle\text{NP}\rangle\right]$
      - V $\left[\text{SLASH} \;\; \langle\text{NP}\rangle\right]$ — loves
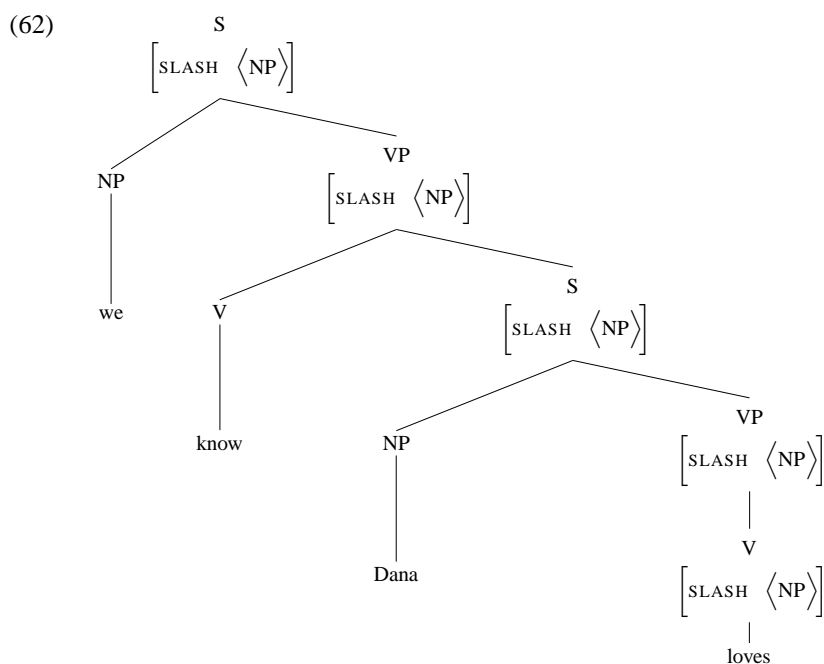
Top   As in GPSG, there is a rule that expands a clausal category to a filler and a slashed head daughter. Sag & Wasow (1999) assume the Head Filler Rule in (63).

(63)   *Head Filler Rule*:

$$
\begin{bmatrix} phrase \\ \text{SLASH} \;\; \langle\; \rangle \end{bmatrix} \rightarrow \boxed{1} \begin{bmatrix} phrase \\ \text{SLASH} \; \langle\; \rangle \end{bmatrix} \;\; \text{H} \begin{bmatrix} phrase \\ \text{FORM} \quad \text{fin} \\ \text{SPR} \qquad \langle\; \rangle \\ \text{SLASH} \quad \langle\boxed{1}\rangle \end{bmatrix}
$$

The Head Filler Rule has to be exempted from the Gap Principle by stipulation (the rule introduces a head and would violate the Gap Principle if it held for it). An example illustrating the working of the Head Filler Rule is given in (64).

(64)
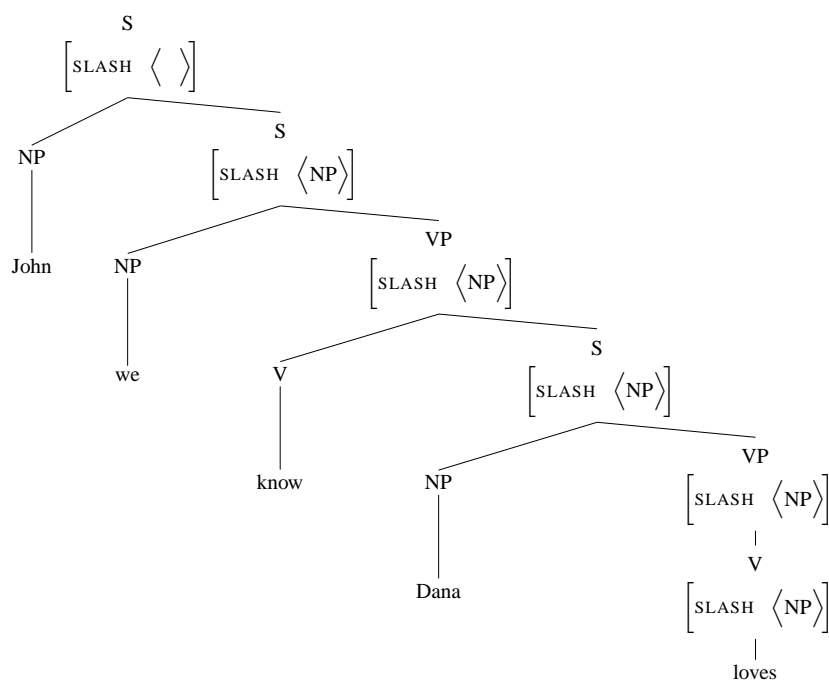
Thus, it seems clear that, technicalities aside, the approach to displacement construc-
tions in HPSG works more or less in the same way as it does in GPSG. Work in GPSG
has also addressed in some detail the role of islands for syntactic displacement (see in
particular Gazdar (1981) and Maling & Zaenen (1982)). However, it seems that the
constraints that are the main focus of this monograph have not extensively been cov-
ered (e.g., GPSG does not seem to have generated a principled approach to CED-type
islands). The case is different with HPSG, in which the Subject Condition has been
tackled.

4.2.3 Islands in HPSG

A simple way to state island constraints in GPSG/HPSG is to formulate them as con-
straints on instantiating SLASH features. Thus, in Pollard & Sag (1994), the Subject
Condition takes essentially the following form.[41]

---

[41] This is a slight simplification. In the original version, the constraint is formulated in such a way that the
the first element of a subcategorization list of a lexical item cannot bear a SLASH feature *unless another
element of the same subcategorization list also bears a* SLASH *feature*. This complication is brought about
by the desire to treat parasitic gaps (which may show up in subject islands) in the same way as regular gaps.
I ignore this complication here because I will remain silent on the treatment of parasitic gaps throughout
this monograph (but, again, see Assmann (2010) for an extension of the approach to CED effects developed
in chapter 4 below to parasitic gaps.

(65)   *Subject Condition* (HPSG):
        The first element of a subcategorization list of a lexical item cannot bear a
        SLASH feature.

The first argument of a subcategorization list is the subject, i.e., the argument of a
predicate that is discharged last. Given (65), a SLASH feature cannot be instantiated
on this kind of argument. Consequently, the propagation of a SLASH feature above
and below a subject argument will be interrupted, and (a constraint like) the GAP
PRINCIPLE will invariably be violated in cases of extraction from a subject.

  Interestingly, Pollard & Sag (1994) also implement Kuno's (1973) Clause Non-
final Incomplete Constitutent Constraint (see (83) of chapter 1). Their version is given
in (66) (in a slightly simplified form).

(66)   *Incomplete Constituent Constraint* (HPSG version):
        $$\begin{bmatrix} \text{INHER} \mid \text{SLASH} & \text{empty-set} \end{bmatrix} < \begin{bmatrix} \text{INHER} \mid \text{SLASH} & \text{nonempty-set} \end{bmatrix}$$

Like Kuno's original constraint, this accounts for the contrast in (67) in English.

(67)   a.   [$_{DP_1}$ Which man] did you buy [$_{DP}$ a picture of t_1] ?
        b. ?*[$_{DP_1}$ Which man] did John give [$_{DP}$ a picture of t_1] to Bill ?

4.2.4 Recent Developments: Competence vs. Performance

More recently, much work within HPSG has moved from a competence-based ap-
proach to locality constraints on displacement to a performance-based approach,
grounded in the hypothesis that restrictions on the human sentence processor can de-
rive island constraints. Levine & Sag (2003, 6) are quite explicit about this radical
change of perspective:

> *From Chomsky 1964 on, considerable effort and ingenuity has been devoted to*
> *deriving these [island] effects from a small set of increasingly abstract syntactic*
> *participles. Nonetheless, there is now increasing evidence that a good number*
> *of these phenomena are due not to constraints of grammar, but rather to pro-*
> *cessing, pragmatic, and discourse effects that can be manipulated in such a*
> *way as to ameliorate the severe unacceptability exhibited by [examples involv-*
> *ing violations of island constraints]. [...] The full range of island phenomena*
> *and the relative weighting of given effects will ultimately be explained by some*
> *combination of grammar-internal and extragrammatical factors whose precise*
> *delineation is yet to be determined.*

  As a case study showing how this programme is pursued in recent HPSG-based
work, consider the approach to CNPC effects in Sag, Hofmeister & Snider (2008),
which is relevant to the discussion in this chapter given that at least some CNPC
effects (viz., those involving relative clauses), perhaps even all CNPC effects (if ar-
gument clauses of N are not merged in complement position, see subsection 3.11 of

chapter 1) can be derived from the CED.[42] The general line of reasoning is as follows. First, if a sentence is not (or not fully) acceptable, it is a priori unclear whether this is due to a competence or a performance factor. Next, it is known that certain factors determine the processing of sentences. Third, on the basis of a self-paced reading experiment, such factors can be shown to be relevant for the processing of sentences with CNPC violations. An acceptability study that accompanies the self-paced reading experiments confirms these findings (at least as a tendency). There is thus good evidence that processing and grammatical theory are correlated: Either differences in grammar are responsible for differences in processing, or differences in processing are responsible for differences in grammar. Since there is independent evidence for the processing factors, or so the argument goes, the latter account is to be preferred.

More specifically, Sag, Hofmeister & Snider (2008) identify the following factors as determining processing difficulty in displacement constructions: distance (should be as small as possible), semantic complexity (should be minimized), informativity (should be maximized), frequency (should be high), similarity (of items should be avoided), finiteness (gives results that are worse than non-finiteness), contextualization (should be easy), and collocational frequency (of items simplifies processing). Thus, in the CNPC construction in (68-a), many such factors that disfavour processing are active; in contrast, in (68-b), several factors are involved that facilitate the processing of displacement (the judgements indicated here are Sag, Hofmeister & Snider's (2008)).

(68)   a.   *This was a puzzle that we met the man who solved t
        b.   This was the only crossword puzzle that we've ever found anyone who
             could solve t

The self-paced reading experiment carried out by Sag, Hofmeister & Snider (2008) then focusses on just two parameters: The *wh*-phrase can be primitive or complex (*who* vs. *which N′*), and the complex noun phrase type can be definite singular, indefinite plural, or indefinite singular (*the* vs. *Ø* vs. *a*). There were seven conditions: six that result from a cross-classification of the two parameters, plus one control condition. For instance, *I saw who Emma doubted the report that we had captured t in the nationwide FBI manhunt* instantiates the first condition (primitive, definite singular); *I saw who Emma doubted reports that we had captured t in the nationwide FBI manhunt* the second (primitive, indefinite plural); and *I saw which convict Emma doubted that we had captured in the nationwide FBI manhunt* the seventh (control). The results of the experiment were as follows. Complex fillers/displaced items (of the type *which N′*) are easier to parse than primitive fillers (of the type *who*); and different

---

[42] Also compare the approach to superiority effects in Sag, Hofmeister, Arnon, Snider & Jaeger (2008).

complex noun phrase types do not behave differently in interesting ways. The acceptability study that was designed as a control experiment involved gradient acceptability judgements by subjects of the same sentences. The results are the following. First, the control sentences (without CNPC configurations) are judged best. Second, sentences with a complex filler are judged as better than sentences with a primitive filler, but they still emerge as significantly degraded compared to the control condition; this latter fact implies that the result is different to what emerged from the self-paced reading experiment. Third, sentences with the indefinite, plural condition are judged as better than sentences with the indefinite, singular condition or with the definite, singular condition; this is a second difference to the first experiment.

Thus, there are massive differences between the results of the two experiments. Nonetheless, Sag, Hofmeister & Snider (2008) conclude that the two experiments have essentially the same outcome; that this cannot be accidental; that, therefore, either reduced grammaticality must be responsible for processing difficulties, or vice versa; and, finally, that since processing difficulties are independently motivated, they should be taken to be responsible for reduced grammaticality. The more general conclusion then is that the CNPC (or a set of more basic assumptions of grammatical theory from which the CNPC follows) can be dispensed with in the theory of grammar.

I take issue with these conclusions. As far as I can see, the experimental studies do not support the far-reaching claim that Sag, Hofmeister & Snider (2008) make. The acceptability study shows that complex noun phrases are islands. However, neither the acceptability study nor the self-paced reading experiment contributes anything to answering the question of why that should be the case. Indeed, the latter experiment seems to suggest that different complex noun phrase types do not play any role whatsoever. As a matter of fact, the only thing that the self-paced reading experiment shows is that *who*-phrases behave differently from *which*-phrases in cases of extraction from islands. This has been known for quite a while, and there are various analyses around in grammatical theory that can derive the asymmetry; see, e.g., Pesetsky's (1987), Cinque's (1990), and Rizzi's (1990) accounts in terms of D-linking or Hornstein & Weinberg's (1990) approach that directly relies on the structural complexity of moved items. All the evidence is therefore compatible with rules of grammar giving rise to CNPC islands (and predicting a variable behaviour for *which*-phrases and *who*-phrases), and processing and acceptability judgments reflecting this. In contrast, processing difficulties as such do not tell us anything about the nature of the island constraint.

I take this result to be typical of processing accounts of island phenomena.[43] For this reason (and this leads us back to the original question that motivated the present excursus), I do not think that it is a promising strategy to look for parsing difficulties when one is confronted with the task of deriving (G)MLC and CED effects from more basic principles; a grammar-internal account should be sought.[44]

4.3   Outlook

In the following two chapters (3 & 4), I will argue that both (G)MLC and CED effects follow from the *Phase Impenetrability Condition* (PIC), in interaction with independently motivated assumptions about movement and structure-building in general – in particular, about the conditions under which edge features (that drive intermediate movement steps) can be inserted on phase heads in derivations. In chapter 5, I address a residual island effect that follows from Relativized Minimality but not from the (G)MLC and the CED (or from what I argue derives (G)MLC and CED effects in chapters 3, 4), viz., what one may call an operator island effect (including *wh*-island and topic island effects; cf. subsection 3.6 of chapter 1). I argue that this residual island effect can also be derived without invoking a separate constraint. More specifically, operator island effects are shown to be derivable by invoking the independently

---

[43] In a similar vein, Kluender & Kutas (1993) argue that Subjacency effects should be reanalyzed as a mere processing phenomenon; their proposal relies on two behavioural studies and one EEG study measuring event-related brain potentials (ERPs). However, as far as I can see, the scope of their investigation is actually much more limited: Focussing on the (somewhat more informative) ERP study only, what Kluender and Kutas show is that (i) displacement in general involves a processing cost (detectable in the form of a left-anterior negativity, LAN); (ii) a *wh*-word *what* or *who* at the left edge of an embedded clause gives rise to a stronger N400 effect than *if*, which in turn creates a more negative ERP than *that* at the left edge of an embedded clause; and (iii) that the two effects may come together in the case of long-distance *wh*-movement from an embedded clause, ultimately inducing the strongest combined effect in *wh*-island contexts. Notwithstanding some potential problems in experimental design (concerning, e.g., the source of the different behaviour of *what/who* vs. *if* vs. *that*, with phonological differences a possible confounding factor, and differences with respect to displacement *within* the embedded clause another one), the only thing that Kluender and Kutas show is that there is a certain processing cost with extractions from *wh*-islands. This result is not extendable to other contexts where the Subjacency Condition has been invoked. Furthermore, it remains completely unclear why this processing cost should suffice to push the *wh*-island construction over the edge, resulting in ungrammaticality. Finally, Kluender & Kutas (1993, 578) explicitly assimilate the deviance of extractions from *wh*-islands to the deviance of cases of recursive center-embedding, which are well known to pose parsing problems; but they ignore that there is an important difference between the variable, and gradient, nature of the latter (which can in addition actively be influenced by factors like enhanced concentration, prosodic means, etc.), and the strict, and categorical, nature of the former.

[44] Yet another strategy that has been pursued is to derive (certain) island constraints from semantic or pragmatic requirements. See, e.g., Szabolcsi & Zwarts (1993) and Szabolcsi & den Dikken (2003) on certain kinds of weak (operator-induced) islands, Beck (1997) and Kim (2002) on intervention effects induced by quantifiers or focussed items, and Truswell (2007) on exceptions to the Adjunct Condition that seem to

motivated concept of feature *maraudage*, given that successive-cyclic movement proceeds in just the way argued for in chapters 3 and 4.

Anticipating the results of chapters 3 and 4, the main claims will be these: (i) Edge features that trigger intermediate movement steps to phase edges can only be inserted when they have an "effect on outcome", in Chomsky's (2001b) terms. A simple way of making precise what this means implies that (G)MLC effects follow from the PIC (chapter 3). (ii) Edge features that trigger intermediate movement steps to phase edges can only be inserted "after the phase is otherwise complete", in Chomsky's (2001b) terms. There is good theory-internal evidence for replacing "after" with "before"; this move implies that CED effects follow from the PIC (chapter 4).

---

be semantically conditioned (more precisely, accountable for in terms of event-structural notions). I have nothing to say about these approaches in what follows, except for noting that it seems highly unlikely that they can be generalized so as to cover all the core cases subsumed under the (G)MLC and the CED. Of course, this does not rule out the possibility that they account for some of the data.

Chapter 3

On Deriving (G)MLC Effects from the PIC

1. Introduction

Given the arguments against the (G)MLC as a primitive of grammar laid out in chapter 2, effects that are standardly attributed to this constraint need to be accounted for differently. In this chapter, I argue that they follow from the Phase Impenetrability Condition (PIC), given some independently motivated assumptions about the nature of successive-cyclic movement in a phase-based approach to syntax.[1] I proceed as follows. Section 2 introduces and justifies the necessary background assumptions; in particular, those that concern (i) the hypothesis that all syntactic operations are driven by features on lexical items (subsection 2.2); (ii) the hypothesis that phases are smaller than is standardly assumed (subsection 2.3); and, most importantly, (iii) the nature of the mechanism of edge feature insertion on phase heads, which is argued to rely on a concept of *phase balance*, as it was first suggested in Heck & Müller (2000; 2003) (subsection 2.4). In section 3, then, I show that the system automatically derives *wh*-intervention effects that involve c-command (i.e., superiority effects), and also accounts for the variable occurrence of these effects in languages like German, which have scrambling. In section 4 it is shown, based on Heck & Müller (2000; 2003), that there are *wh*-intervention effects that do not involve c-command (or dominance), and that therefore must remain a mystery under a standard (G)MLC-based approach; in contrast, the approach adopted here is shown to directly account for the effect. Section 5 presents some refinements. Section 6 investigates the scope of the general result reported in this chapter (i.e., it answers the question to what extent the (G)MLC can be viewed as derived, tackling F-over-F configurations in the course of doing so). Finally, section 7 draws a conclusion.

---

[1] Sections 1–5 of the present chapter are essentially based on Müller (2004a). However, various extensions have been made, and various changes have been carried out, among them one that is quite drastic since it effects the overall architecture of grammar (centered around the core issue of whether all movement is feature-driven or not).

2. Assumptions

2.1 Basic Assumptions

I presuppose that syntax is strictly derivational, generating sentences by bottom-up application of elementary operations like Merge and Agree, as in in Chomsky (2000; 2001b) and subsequent related work. Two fundamental principles ensuring that the part of a derivation that is accessible at any given stage is small are the Strict Cycle Condition and the Phase Impenetrability Condition, both of which have already figured in the preceding chapters. The Strict Cycle Condition can be defined as in (1), based on Chomsky (1973) and Perlmutter & Soames (1979).[2]

(1)   *Strict Cycle Condition* (SCC):
       Within the current XP α, a syntactic operation may not target a position that is
       included within another XP β that is dominated by α.

The version of the Phase Impenetrability Condition (PIC) that I will adopt throughout this study is given in (2) (see Chomsky (2000, 108), Chomsky (2001b, 13), and Chomsky (2008); also recall (17) and (9) of chapters 1 and 2, respectively).

(2)   *Phase Impenetrability Condition* (PIC):
       The domain of a head X of a phase XP is not accessible to operations outside
       XP; only X and its edge are accessible to such operations.

As for the notion of edge, I will assume (deviating minimally from the definition in (18) of chapter 1 for reasons that will become clear below, in subsection 3.1.2) that it is defined as in (3), as the *left-peripheral* minimal residue of specifiers (and adjuncts) outside of X′.

(3)   *Edge*:
       The edge of a head X is the left-peripheral residue outside of X′.

As remarked in chapter 2, the SCC restricts possible positions for the operation-inducing head (e.g., the probe of an Agree operation); and the PIC restricts possible positions in which the derivation can find an item that is affected by an operation (e.g., the goal of an Agree operation). As also noted in chapter 2, the SCC and the PIC impose strong restrictions on *active parts* of derivations. I take both the SCC and the PIC to be first and foremost motivated as constraints that reduce computational complexity by minimizing search space in derivations. The reduction of search space throughout the derivation under (1) and (2) can schematically be illustrated as in (4), where the

---

[2] For somewhat more recent versions of the SCC, see Chomsky (1995; 2001b), Collins (1997), Kitahara (1997), Bošković & Lasnik (1999), and Freidin (1999), among others. For the option of dispensing with the SCC under a certain concept of cyclic spell-out of phase domains, see footnote 5 of chapter 2.

underlined categories are phases and ⌢ and ⌣ indicate the parts of the derivation that are accessible at a given step ((4) = (10) of chapter 2).[3]

(4) *Search space under PIC*:

a. $\overbrace{[_{YP}...Y\ [_{XP}...[_{X'}X}^{\text{SCC}}\ \underbrace{}_{\text{PIC}}\ [_{WP}...W\ [_{UP}...U...\ ]]]]]$

b. $\overbrace{[_{ZP}...Z}^{\text{SCC}}\ [_{YP}...Y\ \underbrace{[_{XP}...[_{X'}X}_{\text{PIC}}\ [_{WP}...W\ [_{UP}...U...\ ]]]]]$

Adopting (versions of) the SCC and the PIC can be taken to be fairly uncontroversial against the background of recent work in derivational minimalism. Things are slightly different with three further assumptions that will be relevant, which I turn to now. The first important assumption is that all syntactic operation are driven by features.

2.2 *All Syntactic Operations are Feature-Driven*

It is often assumed that the operations Agree and internal Merge (i.e., movement) are feature-driven. In addition, I presuppose that external Merge (i.e., Merge applying to items from the numeration, or from the workspace of the derivation, which comprises the numeration and material in trees that have been created earlier) also does

---

[3] The version of the PIC that is assumed here and in what follows is one of the two versions envisaged in Chomsky (2001b) (the stricter one, at least at first sight; but see Richards (2004; 2011a) for extensive discussion). To accomodate certain cases of non-local Agree relations, Chomsky (2001b, 14) also introduces another version of the PIC, which is given in (i).

(i) *Phase Impenetrability Condition* (PIC; alternative version):
      The domain of a head X of a phase XP is not accessible to operations at ZP (the next phase); only X and its edge are accessible to such operations.

As a consequence of (i), the search space is enlarged: Operations in YP can now look for a goal in YP, in XP, in WP, or in the residue or head of UP.

(ii) *Search space under alternative PIC*:

a. $\overbrace{[_{YP}...Y\ [_{XP}...[_{X'}X}^{\text{SCC}}\ [_{WP}...W\ [_{UP}...U}\ ...\ ]]]]]$ $\underbrace{}_{\text{PIC}}$

b. $\overbrace{[_{ZP}...Z}^{\text{SCC}}\ [_{YP}...Y\ [_{XP}...[_{X'}X}\ [_{WP}...W\ [_{UP}...U...\ ]]]]]$ $\underbrace{}_{\text{PIC}}$

Given that the PIC's motivation is primarily tied to its reduction of search space, (i) is inferior to (2), and I will disregard it in what follows. As for the problem created by non-local Agree relations, see the final section of the present chapter, and chapter 4.

not come for free but must be triggered by appropriate features, viz., subcategoriza-
tion features (see Svenonius (1994), Collins (2003), Adger (2003), Lechner (2004),
Kobele (2006), Sternefeld (2006), Pesetsky & Torrego (2006), and Di Sciullo & Isac
(2008), among others). Thus, on this view, all basic syntactic operations are (non-
trivially) feature-driven.[4] More specifically, following Heck & Müller (2007), I will
assume that there are two types of features that drive operations: Structure-building
features (edge features and subcategorization features) trigger (internal and external)
Merge operations; and probe features trigger Agree operations. Here and in what
follows, structure-building features (for both basic Merge and movement) are accom-
panied by bullets ([•F•]), and probe features are embraced by asterisks ([∗F∗]).[5] I
presuppose that just like subcategorization features, edge features that trigger move-
ment do not have to be neutral with respect to the properties of the items they attract.
Thus, [•wh•] is a feature on C that forces movement of a *wh*-phrase (i.e., a phrase with
the feature [wh]) to C's domain; [•top•] induces topic movement (i.e., movement of
an item marked [top]); and so forth for other features that are involved in movement.
I will also assume that scrambling can be viewed as a purely formal operation (with
manifold information-structural effects) that is triggered by an edge feature [•Σ•] on
certain designated heads which matches a feature [Σ] on the scrambled item ([Σ] can
be freely instantiated on all categories that may undergo scrambling in a language);
see Müller (1998), Grewendorf & Sabel (1999) and Sauerland (1999), among others,
for approaches to scrambling along these lines.

   Clearly, the introduction of features like [•F•] and [∗F∗] makes sense only if it is
assumed that these features must be checked, or discharged, thereby driving syntactic
operations. This can be formulated as the *Feature Condition*; see (5).[6]

---

[4] See Chomsky (2001b; 2008) for a different view with respect to external Merge; and Chomsky (2007;
2008) on an extension of this position to internal Merge.

[5] This represents a combination of notational conventions that can be found in Roberts & Roussou (2002),
Adger (2003), Sternefeld (2006), and Heck & Müller (2007).

[6] Throughout this book, I assume a *feature checking* rather than a *feature valuation* approach. There are
various reasons for doing so; here are two: First, the valuation approach is confronted with various kinds
of problems outside the realm of ϕ-features – e.g., it is far from clear what role valuation can play in the
domain of subcategorization, or A-bar related features triggering movement (the relevant information must
be located on both items that participate in the relation). Second, it seems to me that one can make a
case that the concept of valuation is inherently at variance with the Inclusiveness Condition (see Chomsky
(1995; 2000; 2001b; 2008) and below): Copied feature values lead to an increase in material that is part
of the derivation. (Similar considerations apply to the copy theory of movement, and my conclusion will
be the same in this case.) – That said, much of what follows could presumably be construed in a properly
extended valuation approach without too much ado. Another (albeit related) distinction that is made in
much recent work on features in the minimalist program concerns the status of a feature as interpretable or
uninterpretable. This distinction will play no role in what follows. In line with this, I remain uncommitted

(5)   *Feature Condition* (FC):
      A feature [●F●] or [∗F∗] must be discharged by the end of the derivation.

(5) only requires that operation-inducing features are discharged by the end of the derivation, but given the SCC, it follows that such discharge must take place very quickly in the derivation: If such a feature is not discharged on some lexical item X before the next head Y is merged with X's projection, ungrammaticality will arise: Either the feature on X is checked later, in violation of the SCC; or the feature on X remains unchecked throughout the derivation, in violation of the FC.

   In addition, I assume there is a complementary Last Resort condition which requires all syntactic operations to be driven by (structure-building or probe) features; see (6).

(6)   *Last Resort* (LR):
      Every syntactic operation must discharge either [●F●] or [∗F∗].

Do FC and LR meet the requirements for good constraints, in the sense made precise in chapter 1? I think they do. Given the very existence of [●F●] and [∗F∗] features, the two constraints implement what looks like the simplest and most economical way to put them to use. (In addition, they are simple, general, and not complex; they do not increase search space, and they do not give rise to redundancies, all in accordance with the requirements in (11) of chapter 2.)

   The second assumption that will prove important (for a maximal coverage of the available empirical evidence, at least) is that all phrases are phases, not just vP and CP.

2.3   All Phrases are Phases

Standardly, it is assumed that vP and CP are phases (see Chomsky (2000; 2001b; 2008)), plus perhaps DP (see Svenonius (2004), Heck & Zimmermann (2004), Matushansky (2005), Kramer (2007), Heck et al. (2008)), and possibly TP as a (parametrized) alternative (see Richards (2004; 2011a)). In contrast, it has also been argued that all XPs qualify as locality domains for movement (see Koster (1978), Riemsdijk (1978)). Successive-cyclic movement via all intermediate XP edges has been assumed, inter alia, in Sportiche (1989), Takahashi (1994), Agbayani (1998); Chomsky (1995; 2005; 2008), Bošković (2002), Boeckx (2003), and Boeckx & Grohmann (2007).[7] Given the present set of assumptions as background, it seems that

---

as to whether all probe (more generally, operation-inducing) features have to be uniniterpretable or not.

[7] Also compare the concept of SLASH feature percolation in Gazdar (1981); Gazdar et al. (1985), and work in HPSG (also see subsection 4.2.1 of chapter 2); and the related concepts of "gap phrase" and "operator feature percolation" in Koster (2000) and Neeleman & van de Koot (2010), respectively.

the simplest way to bring about highly local successive-cyclic movement via phrase edges is to assume that all phrases are phases; and this is what I will do in what follows. Thus, if every XP is a phase, the PIC forces intermediate movement steps to all phrase edges.[8]

If all phrases are phases, the search space of derivations is reduced even further. Compare (7) with (4) (as before, the convention is that all underlined XPs are phases).

(7) *Search space under PIC given that every XP is a phase*:

                  SCC

a. $\overbrace{[_{\underline{YP}}...Y\ [_{\underline{XP}}...[_{X'}X}^{}\ \underbrace{}_{PIC}\ [_{\underline{WP}}...W\ [_{\underline{UP}}...U...\ ]]]]]$

                SCC

b. $\overbrace{[_{\underline{ZP}}...Z\ [_{\underline{YP}}...Y}^{}\ \underbrace{}_{PIC}\ [_{\underline{XP}}...\ [_{X'}X\ [_{\underline{WP}}...W\ [_{\underline{UP}}...U...\ ]]]]]]$

Furthermore, if all phrases are phases, the PIC is classified as an optimal constraint by *all* of the meta-requirements on good constraints listed in (11) of chapter 2 (also see the pertinent remarks on page 76); categorial selectivity, which might have been viewed as problematic from the point of view of requirement (11-a) (demanding maximal generality) is now abandoned.

Needless to say, such an approach makes it necessary to reconsider a substantial part of the evidence that has been brought forward in support of a less general, category-selective concept of phases. I cannot possibly provide comprehensive discussions of all the relevant issues here; I will confine myself to brief statements (in some cases only a few remarks) concerning the core cases.

First, phases have been correlated with cyclic spell-out domains. However, the first thing to note is that the correlation is not perfect (it is the complement domain of the phase head that is spelled out, rather than the phase as such – edge and head material must be available for operations on the next phase level). Moreover, domains for cyclic spell-out have also been argued to be potentially smaller than the classical phase (cf. Uriagereka (1999) and chapter 2 above), perhaps radically so (see Epstein & Seely (2002)).

Second, closer inspection of the literature on morphological and other reflexes of successive cyclicity reveals that that they are not always confined to the CP or vP domains (as has been argued to be the case with, e.g., complementizer selection in Modern Irish discussed in McCloskey (1979; 2002), or possibly the instances of *wh-*

---

[8] An alternative type of approach that I will not pursue here envisages phases as more flexible objects; see Grohmann (2000), Bobaljik & Wurmbrand (2003), Marušič (2005), Gallego & Uriagereka (2006) (see the previous chapter), den Dikken (2007), and Gallego (2007). Such a view is incompatible with the assumption that the main motivation for phases is to minimize active parts of derivations.

agreement in Chamorro discussed in Chung (1994)). For instance, focus movement in Ewe looks like a clear counter-example (at least at first sight) because the morphological reflex shows up in the TP domain (see Collins (1993; 1994) and (27) of chapter 1 above). Similarly, tonal downstep in Kikuyu is a reflex of successive-cyclic movement that does not exclusively identify CP or vP edges (cf. Clements, McCloskey, Maling & Zaenen (1983)). See Lahne (2009) for more general discussion, as well as chapter 4 below.

Third, as far as I can tell, assuming all intervening phrase edges to be landing sites for intermediate movement steps is fully compatible with evidence from reconstruction (see Fox (2000), Nissenbaum (2000a)).

Fourth, Abels (2003) advances an argument for designated intermediate landing sites that is based on reflexivization options in English, and that excludes the TP domain from the set of intermediate landing sites. I think that this argument can be shown to be inconclusive. Let me dwell on this for a moment. Abels notes that the enriched options of anaphoric binding resulting from successive-cyclic *wh*-movement in examples like (8-a) in English are neutral between an approach that singles out CP and vP as designated phases, and an approach that postulates that all intervening XP edges are targets of intermediate movement steps; all that is required is that the position identified by the box □ is used. However, so the argument goes, since movement from raising infinitives does not give rise to the same enrichment of options for anaphoric binding in English (see (8-b)), this can be taken to argue against the use of the edge of TP as an intermediate landing site (given that raising infinites are TPs, not CPs).

(8) a. Which pictures of himself$_1$ does Jane believe (that) John$_1$ thinks [$_{CP}$ □ (that) she likes t] ?

   b. *Which pictures of himself$_1$ did Mary$_2$ seem to John$_1$ [$_{TP}$ □ t$_2$ to like t] ?

   c. *Which pictures of himself$_1$ did Mary$_2$ seem to Jane$_3$ [$_{TP}$ t$_2$ to have told John$_1$ [$_{CP}$ □ that she likes t]] ?

Still, there are problems with this account. First, a precondition for the effect in (8-a) is that the moved *wh*-phrase with the reflexive in it has left the domain of the lower subject (*she* in (8-a)), which would otherwise continue to block binding from above. Since the lower subject is in SpecT in the most deeply embedded clause in (8-a), and the □ position is outside of this domain, a new binding option can be generated by *wh*-movement. However, things are different in (8-b): When *which picture of himself* in (8-b) moves to SpecT (assuming that it does because every phrase is a phase), the intervening subject *Mary* is still present in the same domain, and it continues to be present when *John* is merged. Therefore, *Mary* will intervene, and block binding of the reflexive by *John* (as a closer potential antecedent) even if *which picture of himself* moves to SpecT, assuming that two specifiers of the same head are in the

same minimal domain and may create intervention effects for each other. Since it is unlikely that there can be a subsequent step in the derivation where *Mary* ceases to be an intervener and permits binding of the reflexive by *John*, the evidence is fully compatible with a phase status of TP.

A second problem with Abels's (2003) argument for designated phases can be found in Boeckx & Grohmann (2007) and Boeckx (2008) (also see Abels & Bentzen (2011)): Sentences like (8-c) also lack the enrichment of binding options by movement to intermediate positions although the most deeply embedded clause is a CP, and movement to the $\square$ position of this CP domain should suffice for creating the new binding option. This would seem to suggest that the correct generalization is that an intervening experiencer blocks the enrichment of binding options, quite independently of the nature of the landing site involved.

Assuming that other arguments for designated phase categories can be refuted in the same way, I would like to contend that there is every reason (not least of all because phases are motivated on the basis of complexity considerations) to assume that all phrases are phases. Consequently, *wh*-movement must proceed via every XP edge domain on its way to its ultimate target position (the $C_{[\bullet wh \bullet]}$ node that attracts it), given the PIC.

This leads me to the third important assumption; it concerns the insertion of edge features.

2.4   The Edge Feature Condition

The PIC forces successive-cyclic movement via phase edges. Accordingly, it must be ensured that the intermediate movement steps that are required by the PIC are indeed permitted. Given the Last Resort principle according to which all syntactic operations are feature-driven (see above), movement to intermediate phase edges must be triggered by appropriate features. For the time being it can be assumed that edge features for intermediate movement steps (unlike the edge features for final movement steps which are relativized with respect to properties of the attracting head, such as [wh]) are blind to the properties of the attracted item. Here and henceforth, I notate edge features for intermediate movement steps as [$\bullet X \bullet$] (but see chapter 4 below for a qualification).[9]

---

[9] Throughout this book, I assume that intermediate movement steps involve genuine movement operations rather than counter-cyclic locality-driven trace insertion, as proposed in Chomsky (1993), Takahashi (1994), and Boeckx (2003); cf. the discussion of Form Chain in chapter 1. If movement to intermediate positions exists, it may be either feature-driven or forced in some other way, in minimal violation of Last Resort; see McCloskey (2002) vs. Heck & Müller (2000; 2003) and Müller (2004a) for the two options. I adopt the *former* approach in what follows. This change of perspective is responsible for most of the relevant differences between the present chapter and Müller (2004a), on which it is based – the latter study assumes

These edge features cannot be obligatorily present on phase heads because they lead to a crash if the number of items that need to be attracted to a phase edge in the course of successive-cyclic movement does not equal the number of movement-triggering features on the phase head. To ensure that edge features for intermediate movement steps are only present on a phase head if they are needed, a condition like (9) can be assumed (see Chomsky (2000, 109), Chomsky (2001b, 34), and Chomsky (2008, 149)). I refer to this condition as the Edge Feature Condition here and in what follows.

(9) *Edge Feature Condition* (EFC):
   The head X of phase XP may be assigned an edge feature after the phase XP is otherwise complete, *but only if that has an effect on outcome*.

Given (9), phase heads can be assigned additional (i.e., non-inherent) edge features in the course of the derivation "if that has an effect on outcome", i.e., if that serves to implement intermediate movement steps required by the PIC.[10] Of course, an important question is what it means for edge feature insertion to have "an effect on outcome". This requirement looks initially like a clear case of a piece of theoretical machinery that requires look-ahead: To find out whether an edge feature can be inserted on a phase head, thereby making attraction of some item to the phase edge possible, one must know whether the moved item will ultimately be needed as the specifier of some movement-inducing head. For instance, movement of some item α to the specifier of vP (as required by the PIC) should only be classified as legitimate if some other head (say, $C_{[\bullet wh\bullet]}$) makes α its specifier later in the derivation (i.e., in the case at hand, if α is a *wh*-phrase that must become a specifier of C).

The question arises of whether there is a sufficiently local way to determine whether features for intermediate successive-cyclic movement are needed or not, without actual look-ahead. The approach in terms of a constraint *Phase Balance* developed in Heck & Müller (2000; 2003), Fischer (2004), and Müller (2004a) is designed to accomplish this.[11] It correctly predicts under which circumstances intermediate movement steps are legitimate. However, it dispenses with edge features in intermediate

---

that Last Resort is minimally violable, which I do not assume here.

[10] Such an introduction of edge features in the derivation will presumably have to be construed as violating the Inclusiveness Condition (see Chomsky (1995; 2000; 2001b; 2008)): Something that is not originally part of the numeration is introduced in the course of the derivation. Thus, this reasoning would seem to imply that whereas the prohibition against non-feature driven movement is strict, the prohibition against feature insertion is not; in other words, the Inclusiveness Condition is minimally violable whereas Last Resort is not; see the last footnote.

[11] Also cf. Biskup's (2009) related concept of Phase Featuring.

posititions. In what follows, I sketch how the Phase Balance approach can be transferred into the present set of assumptions. The most important deviations from Heck & Müller (2000; 2003) are (i) that the new version of the approach relies on edge features for interemediate movement steps, and (ii) that Phase Balance is not understood as a constraint in its own right anymore, but rather as a concept that a revised Edge Feature Condition refers to (obviously the two changes are intimately related).

The concept of a *balanced phase* can be understood as in (10).

(10) *Balanced Phase*:
 A phase is balanced iff, for every movement-inducing feature [•F•] in the numeration, there is a distinct potentially available feature [F].

Thus, to find out whether a phase (i.e., a phrase, given the assumption that all phrases are phases) is balanced or not one has to check whether, for every movement-inducing feature [•F•] on a lexical item in the numeration, there is a matching feature [F] that qualifies as *potentially available*.[12] Potential availability is defined in (11).

(11) *Potential availability*:
 A feature [F] is potentially available at the XP level if (i) or (ii) holds:
 (i) [F] is on X or edgeX of the present root of the derivation.
 (ii) [F] is in the workspace of the derivation.

The workspace of a derivation D comprises the numeration N and material in trees that have been created earlier (with material from N) and have not yet been used in D.[13]

On this basis, the Edge Feature Condition in (9) can be revised as shown in (12). Basically, (12) differs from (9) only in that the informal requirement of *having an effect on outcome* (which requires look-ahead) is made precise as *being the only way to produce a balanced phase* (which does not require look-head).[14]

---

[12] (10) presupposes that one can distinguish between structure-building features that trigger external Merge (i.e., subcategorization features), and structure-building features that trigger internal Merge (i.e., movement-inducing features). Postulating such a difference seems unavoidable, and straightforward (see chapter 4 for a technical execution that relies on the hypothesis that features for structure-building show up on stacks). Still, it is worth bearing in mind that (10) could in principle also be formulated in terms of structure-building (rather than movement-inducing) features, even though this would make some reasonings slightly more complex.

[13] I do not assume that such trees are sent back to the numeration before re-use, as argued in Johnson (2003); see chapter 2.

[14] Or does it? The analysis makes crucial use of Chomsky's (1995; 2000; 2001) concept of a numeration (and the concept of workspace that is based on it). One might argue that this is another form of look-ahead since properties of items that are outside the currently available portions of structure of the derivation need

(12) *Edge Feature Condition* (EFC; revised):

   The head X of phase XP may be assigned an edge feature after the phase XP is
   otherwise complete, but only if that *is the only way to produce a balanced phase*.

Thus, edge feature insertion on phase heads emerges as a repair operation: It can take
place only in order to balance a phase that would otherwise not be balanced, so as to
attract some XP to the edge of the phase that will eventually be needed by some lexical
item that has not yet been merged. To see how the system works, let us go through a
derivation of a simple *wh*-question, first abstractly and then on the basis of an actual
example.

   Suppose that an object *wh*-question is going to be generated. In the case of *wh*-
movement, a *wh*-object must move to the edge domain of every intervening phase (=
phrase), because of the PIC. For these movement steps to be possible, category-neutral
edge features ([●X●]) must be inserted on phase heads, because of Last Resort (by as-
sumption, there is no movement that is not triggered by structure-building features).
Next, edge feature insertion on a phase head is legitimate only if there is no other way
to balance the phase, because of the EFC. To find out whether the phase would other-
wise be balanced or not, a look into the numeration (and the workspace) is necessary
– if it is not balanced without *wh*-movement, an edge feature can be inserted which
attracts the *wh*-phrase. Now, suppose first that the edge feature is not in fact inserted
at this point (the EFC says something about the conditions under which edge features
can be inserted, it does not say that they *have to be* inserted if they can be, so edge
feature insertion will always be optional at any given step of the derivation where it
can apply). Then, the PIC will block further movement of the *wh*-phrase on the next
cycle, and the derivation eventually crashes because [●wh●] on C cannot be checked.
Suppose, in contrast, that the edge feature is in fact inserted. Then, the derivation can
continue, and access the moved item in the edge domain of the previous phase. In gen-
eral, it thus follows that a *wh*-phrase can only undergo an intermediate movement step
if it is required by some (non-local) C head bearing [●wh●]; and that *wh*-movement

---

to be accessed. However, one should be aware that the two procedures are far from identical; here we
are dealing with a "look to the side", so to speak, rather than with look-ahead. There can be little doubt
that this kind of procedure is much more restricted than genuine look-ahead – it utilizes a concept that
has been proposed for independent reasons (viz., the numeration), and it does not have access to structural
information provided by later parts of the derivation. (For the same reason – lack of a syntactic structure
in which search must be carried out, sensitive to configurational notions of this syntactic structure – the
approach based on balanced phases is not subject to the conceptual argument against the (G)MLC based on
massive search space given in chapter 2.) Note that if the numeration is present, and accessible, anyway
throughout the derivation, the simplest assumption is that the derivation is not blind to its properties – if we
want to ensure that the derivation cannot look into the numeration, an additional stipulation to this effect is
needed.

must proceed via every XP on its way to its ultimate target position (the $C_{[\bullet wh\bullet]}$ node that attracts it).

(13) illustrates the relevant parts of a concrete derivation of a *wh*-question in English.[15] Suppose that the derivation has reached the stage where V has combined with the object *wh*-phrase DP_1. At this point, there are two lexical items in the workspace bearing movement-inducing features, viz., C ($[\bullet wh\bullet]$) and T ($[\bullet D\bullet]$, the EPP feature that triggers raising to subject position in English). Is VP balanced without edge feature insertion (and subsequent movement of DP_1 to SpecV)? It is not: Although there is a potentially available feature [D] in the numeration (hence, in the workspace) that satisfies the balanced phase requirement on VP for $[\bullet D\bullet]$ on T (the relevant [D] feature is the categorial feature of *John*, or of an empty D element that ultimately subcategorizes for, and combines with, *John*), there is no potentially available [wh] feature left in the numeration for $[\bullet wh\bullet]$ on C. Consequently, the VP phase is not balanced as such, and edge feature insertion can take place in accordance with the EFC in (12). This triggers the first intermediate movement step of DP_1 (*what*); see (13-a). Note that if no edge feature were inserted at this point, movement of DP_1 could not take place, and the derivation would ultimately violate the PIC (either on one of the next cycles – vP, TP – if an edge feature is inserted for DP_1 on v or T, or on the final cycle, when the Feature Condition demands checking of $[\bullet wh\bullet]$ on C). Thus, extraction from VP can only take place via the edge of V.[16]

(13) (I wonder) what John read

    a.   [$_{VP}$ what$_1$ read$_3$ t$_1$]

                               $\rightarrow$ workspace: {$C_{[\bullet wh\bullet]}$, John, $T_{[\bullet D\bullet]}$, v}

    b.   [$_{vP}$ what$_1$ John$_2$ v+read$_3$ [$_{VP}$ t$'_1$ t$_3$ ~~t$_1$~~]]

                                 $\rightarrow$ workspace: {$C_{[\bullet wh\bullet]}$, $T_{[\bullet D\bullet]}$}

    c.   [$_{TP}$ what$_1$ John$_2$ T [$_{vP}$ t$''_1$ t$_2$ v+read$_3$ ~~[$_{VP}$ t$'_1$ t$_3$ t$_1$]~~]]

                                 $\rightarrow$ workspace: {$C_{[\bullet wh\bullet]}$}

    d.   [$_{CP}$ what$_1$ C [$_{TP}$ t$'''_1$ John$_2$ T ~~[$_{vP}$ t$''_1$ t$_2$ v+read$_3$ [$_{VP}$ t$'_1$ t$_3$ t$_1$]]~~]]

                                   $\rightarrow$ workspace: {−}

The derivation in (13-bcd) proceeds as expected. In (13-b), the external argument DP_2 (*John*) is merged, and DP_1 (*what*) moves to an outer Specv, as permitted by the

---

[15] Material that is crossed out is not accessible anymore for further operations in the derivation, because of the PIC.

[16] It can be noted that this is incompatible with Anti-Locality requirements on movement, as they have been advanced (in slightly different ways) by Grohmann (2000; 2003) and Abels (2003), among others: The first movement step of an object to SpecV is extremely local but indispensable, given that t$_1$ in (13-a) is not part of the edge domain of V.

revised EFC (and required by the PIC). On the TP cycle (see (13-c)), both DP_1 and DP_2 need to undergo movement – in order to directly discharge T's [•D•] feature, in the second case, and in order to make further movement to the ultimate landing site SpecC possible, in the first case (again, edge feature insertion is permitted because this is the only way to create a balanced phase – a balanced TP, in this case). Finally, on the CP cycle (see (13-d)), DP_2 undergoes feature-driven movement to its final landing site, thereby discharging [•wh•] on C and satisfying the Feature Condition.

This approach ensures that in situations where there are two categories bearing the same feature (say, two *wh*-phrases bearing [wh] each) but only one lexical item that has only one matching movement-inducing feature (like interrogative C, which can bear only a single [•wh•] in English-type languages that lack multiple *wh*-movement), only one of the two categories can undergo successive-cyclic movement.[17] What is more, the EFC in (12) and the hypothesis that only very small parts of syntactic derivations are active, and accessible, at any given step conspire so as to derive superiority effects, without recourse to a separate constraint like the (G)MLC: The prediction is that presence of an accessible [wh] feature in the workspace should make edge feature-driven *wh*-movement of a *wh*-phrase impossible, and the PIC should then block any further operations applying to this *wh*-phrase. The next section shows that this prediction is borne out, and that it offers a simple account of superiority effects in English, without recourse to a constraint like the (G)MLC.

3. Deriving the (Generalized) Minimal Link Condition

3.1 Superiority Effects in English

3.1.1 Standard Cases

Examples illustrating superiority effects in English are shown in (14) (for subject and object) and in (15) (for two objects): Given two *wh*-phrases that might in principle qualify as goals for movement operations targetting a single $C_{[•wh•]}$ probe, only the higher *wh*-phrase can undergo movement to the target position (see (78), (79) of chapter 1).

(14) a. (I wonder) who$_1$ bought what$_2$
　　 b. *(I wonder) what$_2$ who$_1$ bought t$_2$

---

[17] In this context, note that whereas some strengthened version of Chomsky's (2001b) Activity Condition might be integrated with the EFC to achieve some of the effects of the present formulation relying on balanced phases ("an edge feature can be inserted only for a category that is active", i.e., has certain kinds of unchecked features), it is hard to see how the Activity Condition could help to ensure that, say, only one of several *wh*-phrases can undergo intermediate steps on the way to a single interrogative C in English-type languages.

(15) a. Who$_1$ did you persuade t$_1$ [$_{CP}$ to read what$_2$] ?

     b. *What$_2$ did you persuade who$_1$ [$_{CP}$ to read t$_2$] ?

The analysis is straightforward under present assumptions. The illegitimacy of *wh*-movement applying to DP$_2$ in (14-b) and (15-b) is determined very early in the derivation, in the VP phase in which DP$_2$ is externally merged. Given the PIC, DP$_2$ needs to reach SpecV first if it is to move on. However, in these cases, reaching SpecV is impossible for DP$_2$. There are two possibilities to be considered. First, movement of DP$_2$ to SpecV (similarly, to Specv, SpecT, ...) might be triggered by inherent movement-inducing features on lexical items ([•F•]). This is not an option: English does not have object shift or scrambling. Second, movement of DP$_2$ to SpecV (similarly, to Specv, SpecT, ...) might be triggered by an edge feature. However, it is not possible to insert an edge feature on V without violating the EFC: VP is balanced because there is another *wh*-phrase in the workspace, viz., DP$_1$; similarly, vP and TP in (14) are balanced because DP$_1$ occupies the respective specifiers (but given the PIC, at this point edge feature insertion would come too late for DP$_2$ anyway). Hence, there is no way for DP$_2$ in (14-b) and (15-b) to move to SpecV, and a subsequent PIC violation that occurs if *wh*-extraction of DP$_2$ takes place from the opaque domain of a phase is fatal. The only way the derivation can proceed in (14) is by leaving the internal argument DP$_2$ in situ, merging the external argument in Specv, moving it to SpecT as a consequence of T's EPP feature, and then on to SpecC, thereby discharging the movement-inducing inherent edge feature [•wh•] on C. This is shown in (16).

(16) a. [$_{VP}$ bought$_3$ what$_2$]

$\rightarrow$ workspace: $\{C_{[•wh•]}, \text{who}_1, T_{[•D•]}, v\}$

     b. [$_{vP}$ who$_1$ v+bought$_3$ [$_{VP}$ t$_3$ ~~what$_2$~~]]

$\rightarrow$ workspace: $\{C_{[•wh•]}, T_{[•D•]},\}$

     c. [$_{TP}$ who$_1$ T [$_{vP}$ t$_1$ v+bought$_3$ ~~[$_{VP}$ t$_3$ what$_2$]]~~]

$\rightarrow$ workspace: $\{C_{[•wh•]}\}$

     d. [$_{CP}$ who$_1$ C [$_{TP}$ t$_1''$ T ~~[$_{vP}$ t$_1$ v+bought$_3$ [$_{VP}$ t$_3$ what$_2$]]~~]]

$\rightarrow$ workspace: $\{-\}$

3.1.2 Double Object Constructions

Double object constructions provide an interesting testing ground for approaches to superiority effects. Constructions with ditransitive verbs basically come in two varieties in English: the *prepositional object construction* and the *dative shift construction*. As far as the prepositional object construction is concerned, it has been observed by Chomsky (1973, 246) and Fiengo (1980, 123) that either object (DP or PP) can move if both are *wh*-phrases (see (17-ab)); however, preposition stranding (which is legitimate if the remaining DP object is not a *wh*-phrase) becomes impossible in this context (see (17-c)).

(17) a. What$_1$ did you give t$_1$ to whom$_2$?

     b. To whom$_3$ did you give what$_1$ t$_3$?

     c. *Who$_2$ did you give what$_1$ [$_{PP_3}$ to t$_2$] ?

The situation is different in dative shift constructions with two *wh*-objects. Barss & Lasnik (1986, 349) note that the second object DP can never move in this context; the first, dative-shifted object DP can marginally move.[18]

(18) a. ?Who$_2$ did you give t$_2$ what$_1$?

     b. *What$_1$ did you give who$_2$ t$_1$?

To account for these data, we need to say something about the structure of double object constructions in English, and about pied piping in *wh*-PPs. Modifying the proposal in Larson (1988), I assume that the direct (i.e., THEME) object is merged in a complement position of V, where it remains in both types of double object constructions (unless it undergoes movement to the clausal periphery). The indirect (i.e., GOAL) object is at the edge of V if it has undergone dative shift (see (19-b)), and in a right-peripheral V′ sister position if it is prepositional (see (19-a)).[19] Note that this right-peripheral position does not belong to the edge of V.[20]

(19) a.
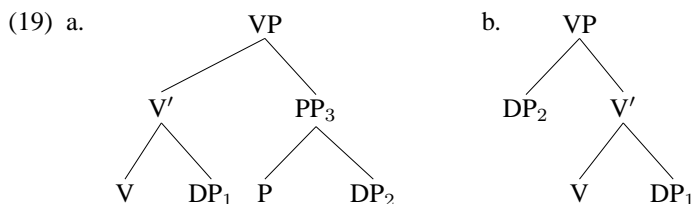
With respect to pied piping, I assume that there is optional percolation of the feature [wh] in *wh*-PPs; for present purposes, this percolation operation can be conceived of as an actual feature displacement.[21]

---

[18] The marginality is due to a general weak ban on A-bar movement of dative-shifted objects in English and thus independent of superiority; see Stowell (1981, ch. 4) and Larson (1988), among others.

[19] To accomodate evidence from binding theory, we must then assume that linear order plus m-command is relevant; see Barss & Lasnik (1986) and Jackendoff (1990).

[20] However, if we follow Chomsky (2002, 133-136), this position will invariably be a specifier (i.e., non-first Merge) position. Thus, as will be shown momentarily, at this point it is crucial that (3) refers to (the left-peripheral domain) edgeX rather than to SpecX.

[21] Heck (2004; 2008; 2009) and Cable (2007) develop more elaborate theories of pied piping that do without feature percolation. These approaches could presumably be reconciled with the present analysis, but I will refrain from attempting that here.

Consider now first the examples involving prepositional object constructions in (17). Given that feature percolation of [wh] from DP to PP is optional, we have to take into account two possibilities. First, suppose that [wh] percolation has taken place, and PP bears [wh]. The two objects are merged in VP-internal non-edge positions; see (19-a). Hence (given that there is no other [wh] waiting in the workspace), edge feature insertion plus movement to SpecV is possible only for one *wh*-phrase so as to balance the VP (there is one [●wh●] on C in the numeration). It does not matter which of the two *wh*-phrases moves to SpecV, but whichever *wh*-phrase moves first forces the other *wh*-phrase to stay in situ, to avoid a violation of Last Resort or the EFC. The *wh*-phrase in SpecV is then passed on through further cycles of the derivation, until CP is reached and [●wh●] on C is checked. This way, (17-a) and (17-b) can both emerge as grammatical.

Consider now the second option: [wh] percolation from DP to PP has not taken place. Then, PP cannot move to SpecV – if it moves as a result of edge feature insertion, the phase will not be balanced because [wh] is not potentially available at the VP level since it is not part of the edge of V (it is dominated by an edge element – PP – but not on an edge element itself). DP_2 in PP cannot move either, though: To leave PP, DP_2 must move to SpecP, given the PIC. However, this operation is not legitimate because there is no [●F●] that might trigger it (English does not have an independent PP-internal preposing operation), and because the phase is independently balanced (with another *wh*-phrase in the workspace). Therefore, the superiority effect in (17-c) is correctly predicted.

Turning next to dative shift constructions as in (18), it follows from structure (b) in (19) that DP_2 is in SpecV for independent reasons.[22] Hence, the phase is balanced from the start, and an edge feature cannot be inserted – therefore, movement of the lower *wh*-phrase is impossible. Consequently, DP_2 can undergo *wh*-movement (see (18-a)), but DP_1 cannot undergo such movement, because of the PIC (see (18-b)).

The analysis makes a further prediction: If both *wh*-phrases are embedded in PPs, preposition stranding is predicted to be blocked throughout: It is impossible for an embedded *wh*-phrase to move to SpecP in this context because the phase is always balanced without such movement, and edge feature insertion is impossible. Here is why. For the first *wh*-phrase DP_i that is merged with P (be it DP_1 or DP_2), the PP is balanced without local inversion of DP_i to the edge of P because there is a [wh] feature on another *wh*-item left in the numeration. For the second *wh*-phrase DP_j that is merged with P, movement to the edge of PP will also be blocked because there is now

---

[22] There are two possibilities: Either DP_2 is merged in SpecV, or it is moved there because of some [●F●] feature that triggers dative shift to that position. The present analysis is compatible with both a base-generation and a movement approach to dative shift constructions.

invariably a tree in the workspace that contains (or is) a *wh*-phrase bearing a [wh] feature. Consequently, no *wh*-phrase can move to SpecP in this context, and subsequent movement of such a *wh*-phrase from its base position will fatally violate the PIC. By and large, this prediction seems to be tenable, as the data in (20) illustrate.[23]

(20) a. ?*Who$_2$ did you give [$_{DP}$ pictures of t$_2$] [$_{PP}$ to whom$_1$] ?
      b. ?*Who$_1$ did you give [$_{DP}$ pictures of whom$_2$] [$_{PP}$ to t$_1$] ?
      c. ?*Who$_2$ did you talk [$_{PP}$ to t$_2$] [$_{PP}$ about whom$_1$] ?
      d. ?*Who$_1$ did you talk [$_{PP}$ to whom$_2$] [$_{PP}$ about t$_1$] ?

To sum up this subsection, the present approach accounts both for standard superiority effects in English, and their absence in certain kinds of double object constructions, without invoking the (G)MLC, by the interaction of the EFC and the PIC. I will now turn to the situation in German.

3.2   Lack of Superiority Effects in German

It has often been observed that German does not exhibit superiority effects with *wh*-phrases that are clause-mates; see Haider (1983; 1993a; 2000b), Grewendorf (1988), and Bayer (1990), among many others. A relevant pair of examples involving a *wh*-subject DP and a *wh*-object DP is given in (21).

(21) a. (Ich weiß nicht) wer$_1$ C t$_1$ was$_2$ gesagt hat
         I   know not   who$_{nom}$ what$_{acc}$ said   has

      b. (Ich weiß nicht) was$_2$ C wer$_1$ t$_2$ gesagt hat
         I   know not   what$_{acc}$ who$_{nom}$ said   has

Similarly, German does not exhibit superiority effects with control infinitives; see Fanselow (1991), Kim & Sternefeld (1997), and Haider (2000b).[24] This is shown in (22):

---

[23] It should be noted, however, that there is some disagreement about the status of these examples. Sentences like (20-d) are classified as ill formed in Jackendoff (1990, 433)), and as well formed in Fiengo (1980, 124). Furthermore, (20-a) and (20-c) are classified as acceptable by Jackendoff; but note that these examples are in fact expected to involve an additional violation of the *Clause Nonfinal Incomplete Constituent Constraint* (or whatever ultimately derives this constraint, which – as we have seen – does not qualify as 'good' in the sense of (11) of chapter 2; see Kuno (1973, 379), Lasnik & Saito (1992, 91), and (83) in chapter 1). Recall that this constraint is operative independently of multiple-*wh* (superiority) contexts; see (i-a) vs. (i-b).

(i) a.   Who$_2$ did you give [$_{DP}$ pictures of Mary] [$_{PP}$ to t$_2$] ?
     b. ?*Who$_1$ did you give [$_{DP}$ pictures of t$_1$] [$_{PP}$ to John] ?

[24] However, see Haider (2000b, 239) for an additional dissimilarity requirement on the two *wh*-phrases.

(22) a. (Ich weiß nicht) wen$_1$ er t$_1$ überzeugt hat [was$_2$ zu kaufen]
          I   know not   whom$_{acc}$ he convinced has   what$_{acc}$ to buy

     b. (Ich weiß nicht) was$_2$ er wen$_1$ überzeugt hat [t$_2$ zu kaufen]
          I   know not   what$_{acc}$ he whom$_{acc}$ convinced has   to buy

Various accounts of the lack of superiority effects with two *wh*-phrases that share a minimal finite clause have been given in the literature. I will here adopt an analysis that has been suggested by Fanselow (1996) and Grohmann (1997) (who assume that the (G)MLC underlies superiority effects). German has scrambling: A lower *wh*-phrase can independently be moved to a higher position, by *wh*-scrambling.[25] Thus, a lower *wh*-phrase cannot move across another *wh*-phrase merged in a higher cycle by *wh*-movement, given the EFC and the PIC; but it can do so by scrambling. To implement this analysis in the present approach, the assumption becomes necessary that scrambling is triggered by a designated optional feature (or feature bundle) that we can refer to as [•Σ•]; accordingly, scrambled items bear [Σ] features (see subsection 2.2 above). For our present concerns, it is immaterial whether [Σ] is a formal feature that is not interpreted, or can in fact be shown to be related to contentful notions that are sometimes viewed as triggers for scrambling (definiteness, specificity, animacy, focus, defocussing, and the like). The derivation of a sentence like (21-b) can then proceed as in (23), where the *wh*-object DP$_2$ first undergoes edge feature-

---

[25] The basic idea can already be found in the analysis of the lack of weak crossover effects in German that is developed in Grewendorf (1988, 320). As noted, there are many other accounts of the lack of superiority effects in German. Here is a (possibly incomplete) list. According to a first view, superiority effects in English result from a lack of (proper) government after LF *wh*-movement (an ECP approach; see chapter 1). Assuming that all arguments are governed in German, it follows that the effect can be avoided (see Haider (1983; 1993a)). A second approach also holds that superiority effects result from a lack of (proper) government (again, an ECP approach). The assumption is that C in verb-second languages can govern the subject position, which makes it possible to circumvent the effect (see Noonan (1988)). Third, it has been argued that superiority effects result from a TP barrier that restricts LF movement of the in-situ *wh*-phrase (this amounts to yet another ECP account). The LF barrier does not exist in German, for independent reasons having to do with the combined verb-second/semi-pro-drop nature of the language (see Müller (1995)). Fourth, it has been argued that superiority effects result from a feature-based version of Relativized Minimality (see chapter 1): A *wh*-element in situ must not c-command the trace of a *wh*-element whose overt licensing features are non-distinct (see Haider (2000b), based on Fanselow (1990a; 1991)). On this view, German can avoid superiority effects because Case is a licensing feature and overt Case forms are often distinct; this is not an option in English, where the Case system is reduced. Fifth, against the background of the (standard) assumption that superiority effects result from the (G)MLC, it has been claimed that German has variable base generation of arguments, which then helps avoid the effect (see Kim & Sternefeld (1997)), based on (an earlier version of) Fanselow (2001a)). Such an approach is close to the one adopted in the main text, the main difference being that the required optional permutation of arguments is brought about by scrambling in one case, by variable base-generation in the other. A sixth view also holds that superiority effects result from the (G)MLC. An interrogative C in English attracts all *wh*-phrases, but only the first one that is attracted by C is PF-realized (see Pesetsky (2000), Grewendorf (2001)). In such an approach, German can be assumed to have a different C that requires only one *wh*-phrase in SpecC and

driven movement to SpecV (because of [•Σ•] on v, not because of [•wh•] on C), and then Feature Condition-driven movement to Specv (because of [•Σ•]). At the vP level, both *wh*-phrases show up at the edge (they are both left-peripheral specifiers); hence, DP$_1$ and DP$_2$ are both in principle eligible for further movement. Such further movement is required by the PIC and permitted by the EFC on the TP cycle, and it is required by the Feature Condition in the final step (CP).[26] The well-formed derivation of (21-b) is given in (23).

(23) a.   [$_{VP}$ was$_{2,\Sigma}$ [$_{V'}$ t$_2$ gesagt]]

$\rightarrow$ workspace: {C$_{[•wh•]}$, wer$_1$, T, [$_v$ hat]$_{[•\Sigma•]}$}

  b.   [$_{vP}$ was$_{2,\Sigma}$ wer$_1$ [$_{VP}$ t$'_2$ ~~[$_{V'}$ t$_2$~~ gesagt]] [$_v$ hat]]

$\rightarrow$ workspace: {C$_{[•wh•]}$, T}

  c.   [$_{TP}$ was$_{2,\Sigma}$ [$_{vP}$ t$''_2$ wer$_1$ ~~[$_{VP}$ t$'_2$ [$_{V'}$ t$_2$ gesagt]]~~ [$_v$ hat]] T]

$\rightarrow$ workspace: {C$_{[•wh•]}$}

  d.   [$_{CP}$ was$_{2,\Sigma}$ C [$_{TP}$ t$'''_2$ ~~[$_{vP}$ t$''_2$ wer$_1$ [$_{VP}$ t$'_2$ [$_{V'}$ t$_2$ gesagt]] [$_v$ hat]]~~ T]

$\rightarrow$ workspace: {$-$}

This concludes the account of the lack of superiority effects in German in the present approach. There are certain restrictions on *wh*-in situ in German that are not covered by the present approach (see Haider (2000b; 2004)); but these effects are of a clearly different nature, and integrating them into the present analysis would not affect the gist of the proposal.[27]

The prediction is that the PIC should give rise to superiority effects after all in German if the lower *wh*-phrase cannot be moved to the domain occupied by the higher *wh*-phrase because scrambling is not available (for whatever reason). The following three subsections highlight three contexts where *wh*-scrambling is impossible in German; and it is in these contexts that superiority effects do indeed occur.

---

attracts the *wh*-features of the others. If feature attraction precedes phrase attraction, superiority violations can occur. Seventh and finally, Fanselow (1991), Wiltschko (1997), Grohmann (1998), Featherston (2001), and Clifton, Fanselow & Frazier (2006) suggest (various kinds of) re-evaluations of the empirical evidence.

[26] As noted in Fanselow (1990b), Müller & Sternefeld (1993), and elsewhere, *wh*-scrambling often leads to reduced acceptability (but not strict ungrammaticality), which is not attested in cases like (21-b) and (22-b). However, reduced acceptability may result not from the application of *wh*-scrambling as such, but from the surface position of a scrambled *wh*-phrase. Since *wh*-scrambling is subsequently undone in a derivation like (23), this restriction will not apply.

[27] English does not have scrambling; but it does exhibit topicalization. Thus, it has to be ensured that the account of superiority effects in English is not undermined by intermediate *wh*-topicalization. Indeed, *wh*-topicalization is independently excluded in English (and other languages); see Epstein (1992) and Müller & Sternefeld (1996) for analyses and further references.

3.3   Superiority Effects with Long-Distance Movement in German

The first such context is well known: As observed by Frey (1993), Büring & Hartmann (1994), Fanselow (1996), Heck & Müller (2000), Pesetsky (2000), and others, German does exhibit superiority effects with long-distance movement. This is shown by the contrast in (24).

(24) a. Wer$_1$ hat t$_1$ geglaubt [$_{CP}$ dass der Fritz wen$_2$ mag] ?
         who$_{nom}$ has   believed   that the Fritz whom$_{acc}$ likes

    b. *Wen$_2$ hat wer$_1$ geglaubt [$_{CP}$ dass der Fritz t$_2$ mag] ?
       whom$_{acc}$ has who$_{nom}$ believed   that the Fritz   likes

The analysis works as follows. First, as before, DP$_2$'s [wh] feature in (24) does not permit movement: the VP phase in which DP$_2$ is externally merged is balanced because of the presence of DP$_1$ in the workspace; therefore, edge feature insertion will be unavailable, and movement of DP$_2$ for the purposes of [wh] will fatally violate Last Resort. Second, and more importantly in the present context, DP$_2$ cannot move by scrambling either: Scrambling cannot leave a finite CP in German. Consequently, an embedded *wh*-phrase is correctly predicted to be stuck in the embedded clause if there is another *wh*-phrase in the workspace that is eventually merged in the matrix clause.[28]

3.4   Superiority Effects with Subject Raising in German

The second context in which there are a priori unexpected superiority effects in German involves subject raising.[29] DP raising to subject position is optional in German (see Grewendorf (1989), Diesing (1992)). In the present approach, this implies that the EPP feature [●D●] is optional on T. However, as shown in Haider (1993a, ch. 8), the evidence cited in Diesing (1992) and much related work in favour of subject raising to SpecT (based on phenomena like particle placement) is far from conclusive. As far as I can see, there is only one context where it is clear that subject raising to SpecT must have occurred in German (see Müller (2001, 296)): Unstressed pronouns must be at the *phonological* border of vP (in the sense of Chomsky (2001b, 34)), i.e., they cannot be preceded by non-pronominal material within TP (in contrast, stressed

---

[28] As a matter of fact, only v and V can tolerate [Σ] in their edge domains in German; no other kind of head provides a scrambling domain (see Müller (1995)). Hence, there can be no [●Σ●] feature on T or C that could trigger movement beyond vP. Still, something extra will ultimately have to be said to derive the ban on long-distance scrambling in German in toto: It must be ensured that a [●Σ●] feature on a matrix V or v cannot attract an XP bearing [Σ] in the lower clause (via successive instances of edge feature insertion). There are various ways of achieving this; but I will not pursue the matter here.

[29] To the best of my knowledge, this is not a standard assumption; but an informal survey has shown that the data are quite robust.

pronouns behave like non-pronominal DPs). There is but one exception: The subject DP, and only the subject DP, can optionally precede these pronouns within TP. This strongly suggests a special position that is available only for subject DPs. Hence, we can conclude that if a subject DP precedes unstressed pronouns, it must have undergone optional movement to SpecT (also see Fanselow (2004)).

Interestingly, there is a superiority effect in exactly this context. Since we need an unstressed object pronoun to ensure that subject raising has taken place, relevant examples involve ditransitive verbs. The contrast in (25) shows that a dative *wh*-object DP cannot undergo *wh*-movement to SpecC if a *wh*-subject occurs in front of an unstressed accusative object pronoun.

(25) a. Wem$_2$ hat [$_{vP}$ es t$'_2$ wer$_1$ t$_2$ gegeben] ?
    whom$_{dat}$ has it$_{acc}$ who$_{nom}$ given

b.?*Wem$_2$ hat wer$_1$ [$_{vP}$ es t$'_2$ t$_1$ t$_2$ gegeben] ?
    whom$_{dat}$ has who$_{nom}$ it$_{acc}$ given

The contrast in (26) shows the same for an accusative *wh*-object DP and an unstressed dative object pronoun.

(26) a. Was$_2$ hat [$_{vP}$ ihm t$'_2$ wer$_1$ t$_2$ gegeben] ?
    what$_{acc}$ has him$_{dat}$ who$_{nom}$ given

b.?*Was$_2$ hat wer$_1$ [$_{vP}$ ihm t$'_2$ t$_1$ t$_2$ gegeben] ?
    what$_{acc}$ has who$_{nom}$ him$_{dat}$ given

This superiority effect follows under present assumptions: Suppose that a subject DP$_{[wh]}$ and an object DP$_{[wh]}$ are both in Specv at some stage of the derivation, and that T has an optional [•D•] feature. Then, TP is balanced (for [•wh•]) by feature-driven subject raising, and movement of the object DP would incur a fatal Last Resort violation because an edge feature cannot be inserted. Consequently, object movement to SpecC is blocked by the PIC.[30]

If this analysis is on the right track, we expect that a non-*wh*-subject DP in SpecT should, ceteris paribus, not block movement of an object *wh*-phrase. This is the case: Only a *wh*-subject DP in SpecT blocks *wh*-movement of an object DP; see (27).

(27) a. Wem$_2$ hat t$''_2$ der Fritz$_1$ [$_{vP}$ es t$'_2$ t$_1$ t$_2$ gegeben] ?
    whom$_{dat}$ has the Fritz$_{nom}$ it$_{acc}$ given

b. Was$_2$ hat t$''_2$ der Fritz$_1$ [$_{vP}$ ihm t$'_2$ t$_1$ t$_2$ gegeben] ?
    what$_{acc}$ has the Fritz$_{nom}$ him$_{dat}$ given

---

[30] Haider (2002) argues that Icelandic has optional subject raising, and that it exhibits superiority effects with subject DPs only when the subject DP is in SpecT, not when it is in Specv. This generalization can be derived in the same way.

Similarly, replacing the unstressed object pronoun with a non-pronominal object DP should void the superiority effect (other things being equal). The reason is that the *wh*-subject does not have to be in SpecT in this context (non-pronominal DPs do not have to be at the phonological border of vP). (28) shows that this prediction is borne out, too.

(28) a. Wem$_2$ hat [$_{vP}$ t$'_2$ wer$_1$ t$_2$ das Buch   gegeben] ?
        whom$_{dat}$ has   who$_{nom}$ the book$_{acc}$ given

    b. Was$_2$ hat [$_{vP}$ t$'_2$ wer$_1$ dem Fritz   t$_2$ gegeben] ?
        what$_{acc}$ has   who$_{nom}$ the Fritz$_{dat}$ given

3.5   Superiority Effects with Scrambling from Wh-XP in German

A third context in which superiority effects arise in German has been noted in Fanselow (1996). The construction involves a configuration where the two *wh*-phrases are initially not in a c-command relation (as they are in all the examples discussed thus far); rather, one dominates the other. More specifically, suppose that a *wh*-phrase PP$_1$ is dominated by a *wh*-phrase DP$_2$, as in *wieviele Bücher über wen* ('how many books about whom'). Suppose furthermore that PP$_1$ can be moved out of DP$_2$ without violating locality constraints. This implies that DP$_2$ is in object position when extraction takes place (otherwise, the CED would be violated that permits extraction from XP only if XP occupies a complement position; see chapter 2 and chapter 4). It also implies that DP$_2$ is embedded by a certain kind of verb (verbs like *lesen* ('read') permit extraction from DP, verbs like *zerstören* ('destroy') do not); that DP$_2$ is sufficiently non-specific (highly specific DPs like *welches Buch* ('which book') tend to block extraction, non-specific DPs like *wieviele Bücher* ('how many books') do not); etc. Then, a *wh*-PP$_1$ can be scrambled from a *wh*-DP$_2$ if there is a [Σ] feature on PP$_1$ and a [$\bullet\Sigma\bullet$] feature on a higher head (V or v). PP$_1$ moves to the edge of V (via the edges of N and D), driven either by the Feature Condition (if [$\bullet\Sigma\bullet$] is on V) or by edge feature insertion (made possible because it is the only way to balance the phase if [$\bullet\Sigma\bullet$] is on v). This stage of the derivation is depicted in (29).

(29) [$_{vP}$ [$_{PP_1}$ über wen] [$_{V'}$ [$_{DP_2}$ t$''_1$ wieviele   [$_{NP}$ t$'_1$ Bücher t$_1$]] lesen]]
            about whom   how many   books   read

Here, PP$_1$ occupies an edge position of VP, and DP$_2$ a complement position. As shown by the contrast in (30), it is indeed the case that only PP$_1$ can undergo further movement, as one might expect: Movement of DP$_2$ on the next (vP) cycle will have to violate the PIC, or so it seems.

(30) a. (Ich weiß nicht) [$_{PP_1}$ über wen] er [$_{DP_2}$ wieviele   Bücher t$_1$] lesen
         I   know not   about whom he   how many books   read
        will
        wants

b. ?*(Ich weiß nicht) [$_{DP_2}$ wieviele   Bücher t$_1$] er [$_{PP_1}$ über wen] t$_2$ lesen
    I   know not   how many books   he   about whom   read
    will
    wants

However, there is a gap in this reasoning: (29) closely mirrors the situation found with two *wh*-phrases that are co-arguments in German, which do not normally exhibit superiority effects. Thus, if DP$_2$ can also undergo scrambling in (29), it should be able to undergo further *wh*-movement to SpecC after all. Fanselow (1996) solves this problem by showing that a derivation of (30-b) that involves intermediate scrambling of DP$_2$ will invariably violate another constraint: the principle of *Unambiguous Domination* (see Müller (1998, 271), and page 42 above). Unambiguous Domination is essentially a constraint on the movement of remnant XPs, i.e., XPs from which movement has taken place. This constraint states that α-traces must not be α-dominated (in the domain of the head of the chain). I will come back to this issue at the end of this chapter, and then again in chapter 4. For now, we can confine ourselves to noting a consequence: A scrambling trace like t$_1$ in (30) must not be dominated by a category that has itself undergone scrambling. This precludes intermediate scrambling of DP$_2$ in (29). Consequently, any derivation of (30-b) will have to violate either Unambiguous Domination or the PIC, depending on whether intermediate scrambling of DP$_2$ does or does not take place.

No such effect is predicted to occur if PP$_1$ is not a *wh*-phrase. Now, DP$_2$ can (in fact, must, given the PIC) move to SpecV in (29); this movement is not an instance of scrambling because there is no [•Σ•] involved (be it directly or indirectly). Compare (30-b) with (31).

(31) (Ich weiß nicht) [$_{DP_2}$ wieviele   Bücher t$_1$] er [$_{PP_1}$ über die Liebe] t$_2$ lesen
    I   know not   how many books   he   about love   read
    will
    wants

To sum up, German does exhibit superiority effects in certain contexts. These contexts have in common that intermediate scrambling of the second *wh*-phrase is not available, for independent reasons (scrambling in German cannot leave a finite clause, cannot target TP, and cannot apply to XPs from which scrambling has taken place). The effects are then derivable from the PIC.

4.   Intervention Effects that do Not Follow From the (G)MLC

4.1   *Long-Distance Intervention without C-Command in German*

The three types of superiority effects in German that have been discussed in subsections 3.3, 3.4, and 3.5 as such do not differentiate between PIC-based and (G)MLC-based analyses. However, it is worth noting that, in stark contrast to what is the case

with an (G)MLC account, there is nothing in the PIC-based account that would tie the intervention effect incurred by a *wh*-phrase wh_1 for another *wh*-phrase wh_2 to a *c-command* (or *dominance*) relation between the two. All that is needed for an intervention effect to arise in the PIC-based analysis is that wh_1 enters the derivation that wh_2 is part of at a later stage, and wh_2 cannot end up in the same edge domain as wh_1 by some independently motivated movement operation. Consequently, we expect that there should be *wh*-intervention effects without c-command.

As noted in Heck & Müller (2000; 2003) (originally in response to a question by Gisbert Fanselow (p.c.)), it is indeed the case that non-c-commanding *wh*-phrases in a matrix clause block long-distance *wh*-movement in German. This superiority-like effect without c-command is exemplified by the contrast in (32). In (32-a), there is clause-bound *wh*-movement of DP_1 across an adverbial CP that contains another *wh*-phrase DP_2, and that is merged later; here, an intervention effect can be avoided because DP_1 can reach a position in the same edge domain as the adverbial CP by scrambling. However, the option of intermediate scrambling is not available for long-distance *wh*-movement; scrambling must stop in the embedded vP domain. Thus, the presence of DP_2 in the workspace blocks edge feature-driven movement of DP_1, and (32-b) emerges as ungrammatical because of the PIC.[31] In (32-c), it is DP_2 rather than DP_1 that undergoes *wh*-movement; the result is also ill formed. As in (32-b), a PIC violation cannot be avoided here: No matter whether the adverbial CP is created before or after the object CP, the need to balance a phase cannot trigger edge feature insertion and concomitant successive-cyclic movement of DP_2 because DP_1's [wh] feature is potentially available for $C_{[\bullet wh \bullet]}$ in the workspace. In addition, sentences like (32-c) are ruled out by the CED (or whatever derives it; see chapter 4): Movement of DP_2 takes place from an adverbial CP that does not occupy a complement position.[32] The overall result is that the numeration underlying (32-b) and (32-c) cannot yield a well-formed output.[33]

---

[31] This presupposes that *wh*-features are visible in the workspace even if they show up in a position which would not be accessible under the PIC. This is unproblematic since checking whether a phase is balanced is not a syntactic "operation outside XP" in the sense of (2) (like external or internal Merge).

[32] Does this instantiate a case of unwanted redundancy, in the sense of (11-e) of chapter 2? The answer is no: As I will argue in chapter 4, CED effects are also derivable from the PIC. So minimally one can conclude that it is one and the same constraint which is violated here – albeit in two different ways. What is more, it will turn out that the CED-type PIC effect in (32-c) could only arise at the CP (adjunct clause) level; however, at this point, the fatal (G)MLC-type PIC effect has long occurred.

[33] This means that, if nothing else is said, examples of the type in (32-c), where a *wh*-phrase XP_2 is embedded in some other phrase that c-commands the *wh*-phrase XP_1, are predicted to be ill formed even if movement of XP_2 to $SpecC_{[\bullet wh \bullet]}$ does not violate the CED or another locality constraint, as long as XP_2 cannot reach the main branch by some other movement operation like scrambling. Relevant examples

(32) a. Wen$_1$ hat Fritz [$_{CP}$ nachdem er was$_2$ gemacht hat] t$_1$ getroffen ?
         whom has Fritz   after   he what done   has   met

   b. *Wen$_1$ hat Fritz [$_{CP}$ nachdem er was$_2$ gemacht hat] gesagt [$_{CP}$ dass Maria
       whom has Fritz   after   he what done   has said   that Maria
       t$_1$ liebt] ?
          loves

   c. *Was$_2$ hat Fritz [$_{CP}$ nachdem er t$_2$ gemacht hat] gesagt [$_{CP}$ dass Maria
       what has Fritz   after   he done   has said   that Maria
       wen$_1$ liebt] ?
       whom loves

Exactly the same reasoning applies in (33), where DP$_2$ shows up in a relative clause CP that is in turn dominated by an DP:

(33) a. Wen$_1$ hat Fritz [$_{DP}$ einem Mann [$_{CP}$ der was$_2$ kennt]] t$_1$
         whom$_{acc}$ has Fritz   a   man$_{dat}$ that what knows
         vorgestellt ?
         introduced

   b. *Wen$_1$ hat Fritz [$_{DP}$ einem Mann [$_{CP}$ der was$_2$ kennt]] gesagt [$_{CP}$
       whom$_{acc}$ has Fritz   a   man$_{dat}$ that what knows   said
       dass er t$_1$ einladen soll] ?
       that he invite   should

   c. *Was$_2$ hat Fritz [$_{DP}$ einem Mann [$_{CP}$ der t$_2$ kennt]] gesagt [$_{CP}$ dass er
       what$_{acc}$ has Fritz   a   man$_{dat}$ that knows   said   that he

---

are hard to find, though. In most pertinent cases, XP$_2$ will have to cross an island, and in the few well-formed constructions where locality constraints can be respected, XP$_2$ can usually undergo scrambling first (given some proviso concerning the coherence/incoherence distinction with control infinitives). Also recall the discussion of the examples in (20) in English. However, should there turn out to be clear cases of well-formed instantiations of the structure in (i) (where α is not an island and wh_2 cannot reach γ by an independently available non-*wh*-movement operation), the present approach would be in need of a modification.

(i)   ... wh$_2$ ... [$_\gamma$ [$_\alpha$ ... t$_2$ ...] ... [$_\beta$ ... wh$_1$ ...] ... ...] ...

One possibility would be to make the definition of workspace of a derivation sensitive to the distinction between main and minor branches (such that features on the main branch would not count as potentially available when a derivation proceeds in a minor branch). Then, structures like (32-c) would not (have to) violate the PIC anymore, and *wh*-movement in (i) would be predicted to be legitimate if α is not an island. – It might also be worth noting at this point that replacing requirement (ii) in the definition of potential availability in (11) by the stricter requirement (ii)$'$ would lead to an approach that is very close to a (G)MLC-based system, with *wh*-intervention effects reduced to c-command environments.

(ii)$'$ [F] is on X or edgeX of a root in the workspace of the derivation (lexical items are trivial roots).

wen$_1$ einladen soll] ?
whom$_{acc}$ invite   should

Yet another set of examples that illustrates the same pattern is given in (34); here the intervening *wh*-phrase that blocks long-distance *wh*-movement is embedded in a simple DP.[34]

(34) a. Wen$_1$ hat Fritz [$_{DP}$ einem Freund   von wem$_2$] t$_1$ vorgestellt ?
          whom$_{acc}$ has Fritz   a   friend$_{dat}$ of   whom   introduced

      b. ?*Wen$_1$ hat Fritz [$_{DP}$ einem Freund   von wem$_2$] gesagt [$_{CP}$ dass Maria
          whom$_{acc}$ has Fritz   a   friend$_{dat}$ of   whom   said   that Maria
          t$_1$ liebt] ?
              loves

      c. *[$_{PP_2}$ Von wem]   hat Fritz [$_{DP}$ einem Freund   t$_2$] gesagt [$_{CP}$ dass
              of   whom$_{acc}$ has Fritz   a   friend$_{dat}$ said   that
          Maria wen$_1$ liebt] ?
          Maria whom$_{acc}$ loves

The prediction is that the same kind of superiority-like effect without c-command should not be detectable in subject raising constructions, in contrast to the intervention effect with c-command in examples like (25-b) and (26-b) in German. The reason is that a *wh*-object can always reach the edge of vP by scrambling; and whereas a *wh*-DP in SpecT (as in (25-b) and (26-b)) renders vP balanced, a complex non-*wh*-DP in SpecT that includes a *wh*-phrase does not. Consider the examples in (35-a) and (35-b). It is not fully clear whether there is a contrast in acceptability here; I will assume that there is not.[35] At any rate, we can note that the present approach does not predict a different status of (35-a) and (35-b). In (35-a), the subject DP does not have to be in SpecT, and the *wh*-object DP can therefore move to a position in front of it by scrambling. This option is not available in (35-b), where the subject DP must be in SpecT (because of the presence of the unstressed object pronoun), i.e., in a domain that cannot be reached by scrambling. However, this does not preclude edge feature insertion on T for the object DP because a *wh*-item that is embedded in a specifier does not ensure phase balance. (An item α that is included in an item β which is in turn part of the minimal residue of some γ is not itself part of the minimal residue of γ; also see chapter 4.) Finally, note that (35-c) is excluded by the PIC; in addition, it

---

[34] In chapter 4, I will argue that sentences involving (apparent) *of/by*-phrase displacement are often not reliable extraction data: They may sometimes be given parses that do not involve extraction. Such an option does not exist in (34-c).

[35] The assumption in Müller (2004a, 311) was that there is, but it was noted there already that a possible distinction "may not be one of perfect wellformedness vs. absolute ungrammaticality". An informal survey among native speakers yielded inconclusive results.

is blocked by the CED (or what derives it) because the subject DP does not occupy a complement position.[36]

(35) a. ?Wem$_1$ hat [$_{DP}$ die Frau   [$_{PP_2}$ mit welchem Mantel]] t$_1$ ein Buch
       whom$_{dat}$ has   the woman   with which   coat   a book$_{acc}$
       gegeben ?
       given

    b. ?Wem$_1$ hat [$_{DP}$ die Frau   [$_{PP_2}$ mit welchem Mantel]] es   t$_1$
       whom$_{dat}$ has   the woman   with which   coat   it$_{acc}$
       gegeben ?
       given

    c. *[$_{PP_2}$ Mit welchem Mantel] hat [$_{DP}$ die Frau   t$_2$] es   wem$_1$
             with which   coat   has   the woman   it$_{acc}$ whom$_{dat}$
       gegeben ?
       given

4.2   Clause-Bound Intervention without C-Command in English

The system developed so far makes yet another prediction: There is no clause-bound intervention effect in the examples in (32-a), (33-a), (34-a), and (35-a) because German has scrambling. Since English does not have scrambling, we expect clause-bound intervention effects with non-commanding *wh*-phrases to occur. At first sight, this seems to contradict the standard view that argument *wh*-in situ in English does not obey any island constraints (see Chomsky (1981), Huang (1982), Huang (1995), Lasnik & Saito (1992), and Hornstein (1995), among others). However, most of the pertinent examples in the literature do not involve intervention without c-command: The typical kind of multiple *wh*-question that is taken to argue for non-island-sensitivity of *wh*-in situ in English has one *wh*-phrase embedded in an island, and a second *wh*-phrase merged in a higher position. This latter *wh*-phrase then undergoes movement to SpecC, as in the examples in (36), where a *wh*-phrase dominated by an object DP or by an adjunct PP does not block *wh*-movement of a subject *wh*-phrase merged later.

(36) a.   Who$_1$ t$_1$ saw [$_{DP}$ the man that bought what$_2$] ?
    b.   Who$_1$ t$_1$ likes [$_{DP}$ books that criticize who$_2$] ?
    c.   Who$_1$ t$_1$ bought [$_{DP}$ the books on which table$_2$] ?
    d.   Who$_1$ t$_1$ met [$_{DP}$ friends of whom$_2$] ?
    e.   I wonder who$_1$ t$_1$ heard [$_{DP}$ the claim that John had seen what$_2$]

---

[36] In principle, one would expect a superiority-like effect to also be absent with examples involving scrambling of a category that includes a *wh*-phrase, from a *wh*-XP, as in (30). However, relevant examples that would instantiate this pattern are difficult to construct because they would have to involve multiple embedding within DP, which creates difficulties of various kinds in multiple questions, for (presumably) independent reasons.

     f.   I wonder who$_1$ t$_1$ heard [$_{DP}$ John's stories about what$_2$]

     g.   Who$_1$ t$_1$ left [$_{PP}$ despite which warning$_2$] ?

Similarly, a *wh*-phrase that is part of a subject DP does not block movement of a *wh*-phrase that is merged in a higher clause, as in (37).

(37)   Who$_1$ t$_1$ thinks that [$_{DP}$ pictures of who$_2$] are on sale ?

All this is expected under present assumptions: DP$_2$ cannot undergo edge feature-driven movement in (36) or (37) early in the derivation because there is another item bearing [wh] left for [•wh•] of C in the workspace, and the PIC precludes Feature Condition-driven movement at the end of the derivation. DP$_1$, in contrast, undergoes movement from the edge of v to the edge of T in accordance with the EFC, and is then forced to end up at the edge of C by the Feature Condition.

    However, consider now the case where *wh*-intervention without c-command occurs. In the examples in (38), an object *wh*-phrase that is merged first (DP$_2$) moves across a subject DP containing another *wh*-phrase (DP$_1$). Such movement results in significantly reduced acceptability, as predicted under the present PIC-based approach.[37]

(38)   a. ?*Who$_2$ did [$_{DP}$ the man that bought what$_1$] see t$_2$?

       b. ?*Who$_2$ did [$_{DP}$ books that criticize who$_1$] impress t$_2$?

       c. ?*What$_2$ did [$_{DP}$ the books on which table$_1$] cost t$_2$?

       d. ?*Who$_2$ did [$_{DP}$ friends of whom$_1$] meet t$_2$?

       e. *Who$_2$ did [$_{DP}$ friends of whom$_1$] say that we should invite t$_2$?

The explanation is completely analogous to that given for the ungrammatical German examples involving intervention without c-command in the last subsection: Edge feature-driven movement of DP$_2$ to a specifier of V at an early stage in the derivation is impossible (because the VP is balanced anyway, with DP$_1$'s [wh] feature matching the matrix C's [•wh•] feature in the workspace), and therefore excluded by Last

---

[37] The data in (38) were checked with various native speakers, who unanimously declared them to be ill formed, and who all found a sharp contrast in the minimal pairs that can be formed on the basis of (36) and (38). However, I am aware of a few exceptions to the apparent general neglect of constructions like those in (38) in the literature: Such examples are discussed in Fiengo et al. (1988) and, following them, Fitzpatrick (2002), and judged grammatical; also see Fanselow (2004, 112). I have nothing to say here about the source of the diverging judgements, except for the observation that Fiengo et al. (1988) are primarily concerned with contrasting the construction in (38), with a *wh*-phrase embedded in a subject DP and an object *wh*-phrase ending up in front of it, with one in which the subject DP-internal *wh*-phrase undergoes movement (in violation of the CED) and the object *wh*-phrase stays in situ – and not with one in which a *wh*-phrase is embedded in an object DP and a subject *wh*-phrase undergoes movement. In other words: One might speculate that judgement differences arise in this domain because different kinds of minimal pairs are taken into account, and judgements are taken to be relative rather than absolute.

Resort. Consequently, any movement operation applying to DP_2 at later stages of the derivation fatally violates the PIC. In contrast, it is worth pointing out that the (G)MLC would not make the right predictions here. If the (G)MLC is defined in terms of asymmetric c-command, all sentences in (38) are ceteris paribus predicted to be well formed; if it is defined in terms of closeness (as in (1) of chapter 2), it will also wrongly permit *wh*-movement of DP_2 in (38), at least in those cases where DP_1 is deeply embedded.[38]

5. Some Refinements

The approach developed in this chapter imposes severe restrictions on *wh*-movement; as a matter of fact, it turns out to be slightly too restrictive in two domains.

5.1 *Multiple $C_{[\bullet wh \bullet]}$ Domains and Intervention*

The first problem concerns sentences like (39) in German (see Heck & Müller (2000; 2003)).

(39) [$_{DP}$ Die Frage   [$_{CP_6}$ wer$_1$ C t$_1$ was$_2$ mitbringt]]] ist relevant für die Frage
           the question   who   what brings   is relevant to the question
        [$_{CP_5}$ wie$_3$ C Fritz denkt [$_{CP_4}$ t$'_3$ dass die Party t$_3$ wird]
           how   Fritz thinks   that the party   will be

(39) is clumsy, but well formed. Consider the underlying numeration. There are two C heads bearing [$\bullet$wh$\bullet$] features (C_6 and C_5), and there are three *wh*-pronouns bearing [wh] features – hence, one of the two $C_{[\bullet wh \bullet]}$ heads will have to give rise to a multiple question. The example has been designed in such a way that CP_6 is the multiple question, and CP_5, which is merged earlier, is a simple question. The task now is to ensure that the *wh*-phrase *wie* ('how') can undertake steps of successive-cyclic movement until it reaches the edge of T of CP_5, where it is attracted by $C_{[\bullet wh \bullet]}$. Unfortunately, successive-cyclic movement of *wie*$_3$ turns out to be blocked at the very first stage under present assumptions: At the point where it must be decided whether an edge feature can be inserted for *wie*$_3$ movement, the phrase is wrongly predicted to be balanced: There are two C heads bearing [$\bullet$wh$\bullet$] features, and there are two remaining items in the workspace that bear corresponding [wh] features. Closer scrutiny reveals that the same kind of problem also shows up in simpler sentences in German (see (40-a)) and in English (see (40-b)).

---

[38] A question may arise with respect to sentences like (i) in English.

(i)   Who$_2$ did you talk to t$_2$ [$_{CP}$ after meeting who$_1$] ?

In order to derive well-formedness here, one might be led to postulating a low base-position for the adjunct clause (see Larson (1988)).

(40) a. $[_{CP_5}$ Wer$_1$ hat t$_1$ wen$_2$ gefragt $[_{CP_4}$ was$_3$ Fritz   t$_3$ mag]] ?
           who$_{nom}$ has   whom$_{acc}$ asked   what$_{acc}$ Fritz$_{nom}$ likes
      b. $[_{CP_5}$ Who$_1$ t$_1$ asked whom$_2$ $[_{CP_4}$ what$_3$ C John likes t$_3$]] ?

As it stands, the *wh*-phrase DP$_3$ (*was/what*) cannot reach the edge of T, from where
it can be attracted by C$_4$ bearing [●wh●] in accordance with the PIC. The highest
position that DP$_3$ can be in prior to *wh*-movement to SpecC is the edge of v in (40-a)
in German (due to this language's scrambling options), and the complement position
of V in English (due to this language's lack of scrambling options).

Intuitively, the problem with (39), (40-a), and (40-b) is clear: A *wh*-phrase that is
part of the workspace must not interact with a *wh*-phrase in a given derivation if the
two *wh*-phrases target different C$_{[●wh●]}$ domains. Following Heck & Müller (2000;
2003), this problem can be solved by minimally enriching the representation of *wh*-
features. Thus, suppose that both [●wh●] and [wh] features are accompanied by scope
indices in the numeration, and that *wh*-phrases can only be interpreted with a given C
node if they share a scope index. Under this assumption, a feature [wh]$_i$ on a *wh*-item
can never count as potentially available for a feature [●wh●]$_j$ on a C in the workspace,
due to feature mismatch. For the cases at hand, this means that the [wh] feature of DP$_2$
must be accompanied by the same scope index as the [●wh●] feature of C$_6$ in (39), and
of C$_5$ in (40-a) and (40-b), in order to be interpretable as part of the multiple question.
Therefore, at the point where the question of edge feature insertion for an intermediate
movement step of DP$_3$ must be decided, DP$_2$ does not intervene anymore: A feature
like [wh]$_5$ on DP$_2$ in (40-ab) can never lead to a balanced phase for a feature like
[●wh●]$_4$ on C$_4$. Only a feature like [wh]$_4$ on DP$_3$ can do so; accordingly, the EFC
permits feature insertion and thereby movement of DP$_3$, as required by the PIC.

Examples like those in (39) and (40) do not pose comparable problems in standard,
(G)MLC-based approaches to *wh*-intervention. Hence, at first sight it seems as though
the additional assumption motivated by these data (viz., scope indices that accompany
wh-features) might provide an argument against the present approach. However, such
a conclusion would be premature. Closer inspection reveals that the same kind of
problem shows up for (G)MLC-based approaches in slightly different constructions
like the one in (41), which were first investigated in detail in Baker (1970).[39]

(41) (*)Who$_1$ C$_{[●wh●]}$ t$_1$ wonders $[_{CP}$ what$_3$ C$_{[●wh●]}$ who$_2$ bought t$_3$] ?

As noted by Lasnik & Saito (1992), (41) is ill formed (on a par with examples that
instantiate a typical superiority effect, like (14-b)) only if *who$_2$* takes embedded scope;
but it is well formed if *who$_2$* takes matrix scope. This is a long-standing problem for

---

[39] Of course, strictly speaking, the [●wh●] features on the two C heads have already been discharged (and
deleted) in the output structure in (41); they are listed here only so as to simplify exposition.

(G)MLC-based approaches because it is by no means clear how it can be ensured that a constraint like the (G)MLC "should choose between two *wh*-phrases only when they are to wind up in the same Spec of CP at LF", as Lasnik & Saito (1992, 120) put it. In view of this problem, approaches that set out to reconcile the superiority effect with embedded scope of who_2 with the absence of such an effect with matrix scope of who_2 in cases like (41) either rely on the assumption that the constraint is not checked locally in the derivation, but globally, by taking into account the semantic interpretation of the whole sentence (see Kitahara (1993) and Grimshaw (1994) for suggestions of this kind); or they dispense with an (G)MLC-type constraint in toto (see Lasnik & Saito (1992), who suggest a constraint on *wh*-absorption that also presupposes global look-ahead).[40]

In contrast, as noted in Heck & Müller (2003, 127), the present approach in terms of scope indices on *wh*-features accounts for (41) straightforwardly, and in a strictly local manner: who_2 does not block movement of $what_3$ if the two *wh*-items have different scope indices, but who_2 blocks movement of $what_3$ if the indices are identical (because the VP phase is then automatically balanced, and an edge feature cannot be inserted). Thus, it turns out that the solution to what initially looks like a problem that is confined to the present approach actually provides the solution to a long-standing problem with standard approaches to *wh*-intervention.

5.2 D-Linking and Intervention

Wh-phrases that qualify as D(iscourse)-linked behave differently from other *wh*-phrases in a number of respects; see Pesetsky (1987) and much subsequent literature. One well-known pecularity of D-linked *wh*-phrases is that they do not induce intervention effects in English if they show up in situ; see the contrast in (42).[41]

(42) a. *I know [$_{DP_2}$ which books] who_1 read t_2
       b.   I know $what_2$ [$_{DP_1}$ which people] read t_2

The present analysis can accomodate standard accounts of this phenomenon. Thus, assume that D-linked *wh*-phrases in English can optionally lack a (proper) [wh]-feature that would make them accessible for a [●wh●] feature on C. Then, if a D-linked *wh*-phrase lacks the [wh] feature in the workspace, the only way to balance the phase is to insert an edge feature, thereby triggering movement of the remaining *wh*-phrase. Thus, an intervention effect can be avoided. (Of course, a [wh] feature must be present in those cases where the D-linked *wh*-phrase itself undergoes *wh*-movement.)

---

[40] But also cf. Fanselow (2004) for an account that tries to evade both these consequences.

[41] The situation is different in German, where D-linking does not seem to have such effects; see, e.g., the examples in (35).

6. Scope of the Account

6.1 (G)MLC vs. PIC

Recall that the (G)MLC (repeated here for convenience in (43)) emerges as a generalization of (i) the F-over-F Principle (i.e., the feature-based A-over-A Principle) which talks about *dominance* (see (60) of chapter 1), and (ii) the original MLC (essentially, the Superiority Condition) which talks about *c-command* (see (80) of chapter 1).

(43) *Generalized Minimal Link Condition*:
      In a structure $\alpha_{[\bullet F\bullet]}$... [... $\beta_{[F]}$ ... $\gamma_{[F]}$ ...] ..., movement to $[\bullet F\bullet]$ can only affect the category bearing the [F] feature that is closer to $[\bullet F\bullet]$.

In this chapter, I have considered a number of intervention effects which all center around the feature [wh] that is involved in *wh*-movement, and which involve either a c-command relation between two competing items, or no (relevant) structural relation at all, but not a dominance relation between two competing items (with the exception of the construction in section 3.5); and I have shown that given a strict version of the PIC (with all phrases as phases) and an independently motivated revision of Chomsky's Edge Feature Condition that regulates the insertion of edge features for intermediate movement steps on phase heads, the intervention effects follow without recourse to a constraint like the (G)MLC (which, furthermore, turned out to be inadequate in cases where intervention took place without c-command). In view of this, the question arises whether the approach can be generalized.

First, it should be clear that although the analysis presented here focusses on [wh] features, it is by no means construction-specific. Intervention constraints are predicted to occur with movement whenever there is a head in the numeration bearing a feature $[\bullet F\bullet]$; two categories could in principle be affected by movement because they are both marked [F]; and one of these two categories enters the derivation in a phase before the other one does (and cannot subsequently move across it because of another feature [F']). Thus, the approach is general in the sense that it is not confined to [wh] features.

Second, what about intervention effects that involve dominance (and that are covered by the F-over-F Principle part of the (G)MLC)? Here the issue is somewhat more involved. Consider again a relevant ungrammatical example (from subsection 3.2 of chapter 1) that involves scrambling from an infinitive followed by scrambling of the (now remnant) infinitive to a higher position (and thereby violates the Unambiguous Domination requirement); see (44-a) (remnant scrambling) vs. (44-b) (scrambling of a complete category).

(44) a. *dass $[_{\alpha_2}$ $[_{VP_{[\Sigma]}}$ t_1 zu lesen]] $[_{DP_{1,[\Sigma]}}$ das Buch] keiner t_2 versucht
          that   to read   the book$_{acc}$ no-one$_{nom}$ tried
          hat]
          has

b. dass [$_{\alpha_2}$ [$_{VP_{[\Sigma]}}$ das Buch$_1$ zu lesen]] keiner   t$_2$ versucht hat]
    that   the book$_{acc}$ to read   no-one$_{nom}$ tried   has

Recall that according to the (G)MLC approach, (44-a) is ungrammatical because the more inclusive category bearing [Σ] (viz., α_2) is closer to the target head that bears [$\bullet\Sigma\bullet$] than DP$_1$, which is also marked [Σ] (by assumption) but is contained in α_2; therefore, α_2 must move first, and subsequent movement of DP$_1$ to a lower position will violate the Strict Cycle Condition (see the next chapter on freezing effects that arise if DP$_1$ is moved to a higher position).

Before we try to determine whether the illformedness of the particular example in (44-a) can be derived under present assumptions, let us look at the predictions that the system makes for dominance-based intervention more generally.

6.2   Dominance-Based Intervention and the PIC-Based Approach

6.2.1   One Landing Site

Consider first the structure in (45), where α_2 and β_1 are equipped with a feature [F] that is required by a head ζ which has a unique movement-inducing feature [$\bullet$F$\bullet$]. Thus, two items that are in a dominance relation compete for movement to a single landing site.

(45) $\zeta_{[\bullet F\bullet]}$ ... [$_{\alpha_{2,[F]}}$ ... ([$_Z$] ... $\beta_{1,[F]}$ ... (])] ...

Two possibilities need to be considered. First, suppose that there is a category Z that dominates β_1 and is dominated by α_2. In this case, the PIC implies that β_1 can be extracted out of α_2 only if it undergoes an intermediate movement step to Specα_2 first (plus perhaps other intermediate movement steps if Z is a phase (i.e., an XP), or dominates a phase that in turn dominates β_1, etc.). Such an intermdiate movement step to Specα_2 could in principle be triggered in three ways. First, it could be that the head of α_2 has a movement-inducing feature [$\bullet$F$'\bullet$] that directly attracts β_1. Second, it could be that there is some other head that is outside of α_1 (i.e., still part of the numeration, at the relevant step of the derivation) that bears a feature [$\bullet$F$'\bullet$] (which is shared by β_1 but not by α_2) that indirectly attracts β_1 to Specα_2, via legitimate edge feature insertion. This corresponds to the situation encountered in section 3.5 above. (Recall that a *wh*-PP may undergo scrambling out of a DP, and may then legitimately move on to satisfy the [$\bullet$wh$\bullet$] feature of an interrogative C, as in (30-a).) However, the third a priori conceivable possibility of effecting intermediate movement of β_1 to Specα_2 in (45) in the presence of Z is predicted to be impossible under present assumptions: Given the EFC, it is not possible to insert an edge feature on the head of α_2 so as to balance the α_2 phase for [$\bullet$F$\bullet$] on ζ in the numeration: α_2 is balanced because it itself makes the [F] feature potentially available for [$\bullet$F$\bullet$] on ζ. Thus, it follows that intermediate movement of β_1 cannot be indirectly triggered by [$\bullet$F$\bullet$] on ζ

in (45). Consequently, if there is no other (direct or indirect) trigger, extraction of β_1 from α_2 is predicted to be impossible.

All this only holds for the case that there is a category Z that dominates β_1 and is dominated by α_2. The second possibility that needs to be considered for (45) is that there is no such Z; i.e., β_1 is a specifier of α_2 to begin with. In that case, extraction of β_1 from α_2 to $\zeta_{[\bullet F \bullet]}$ should be possible. Moreover, under both basic scenarios, movement of α_2 to $\zeta_{[\bullet F \bullet]}$ should be unproblematic. Thus, the predictions for (45) are clear enough. Are they also supported by empirical evidence?

Basically, the prediction should be that since German has scrambling and English does not, English examples that are analogous to (30-a) in German should ceteris paribus qualify as ill formed. Indeed, it appears that native speakers can detect a contrast between (46-a), (47-a) (which involves movement of the more inclusive *wh*-phrase and may sound clumsy, but is generally not regarded as ungrammatical) and (46-c), (47-c) (which involves only one *wh*-phrase), on the one hand, and (46-b), (47-b) (which involves two *wh*-phrases, with the lower *wh*-phrase moved out of the DP that acts as the higher *wh*-phrase, resulting in deviance), on the other hand.

(46) a. $[_{DP_1}$ Which picture of $[_{DP_2}$ who(m) $]]$ do you like t_1 ?
      b. ??$[_{DP_2}$ Who $]$ do you like $[_{DP_1}$ which picture of t_2 $]$?
      c. $[_{DP_2}$ Who $]$ do you like $[_{DP_1}$ a picture of $]$?

(47) a. $[_{DP_1}$ What kind of book about $[_{DP_2}$ who(m) $]]$ did you read t_1?
      b. ??$[_{DP_2}$ Who $]$ did you read $[_{DP_1}$ what kind of book about t_2 $]$?
      c. $[_{DP_2}$ Who $]$ did you read $[_{DP_1}$ a book about t_2 $]$?

Still, it is unclear whether the contrast between, say, (47-b) and (47-c) can ultimately be viewed to be clear enough to directly support the idea of a dominance-based *wh*-intervention effect. It might be that the deviance of the (b)-examples can be traced back to some independent factor, and that the derivations that generate the sentences in question actually can proceed in a way that respects the PIC.[42]

---

[42] To mention just one possibility: It might be that intermediate extraposition might feed *wh*-movement of DP_2; see Heck (2008) for a discussion of this general possibility. This hypothesis might be further substantiated by the relative well-formedness of data such as (i), which may involve PP extraposition.

(i) $[_{DP_1}$ Which picture t_2 $]$ do you like $[_{DP_2}$ of whom $]$?

Also worth mentioning in this context are sentences of the general type in (ii) (also compare the reverse construction in (64) of chapter 1).

(ii) ??/*$[_{DP_2}$ Who $]$ do you wonder $[_{DP_1}$ which picture of t_2 $]$ Mary bought t_2 ?

Examples of this type are sometimes viewed as not being fully ill formed (see, e.g, Lasnik & Saito (1992)). Under present assumptions, they violate the prohibition against freezing that is reducible to the CED (see chapter 4). In addition, we expect (ii) to violate the *Wh*-Island Condition (or whatever derives it; see chapter

As for the prediction that β_1 should be able to extract out of α_2 in (45) if it is a specifier, there are again intervening factors that blur the simple picture arising under the assumptions made in the present chapter. One is that extractions of specifiers typically give rise to violations of the Left Branch Condition (at least in languages that are like English in that they respect the Left Branch Condition); and even in languages that seem to freely violate the Left Branch Condition with extraction from DP (like Russian), the situation is far from simple because of the existence of various intervening factors (like, e.g., constraints on the possible shape of *wh*-DPs and the availability of scrambling).

Finding evidence for or against intervention effects in a structure like (45) in other domains (than *wh*-movement) is no trivial task either. In many cases of optional movement (topicalization, focus movement, scrambling), there is no inherent (e.g., morphological) indication that a category is subject to the movement operation; one can only see that it has undergone the movement when it shows up in the designated position (under the circumstances that are typical of the respective movement type). Furthermore, with semantically oriented movement-inducing features, it may not be clear that the recursive A-over-A structure is legitimate in the first place. For instance, suppose that a language has focus movement that is driven by a feature [foc]. With respect to (45), this would imply a [foc]-over-[foc] structure that is legitimate, and semantically interpretable, to begin with. Relevant evidence may exist in languages with focus movement (it may be related to phenomena like recursive givenness marking or second occurrence focus; see Wagner (2008)), but it is not entirely straightforward. I will have to leave this issue for further research. For the time being, we can note that there may well be empirical evidence for dominance-related intervention effects with single movement to a unique landing site (as predicted under the (G)MLC); and if there is, the present PIC-based analysis can derive most of the kinds of effects that may occur.[43]

---

5). However, note that (ii) is not predicted to be excluded by the PIC as an instance of (45), on a par with (46-b) and (47-b) (with the qualifications just made); (ii) differs from these examples in that there are two interrogative C heads, not just one; see the following subsection.

[43] Note in passing that the partial focus movement construction discussed by Fanselow & Lenertová (2011) for German and Czech, and by Hartmann & Zimmermann (2007) for Hausa does not provide a counterexample to the expected dominance-based intervention effect. Relevant data (from German and Hausa, respectively) are given in (i) and (ii).

(i)   a.   Was hat Fritz gemacht?
             'What has Fritz done'?
       b.   [$_{DP_1}$ Ein Buch] hat Fritz [$_{VP}$ t$_1$ gelesen]
                   a   book   has Fritz   read

(ii)   a.   Mèe ya fàaru?

6.2.2 Two Landing Sites

Returning to the effect visible in (44-a), we can note that the situation shown in (45) is not exactly the same as the one underlying this construction. Rather, the basic structure looks as in (48), where again α_2 and β_1 both have a feature that is sought by some external head ζ, but this time, ζ is equipped with *two* $[\bullet F\bullet]$ features, which effects multiple movement of the same type.[44]

(48) $\zeta_{[\bullet F\bullet],[\bullet F\bullet]}$ ... $[_{\alpha_{2,[F]}}$ ... $([_{Z}$ ... $\beta_{1,[F]}$ ... $(])]$ ...

Irrespective of how deeply embedded β_1 is within α_2, there is now nothing (according to present assumptions) that would preclude local edge feature insertion for β_1 on all heads that might intervene between β_1 and α_2: There are two movement-inducing features in the numeration, and consequently, a phase can only be balanced if there are potentially available matching features. Suppose, e.g., that β_1 has been merged in a separate phase Z; then the head of Z can be assigned an edge feature, and it can attract β_1 to its edge domain (with α_2 in the numeration providing one potentially available matching feature for ζ, and the moved β_1 the other one that is needed to get a balanced phase). Similarly, when the head of α_2 is merged with a category containing β_1 (in its edge domain, given the PIC), the head of α_2 can also be assigned an edge feature for attraction of β to its specifier because this is the only way to balance the α_2 phase. Thus, derivations based on a pattern like (48), where β_1 first moves up to α_2's edge domain, then extracts out of α_2, and, finally, α_2 undergoes remnant movement of the same type, are in fact not blocked under present assumptions.

As for the concrete example in (44-a), the same consequence arises: The ill-formedness of (44-a) does not follow from the PIC. Again, suppose that there are two $[\bullet \Sigma \bullet]$ features, one for DP_1, one for α_2. Then, there should be a well-formed derivation for (44-a), with DP_1 undergoing edge-feature driven movement to SpecV

       'What happened'?

b.   <u>ɓàràayii</u> nèe su-kà   yi mîn saatàa
     robbers PRT 3.PL-REL.PERF do to me theft
     'ROBBERS have stolen from me!'

As the contexts in (i-a) and (ii-a) make clear, the focussed constituents are VP in (i-b) and vP in (ii-b). If these categories bear a feature [foc] that is required by a $C_{[\bullet foc\bullet]}$ triggering focus movement, and the moved items also bear the feature [foc], this would be an unexpected exemption from the non-intervention requirement embodied both in the (G)MLC-based approach and the PIC-based approach. However, in these cases, we are not dealing with a regular case of focus movement: The fronted items are *parts* of focussed phrases, but they are not focussed phrases themselves. Accordingly, Fanselow & Lenertová (2011) argue that the feature that triggers movement operations like those in (i-b) and (ii-b) is not a feature related to focus (but some more general, semantically neutral feature similar to the one that is involved in scrambling).

[44] Alternatively, one might assume that there are two separate heads that each have one $[\bullet F\bullet]$ feature. This would not make a difference in the present context.

first, followed by regular feature-driven movement of DP_1 to the edge domain of the head that bears the $[\bullet\Sigma\bullet]$ features (which discharges one of them), and then of α_2 to the edge domain of the same head (which discharges the second movement-inducing feature). In such a derivation, all instances of edge feature insertion respect the Edge Feature Condition, and all instances of movement are compatible with the PIC.

More generally, we can conclude at this point that whereas the PIC-based approach that dispenses with the (G)MLC can derive a number of (G)MLC effects (as well as other intervention effects), it is not fully equivalent to the (G)MLC because it covers only a subset of its domination-related part: In cases where the two competing dominance-related categories (bearing [F]) correspond to only *one* movement-inducing feature $[\bullet F\bullet]$ on a head in the numeration, the Edge Feature Condition blocks edge feature insertion (hence, movement) of the more inclusive category and thus correctly derives an intervention effect. However, in cases where the two competing dominance-related categories (bearing [F]) correspond to *two* movement-inducing features $[\bullet F\bullet]$ on a head in the numeration, the Edge Feature Condition does not block edge feature insertion (hence, movement) of the more inclusive category; therefore, the intervention effect that also shows up in this configuration cannot yet be derived.

The problem can be fixed by postulating what one might call a "Residual F-over-F Condition" (RFOF), as in (49).

(49)   *Residual F-over-F Condition* (RFOF):
       If $[\bullet F\bullet]$ can be checked either with a head, or with an edge element, it must be
       checked with the head.

This forces α_2 movement to apply first in (48), even if β_1 has reached α_2's edge domain via legitimate edge feature insertion and could in principle extract from α_2 in accordance with the PIC. Subsequent β_1 lowering in Unambiguous Domination contexts then violates (at least) the Strict Cycle Condition.[45]

(49) is arguably conceptually superior to the (G)MLC, and it certainly demands much less search space than the (G)MLC. Still, one might hope that one could eventually do without this constraint. For the time being, I will leave it at that. However, I will come back to this issue in the next chapter, where I will argue that the minimal residual stipulation in (49) can be dispensed with (given an independently motivated further restriction on the insertion of edge features).

---

[45] Note incidentally that the distinction between head and edge element in the definition of (49) is the only case where a minimal structural differentiation of the bundle of categories accessible for further operations seems necessary; recall the discussion in subsection 2.3.2 of chapter 2.

7.   Conclusion

7.1   Main Results

Let me summarize the main results of this chapter. I have argued that typical (G)MLC effects can be derived from the PIC, given an independently motivated modification of Chomsky's (2001b) Edge Feature Condition. This implies that the (G)MLC can be dispensed with (except, so far, for a residue, see (49)). In view of the arguments against the (G)MLC that were presented in chapter 2, I take this to be a welcome result. The analysis derives both superiority effects in English-type languages, and the general lack of superiority with clause-bound *wh*-movement in scrambling languages like German. Furthermore, it is correctly predicted that under those circumstances where scrambling is independently excluded (long-distance effects, subject position effects, remnant movement effects), superiority effects show up again in German. Finally, unlike the (G)MLC, the system based on the revised EFC and the PIC predicts superiority-like intervention effects without c-command, which are indeed attested.

7.2   Core Assumptions Reconsidered

At this point, it may be worthwile to reconsider the assumptions that are needed to derive (G)MLC effects from the PIC in the present analysis. They are repeated in (50).

(50)   a.   All syntactic operations are driven by features of lexical items.
　　　 b.   Edge features that trigger intermediate movement steps can be added only
　　　　　 if that is the only way to produce a balanced phase.
　　　 c.   All phrases are phases.

With these assumptions in place, (G)MLC effects can be derived from the PIC. Still, it turns out that weaking one or more of the assumptions, or, in fact, abandoning them in toto, does not necessarily mean that no (G)MLC effects whatsoever can be derived from the PIC anymore. Let me go through the three assumptions in turn, beginning with (50-a).

   In the present context, (50-a) mainly serves the purpose of rendering all syntactic movement operations costly (in chapter 4, an additional justification for this assumption will be provided): Movement can only take place if it is triggered by designated structure-building features ([•F•]), via the Feature Condition. Accordingly, since inherent features cannot plausibly be assumed to be present to induce intermediate movement steps on lexical items, these features must be inserted on the item in the course of the derivation (in minimal violation of Inclusiveness). The question arises of whether an alternative analysis might also be possible that captures the gist of the present approach but does without the assumption that all syntactic operations are feature-driven. That is the case. The PIC-based approach to (G)MLC effects could

also be executed in an approach in which, e.g., movement is triggered by a class of *Criterion* constraints, like the *Wh*-Criterion (see Lasnik & Saito (1992), Rizzi (1996), among many others) for *wh*-movement, or (one clause of) the Extended Projection Principle ("Clauses have subjects"; see Chomsky (1982b)) for movement to SpecT. The only crucial assumption would be that intermediate movement steps must also be forced, and this has to be done in a way that is independent of the PIC.

As regards (50-b), as it stands the analysis rests on a modification of the EFC that replaces Chomsky's original requirement of "having an effect on outcome" with the more principled, locally verifiable requirement of being the only way to bring about phase balance. Two questions need to be answered: First, is the modification needed? And second, is the EFC itself needed to derive (G)MLC effects from the PIC? (Alternatively, the phase balance requirement could perhaps also be implemented without recourse to edge feature insertion.)

The answer to the first question is yes, for the simple reason that "to have an effect on outcome" is too vague a condition to be unambiguously verifiable at each step of the derivation – in fact, as noted above, the modification in terms of balanced phases envisaged here can to a large extent be viewed as nothing more than a more precise version of the "effect on outcome" requirement. Such a condition is needed to determine whether an intermediate movement step is legitimate or not; and if it is not, the non-moved item will invariably be too far removed from the next phase edge to be accessible for further operations, which eventually accounts for the occurrence of (G)MLC effects.

Turning to the second question, the answer is no. The EFC is the locus that suggests itself for integrating the phase balance requirement if assumption (50-a) is made, and all movement must be triggered by appropriate features. However, as shown in Heck & Müller (2000; 2003) and Müller (2004a), it is in principle also possible to formulate a constraint Phase Balance in its own right that forces or prohibits intermediate movement steps in a system that does not rely on edge features as triggers for successive-cyclic movement.

Finally, consider the third assumption that was employed in the derivation of (G)MLC effects, viz., (50-c). For the reasons laid out in subsection 2.3, I think that assuming all phrases to qualify as phases is both conceptually attractive (since search space is reduced in derivations, and the goal of maximizing efficient computation is furthered) and empirically justifiable (e.g., the assumption is compatible with evidence from morphological reflexes of successive-cyclicity, reconstruction, and binding theory). However, suppose that we were to postulate that only vP and CP qualify as phases, as is standardly assumed (see Chomsky (2000; 2001b; 2007; 2008)). Would this mean that no (G)MLC effects could be derived anymore? It would not. A significant subset of the (G)MLC effects addressed in the present chapter would still follow (others would not, though, and this can be taken to support the more restrictive notion

of phase adopted here). Let me address the effects one by one.

(i) The superiority effect involving subject and object in English (see (14)) would still be derivable from the PIC if only vP and CP were phases: Given that external arguments are externally merged in Specv, and internal arguments in SpecV, the former have an inherent advantage for further *wh*-movement because the vP phase is balanced (for a C$_{[\bullet wh \bullet]}$ in the numeration) with a *wh*-subject in its base position; object fronting to Specv is not required, and therefore excluded.

(ii) The biclausal superiority effect involving object and object in English (see (15)) would also still follow because the two object *wh*-phrases are not externally merged as phase-mates: When the lower *wh*-object is merged in the embedded infinitive, the infinitive's vP is balanced without edge feature insertion and object movement (for C$_{[\bullet wh \bullet]}$ in the numeration, there is a potentially available feature [wh] on the second *wh*-phrase, which is part of the workspace at this point).

(iii) The superiority effects involving object and object in English double object constructions (see (17-c) and (18-b)) do not follow anymore if only vP and CP are phases. The reason is that the crucial domain in which the decision on whether intermediate movement is possible must be taken is either PP (in (17-c)) or VP (in (18-b)). If it is the minimally dominating vP that acts as the first domain that needs to be balanced, movement of the lower *wh*-phrase would be predicted to be just as well possible as movement of the higher *wh*-phrase.

(iv) The superiority effect with P stranding in PP/PP configurations in English (see (20)) is also not derivable from the PIC if only vP and CP are assumed to be phases. The reason is the same as before: The domain in which the decision must be taken (whether intermediate movement is possible or not) is the minimally dominating PP; this category is always balanced in PP/PP structures containing two *wh*-phrases. However, on the vP level, both PPs are part of the structure, and either *wh*-item could move to Specv, in accordance with the EFC.

(v) The absence of superiority effects in German data like those in (21-b) and (22-b) would still be predicted under a vP/CP approach to phases; this does not come as a surprise because such an approach is more liberal to begin with.

(vi) The cases of long-distance superiority effects in German (see (24-b)) could still be derived because the two *wh*-items that are involved are not phase-mates in their base positions.

(vii) The superiority effect with subject raising in German (see (25-b), (26-b)) would be unexplained. Here, the decision about the legitimacy of the intermediate movement step must be taken on the TP level (given that *wh*-scrambling can independently move an object to the edge of vP); on the CP level (i.e., the minimal phase for SpecT if only vP and CP are phases), subject and object movement cannot be distinguished (given that there is no (G)MLC).

(viii) The superiority effect with scrambling from *wh*-XP in German (see (30-b)) does not follow under the assumption that only vP and CP are phases: The decision must be made on the VP level.

(ix) The long-distance intervention effects without c-command in German (see the data in (32), (33), and (34)) would still be accountable for with vP and CP as the only phases.[46]

(x) Finally, the intevention effects without c-command in English in (38-a)–(38-d) do not follow if only vP and CP are phases: Movement of the *wh*-object to Specv would be legitimate here because the insertion of an edge feature on v is the only way to produce a balanced vP phase (the other *wh*-phrase that is embedded in the subject DP in Specv is not potentially available on the vP level); recall that the analysis given on page 146 relies on the decision about the legitimacy of edge feature insertion being taken on the VP level (where the external argument DP is not yet part of the structure, and a VP phase counts as balanced without edge feature insertion and *wh*-object movement to SpecV). In contrast, the intervention effect in (38-e) *does* follow under the vP/CP approach: The *wh*-object is included in an embedded CP, and due to the presence of the other *wh*-phrase in the workspace, it can never make the first step to Specv (let alone the second one to the embedded SpecC).

In sum, several (G)MLC effects could still be derived if the standard assumption were made that only vP and CP are phases. However, certain other effects would not follow anymore, viz., those where the evidence requires that the decision about edge feature insertion must be made in a non-vP, non-CP domain (e.g., on the PP, VP, or TP levels). In a nutshell, an intervention effect cannot be derived if the two competing items are part of the same phase, and neither occupies the edge domain of the phase for independent reasons – and the fewer phases there are, the more occurrences of phase-mateness there will be. Hence, the assumption that all phrases are phases permits an empirically superior analysis of intervention effects.

7.3   Some Consequences

Needless to say, given the assumption that all phrases are phases, the PIC has important consequences for many other phenomena outside the domain of intervention effects with movement, especially if we we adopt the following, even stronger, hypothesis, which the approach assumed here would seem to lend itself to naturally (given some obvious reformulation of (5)):

(51)   Once rendered inaccessible by the PIC, syntactic structure does not become accessible again when the syntactic derivation terminates ("at LF"). Hence, there

---

[46] The (presumed) absence of an intervention effect in (35-b) would also follow: The *wh*-object can reach the vP domain by scrambling.

can be no constraints on representations ("bare output conditions").

Let me briefly (and tentatively) explore some of the consequences of (51). First, (51) implies that there is no reason left to assume the existence of traces (neither as t, nor as a copy). This might then suggest a derivational, cyclic approach to semantic interpretation (see, e.g., the sketches in Sternefeld (1996b), Adger & Svenonius (2003), and Stechow (2005)). More generally, it is clear that if (51) is taken seriously, virtually all instances of semantic binding of variables (traces or pronouns) raise significant problems. Arguably, a variable-free semantics (see Jacobson (1999), also Büring (2005)) would seem to be an ideal modelling of the semantics component to go along with the kind of strictly derivational syntax envisaged here, with extremely small accessible parts of the derivation left.[47]

Furthermore, (51) also raises interesting problems for (the syntactic) *binding of reflexives and reciprocals* (at least for those cases that are not strictly local and can be covered by Reinhart & Reuland's (1993) reflexivity constraints) and *personal pronouns*, for *control*, for *long-distance agreement*, etc. It seems that apparent non-local relations must be accounted for by successive-cyclic local [F] feature movement from head to head (mediated by concepts like balanced phases or motivated by independent features; see Pesetsky (2000) on the viability of feature movement). [F] must encode the relevant properties of the in-situ element; e.g.: anaphor, PRO. For (syntactic) *binding*, this strategy would be a natural extension of proposals like Chomsky's (1986b) LF movement of anaphors, and a version of it is in fact pursued by Reuland (2001) in his account of A-chain condition effects, and by Fischer (2004) more generally. For (obligatory) *control*, the strategy would amount to a decomposition of Landau's (2000) Agree relation into small steps of feature movement (or, indeed, a version of Hornstein's (2001) A-movement approach).

Finally, *long-distance agreement* may or may not suggest the same kind of analysis; here there is substantial disagreement in the literature. First, it has been argued that the phenomenon of long-distance agreement is indeed real; but the operation (unlike other syntactic operations) may violate standard locality restrictions like the PIC (see Stjepanović & Takahashi (2001), Sells (2006), and Bošković (2007)). Second, it has been claimed that long-distance agreement may only affect two items if they are fairly

---

[47] All this presupposes that products of syntactic derivations are subject to semantic interpretation. A radical alternative that has more recently been pursued in some detail by Unger (2010) is based on the assumption that there is actually no relevant interaction between syntactic displacement and semantic interpretation (including phenomena like relative scope) at all. Against the background of a minimalist, local derivational approach to syntactic displacement (partly based on earlier work by Broszewski (2003)), Unger (2010) shows that a substantial set of data involving movement and scope-taking of *wh*-items (and other quantified items) can be given a coherent and satisfactory account where syntax and semantics are (almost) completely independent from one another.

local (phase-mates) after all (see Boeckx (2004), Bhatt (2005), Richards (2011b));
however, this may require a somewhat broader concept of what a phase is. Third,
it has been proposed that long-distance agreement involves (possibly covert) move-
ment into the higher local domain (phase) (see Polinsky & Potsdam (2001), Polinsky
(2003), and Chandra (2005)). Finally, Butt (1995), Chomsky (2001b), Legate (2005)
and Keine (2008) argue that long-distance agreement involves cyclic Agree. Except
for the second analytic option, all these approaches are in principle compatible with
the assumption that all phrases are phases. However, only the last two possibilities are
compatible with the more far-reaching hypothesis in (51). The approach to interven-
tion effects developed in this chapter will undoubtedly have further repercussions on
the analysis of (local and long-distance) Agree relations. However, since the present
monograph focusses on overt displacement operations, I will leave this issue for future
work.[48]

   The last thing I would like to do in this chapter is to emphasize the importance of a
strictly derivational framework for the enterprise of accounting for (G)MLC effects in
terms of the PIC. In particular, the assumption of a total order of syntactic operations
in a derivation is indispensable. If the hypothesis that syntactic operations are strictly
ordered is given up, and a partial ordering is permitted (e.g., by assuming simultaneous
application of operations at the phase level, as envisaged in Chomsky (2007; 2008)),
this not only amplifies the conceptual problems with respect to weak or strong residual
representationality pointed out by Brody (2002) (and discussed in subsection 2.3.2 of
chapter 2 above); it also makes a systematic reduction of (G)MLC effects to the PIC
impossible: Whether an item is or is not yet part of the syntactic structure may have
important consequences: The order in which two competing (e.g., *wh-*) items enter
the structure is crucial for the analysis.

   With (G)MLC effects accounted for by the PIC, I will now turn to the second main
task of the present monograph: that of deriving CED effects from the PIC.

---

[48] The issue will come up again briefly in subsection 3.2.3 of chapter 4, though.

Chapter 4

On Deriving CED Effects from the PIC

1. Introduction

The goal of this chapter is to answer the question of how the effects of the Condition on Extraction Domain (CED; Huang (1982), Chomsky (1986a; 1995; 2008), Cinque (1990), Manzini (1992)) can be made to follow under minimalist assumptions, according to which concepts like government, L-marking, and barrier are not available and specific constraints on movement (like the CED) are to be abandoned if possible.[1] A simple version of the CED that is based on the notion of barrier is given in (1) (see chapters 1 and 2).

(1) *Condition on Extraction Domain* (CED):
    a.   Movement must not cross a barrier.
    b.   An XP is a barrier iff it is not a complement.

I have argued in chapter 2 that none of the existing attempts to derive CED effects from more basic assumptions in a minimalist, derivational framework can be viewed as fully successful. In this chapter, I will argue that CED effects follow from the Phase Impenetrability Condition (PIC; see Chomsky (2000; 2001b; 2008)) which, as noted before, is first and foremost motivated as a constraint that minimizes search space in derivations. The version of the PIC that I adopt throughout this monograph is repeated in (2).

(2) *Phase Impenetrability Condition* (PIC):
    The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

---

[1] The first five sections of the present chapter correspond to Müller (2010b), with various minor deletions, additions, and modifications.

For present purposes, we can assume that an item is in the edge domain of a head X iff it is a (possibly non-unique) specifier of X, where a specifier is a non-first merged (i.e., non-complement), non-head item.[2] The PIC forces successive-cyclic movement via phase edges. Accordingly, it must be ensured that the intermediate movement steps that are required by the PIC are indeed permitted. Given a general Last Resort principle according to which all syntactic operations are feature-driven (see (6) of chapter 3), movement to intermediate phase edges must be triggered by appropriate features.[3] Edge features are not obligatorily present on phase heads; this would create a crash if the number of items that need to be attracted to a phase edge in the course of successive-cyclic movement is different from the number of movement-triggering features on the phase head. The Edge Feature Condition (EFC) guarantees that edge features for intermediate movement steps can only show up on a phase head if they are needed.

(3) *Edge Feature Condition* (EFC):
     The head X of phase XP may be assigned an edge feature after the phase XP is otherwise complete, but only if that has an effect on outcome.

Given (3), phase heads can be assigned additional (i.e., non-inherent) edge features in the course of the derivation "if that has an effect on outcome", i.e., if that serves to implement intermediate movement steps required by the PIC and, indirectly, the Feature Condition (see (5) in chapter 3). In the preceding chapter, I have argued that replacing the informal requirement of "having an effect on outcome" with the more specific condition that edge features can only be inserted if that is the only way to balance a phase is the key to an account of (G)MLC effects in terms of the PIC. Thus, the version of the Edge Feature Condition that we have now is the one in (4) (= (12) in chapter 3).

(4) *Edge Feature Condition* (EFC; revised):
     The head X of phase XP may be assigned an edge feature after the phase XP is otherwise complete, but only if that is the only way to produce a balanced phase.

The crucial assumption that I would like to put forward in this chapter is that a minimal change in (4) also derives a version of the CED (in interaction with certain other, independently justified assumptions, some of which have already been introduced in

---

[2] See (3) in chapter 3 for a slightly more elaborate definition of phase edge that is fully compatible with the approach outlined below but not actually needed in the present context.

[3] As noted before, throughout this book, I assume that intermediate movement steps involve genuine movement operations rather than counter-cyclic locality-driven trace insertion, as proposed in Chomsky (1993), Takahashi (1994), and Boeckx (2003).

the preceding chapter). In a nutshell, I suggest to replace *after* with *before* in (4). More specifically, the claim is that CED effects can be derived from the PIC if the following four assumptions are made: First, all syntactic operations are driven by features of lexical items. Second, these features are ordered on lexical items. Third, all phrases are phases. And, most importantly, fourth, edge features that trigger intermediate movement steps can be added only as long as the phase head is still active. The version of the CED that can be derived under these assumptions differs from the one given in (1) in a subtle but interesting respect: It is not simply non-complement XPs that are barriers; rather, those XPs are barriers that come into being as a result of the very last operation taking place within the phase. See (5).

(5) *Condition on Extraction Domain* (new version; to be derived from the PIC):

    a.   Movement must not cross a barrier.

    b. α is a barrier if the operation that has merged α in a phase Γ is the final operation in Γ.

I proceed as follows. In section 2, I list and discuss the four assumptions that are needed to derive (5).[4] In section 3, I show how CED effects can be derived from the PIC under the assumptions made in section 2. In section 4, the analysis is extended to cover *freezing* effects; in this section, it is also shown that the Residual F-over-F Condition (RFOF) (see (49) of chapter 3) that was designed to account for some (G)MLC effects that could not be derived from the PIC under a phase balance-based approach to edge feature insertion can be dispensed with under the approach to freezing adopted here. As already noted in chapter 2, one of the most important characteristics of the new approach to CED effects, which is not shared by any of the existing approaches, is that it predicts the existence of the phenomenon of *melting*. In section 5, I show that melting effects systematically occur in German and Czech: Local scrambling in front of what would otherwise qualify as a last-merged specifier renders the specifier transparent for extraction. Section 6 contains an outlook and a conclusion. Finally, there is an appendix to the present chapter in which the analysis is tentatively augmented with a new approach to reflexes of successive-cyclic movement (which turn out to be no "reflexes" at all on the analysis proposed there).

---

[4] Just as we have seen in the case of deriving (G)MLC effects from the PIC (see subsection 7.2 of chapter 3), it turns out that the four assumptions are not equally important. In fact, only the last assumption, concerning the timing of edge feature insertion on phase heads, is really indispensable. The first three assumptions could in principle be replaced by alternative (for instance, weaker) statements, which would still yield some version of (5). I will address this issue in the conclusion of the present chapter. For the time being, I treat the four assumptions as equally important – on the one hand, because they help to derive what seems to me to be the empirically most adequate version of the CED; and on the other hand because they ensure that the overall reasoning is maximally perspicuous and explicit.

2. Assumptions

2.1   All Syntactic Operations are Feature-Driven

I have already introduced and motivated this first assumption in section 2.2 of chapter 3: A general Last Resort condition demands that a syntactic operation must either discharge a structure-building (i.e., subcategorization or movement-inducing edge) feature [•F•] or a probe feature [∗F∗] (by carrying out Agree); the constraint is repeated in (6).

(6) *Last Resort* (LR):
     Every syntactic operation must discharge either [•F•] or [∗F∗].

The second important assumption presupposes the first one; but unlike the first one (with the proviso made towards the end of the preceding chapter, in subsection 7.2), it has not been required for an account of (G)MLC effects.

2.2   Operation-Inducing Features on Lexical Items are Ordered

Consider the question of how linking can be brought about in current derivational syntactic theories, i.e., how the order of Θ-roles in argument structures of predicates in the lexicon is mirrored by argument realization in syntax.[5] For examples like (7), the task is to derive that *John*$_1$ is interpreted as AGENT, *a book*$_2$ as THEME, and *Mary*$_3$ as GOAL.

(7) John$_1$ gave a book$_2$ to Mary$_3$

One possible strategy is to assume that syntactic positions are provided with fixed argument structure correlates, such that, e.g., a DP daughter of vP is interpreted as AGENT, a DP daughter of VP is interpreted as THEME, and a PP daughter of V′ is interpreted as GOAL; such an approach is adopted in Adger (2003, 139).[6] On this view, an example like (8) is classified as ungrammatical because it emerges as uninterpretable: The VP-internal arguments cannot be associated with Θ-roles.

(8) *[$_{vP}$ John   gave+v [$_{VP}$ to Mary t$_V$ a book]]
        AGENT   ?   ?

---

[5] I am assuming here a standard approach to argument structure that relies on ordered lexical entries which predict a hierarchy of Θ-roles (see, e.g., Bierwisch (1988), Wunderlich (1997a), and Heim & Kratzer (1998)). Things would be different under radical approaches like the one developed in Borer (2004), where argument structures are not basic, but are read off syntactic representations.

[6] Of course, assignments of thematic roles to positions would be somewhat different under the approach to ditransitive constructions in (19) in chapter 3.

However, I take it to be far from clear that this is the right approach to take: A sentence like (8) can perfectly well be understood by speakers of English – it just happens to violate formal rules of grammar. I would therefore like to adopt another approach to linking in minimalist grammars, one that has its roots in categorial grammar, and a long tradition in transformational and phrase structure grammar (see Lewis (1972), Pollard & Sag (1994), Wunderlich (1997a), Lechner (2004), Kobele (2006; 2011), Sternefeld (2006), and Pesetsky & Torrego (2006), among many others). According to this approach, Θ-roles are ordered in lexical entries of predicates, and they are mapped onto a list of categorial subcategorization features (i.e., structure-building features: [•F•]) in reverse order, as schematically depicted in (9). There must then be a condition ensuring that subcategorization features are discharged one after the other, beginning at the top of the hierarchy (see below for a specific suggestion).

(9) a. *Θ-roles*:
         $\Theta_1 \gg \Theta_2 \gg \Theta_3$ (AGENT $\gg$ THEME $\gg$ GOAL)
     b. *Subcategorization features*:
         $[\bullet P \bullet]_3 \succ [\bullet D \bullet]_2 \succ [\bullet D \bullet]_1$

Furthermore, I assume that there can also be subcategorization features that do not go back to Θ-grids. These kinds of subcategorization features play a role in structure-building operations involving functional categories, and they also enter argument lists of lexical items. For example, VP is not Θ-marked by v, but v has a subcategorization feature [•V•] that permits Merge of v and VP; and similar reasonings apply in the case of Merge (T,vP) and Merge (C,TP).[7] At this point, the question arises of how an external argument DP comes to be base-generated as the specifier of v. Based on Larson (1988) and subsequent work, two options have been pursued. One possibility (adopted in approaches that correlate v with causation or some related semantic concept; see Hale & Keyser (2002), Harley (1995), Kratzer (1996), Adger (2003), Ramchand (2003), Schäfer (2007), Hole (2008), and references cited in these works) is that AGENT DPs are introduced by v. In the present approach, this would amount to a feature list [•V•] $\succ$ [•D•] on v, with [•V•] discharged by Merge (v,VP) before [•D•] is discharged by Merge (v′,DP). On this view, transitive verbs take only one argument; and ditransitive verbs subcategorize for two arguments (rather than three arguments, as in (9-b)). Alternatively, v is motivated solely by considerations pertaining to phrase structure geometry; external arguments remain true arguments of V,

---

[7] Also see Adger (2003) for discussion of this issue, with a different conclusion. – Strictly speaking, assuming, say, Merge of v and VP to proceed by discharge of a subcategorization feature on v was already required by the Last Resort condition in chapter 3. However, the issue was not really relevant in the context of the discussion of (G)MLC effects in chapter 3; it will become relevant now, though.

even though they are merged in the specifier of v (see Haider (2000a; 2006) for an analysis along these lines, and Georgi & Müller (2010) for an implementation of this idea in terms of head movement by reprojection). For now, I adopt the former view: AGENT DPs are arguments of v. Thus, (agentive) v is equipped with an ordered set of structure-building features $[\bullet V\bullet] \succ [\bullet D\bullet]$ – first, the VP argument of v is merged (discharging $[\bullet V\bullet]$), and then its DP specifier enters the structure (discharging $[\bullet D\bullet]$). Of course, Last Resort must be revised appropriately so as to be sensitive to the order of structure-building features; this is done in (10).

(10)   *Last Resort* (LR, revised):
    a.   Every syntactic operation must discharge either $[\bullet F\bullet]$ or $[*F*]$.
    b.   Only features on the top of a feature list are accessible.

(10-b) presupposes that a discharged feature is removed from the lexical item, and deleted. Issues related to morphological realization of case and agreement features may ultimately demand a slightly more complicated approach (according to which discharged features become inactive for further processes in the sense of (10-b) but remain accessible for morphological realization; see Adger (2003), Müller (2009) for discussion); but for present purposes we may keep to the simplest assumption.[8]

The question arises of whether probe features on a lexical item are treated in the same way as structure-building features. In line with the definition in (10), I assume that this is indeed the case, for reasons of symmetry alone: Probe features show up on a separate stack (though their order will not play a role in this chapter). Under these assumptions, there are two stacks of features on lexical items – one for structure-building features, and one for probe features. Consequently, given Last Resort in (10), lexical heads look like *pushdown automata*: They are characterized by a last-in/first-out property (although so far, I have exclusively been concerned with the first-out part; the last-in part will be highlighted below).

2.3   All Phrases are Phases

This assumption has already been adopted, and argued for, in chapter 3 (see subsection 2.3 there). The status of this assumption in the context of deriving CED effects will turn out to be similar to its status in the context of deriving (G)MLC effects: If it is maintained that fewer XPs qualify as phases (e.g., only vP and CP), this means that fewer types of CED effects can be accounted for; but it does not mean that the enterprise as such must be called into question.

---

[8] It is also worth bearing in mind that potential problems with the hypothesis of discharge of $[\bullet F\bullet]$, $[*F*]$ as deletion for the morphological component do not arise if a pre-syntactic approach to inflectional morphology is adopted; see, among others, Wunderlich (1996; 1997b), Alexiadou & Müller (2008), and Müller (2008).

2.4   Edge Feature Insertion

Finally, as noted, the most important assumption that is needed to derive the CED in (5) from the PIC concerns the insertion of edge features. According to the Edge Feature Condition in (4) (or, for that matter, Chomsky's original version in (3)), an edge feature may be inserted on a phase head so as to trigger an intermediate movement step (and thereby avoid a PIC violation) only *after* the phase is otherwise complete. Chomsky (2000, 109) defines the stage of the derivation at which a phase is "complete" in this sense as follows: A phase is complete if it has "exhaust[ed] the lexical subarray from which it is derived". Assuming that all syntactic operations are triggered by the phase head, this would seem to strongly suggest that the phase head is inert once the phase is complete. Hence, the null hypothesis surely is that the phase is inaccessible for further internal modification at this point. From this perspective, subsequent edge feature insertion (as in (4), (3)) is a peculiar operation. I would therefore like to propose that *after* is to replaced with *before* in the definition of the EFC in (4) (or (3)): Edge features can only be inserted *as long as the phase head is active*. Given the assumption that all syntactic operations are (induced by properties of a phase head and) feature-driven (see subsection 2.1 above), activity of a phase head (in the relevant sense) can be defined as follows: A phase head is active as long as it has (structure-building or probe) features to discharge; otherwise it counts as inactive.[9] This leads to the revised version of the Edge Feature Condition in (11).

(11)   *Edge Feature Condition* (EFC, new version):
       An edge feature [•X•] can be assigned to the head γ of a phase only if (a), (b), and (c) hold:

       a.   The phase headed by γ is otherwise not balanced.
       b. γ has not yet discharged all its structure-building or probe features.
       c.   [•X•] ends up on top of γ's list of structure-building features.

Clause (11-a) is nothing new; it represents the phase balance requirement of (4). (11-b) incorporates the new assumption about the timing of edge feature insertion; and (11-c) just makes precise the assumption that phase heads act like pushdown automata – an edge feature that is assigned to a phase head ends up on top of the stack of structure-building features that is present on the phase head (this then fully accounts for the

---

[9] Thus, access to a concept like "lexical subarray" is not required in the present approach to determine inertness of a phase head. As a matter of fact, the balanced phase approach of chapter 3 depends on the assumption that the whole numeration is accessible at all points of the derivation; and the same goes for the approach to freezing developed in section 4 below. In principle, it would still be possible to postulate designated subarrays of the numeration that correspond to phases; I will leave open the question of whether such a move is warranted.

last-in/first-out property mentioned at the end of subsection 2.2).[10] To sum up, the important consequence of (11) (in interaction with the first three assumptions) is that the assignment of a movement-inducing edge feature [●X●] to a phase head γ can take place only to the top of a feature list on γ, and that γ must have a non-empty feature list at this point.[11]

From these four assumptions, the CED in (5) follows as a theorem. This is shown in the following section.

3.   Deriving the Condition on Extraction Domain

3.1   Merge

3.1.1   The Analysis

Let us ignore probe features on phase heads for the moment, and focus on structure-building features only. Given this proviso, the CED in (5) can be derived as follows. First, if an edge feature feature [●X●] is to be inserted on a phase head γ, γ must contain at least one other feature at this point; otherwise it is inert and cannot be affected by edge feature insertion (this follows from clause (b) of the EFC in (11)). Second, edge feature insertion must go to the top of γ's list of structure-building features (this follows from clause (c) of the EFC in (11)). Third, if an edge feature [●X●] has been inserted on a phase head's stack of structure-building features, it is discharged again immediately, as an instance of the last-in/first-out property of phase heads (this follows from clause (b) of Last Resort in (10)). Fourth, and most importantly, this means that it is impossible to insert an edge feature for a category α that is merged in Γ (i.e., the maximal projection of a phase head γ) as the last operation taking place in Γ (because of the EFC in (11) – in this case, at the point where α is part of the structure, the head of Γ has no feature left, and has therefore become inert, making edge feature insertion impossible). Fifth, this in turn implies that a moved item β in the edge domain of a category α merged last in a phase Γ is not accessible anymore outside Γ: If α is last-merged, an edge feature cannot be placed on the head of Γ anymore, and a category that is part of α cannot be moved out of α, to an outer specifier of Γ. Therefore, once the derivation proceeds by combining Γ with some new item from the numeration (or

---

[10] As observed by Winfried Lechner (p.c.), the fact that feature assignment must always go to the top of a feature stack is arguably just a consequence of a general principle regulating derivational structure generation, viz., the Strict Cycle Condition (SCC) (see (23) of chapter 1 or (8) of chapter 2, and (1) of chapter 3; also the references given in footnote 2 of chapter 3. The SCC will become relevant again below.

[11] In addition, edge feature insertion must be the only way to produce a balanced phase. In what follows, this requirement will play no role since I will only be concerned with data that do not instantiate any kind of (potential) intervention effect. I will therefore generally tacitly presuppose that this third requirement is met whenever the question of edge feature insertion arises.

the workspace), the PIC in (2) renders β in α inaccessible.[12] Sixth, it thus follows that extraction from α is predicted to be impossible if α enters a phase as a result of the last operation taking place in that phase, as a consequence of the phase head's features (because of the PIC). Therefore, and this concludes the argument, given that (outer) specifiers are last-merged in their projections, they are predicted to be barriers for movement: The CED in (5) is derived.

3.1.2 Specifiers

As an illustration, consider the (possibly simplified) model of clause structure in (12).

(12)   ...
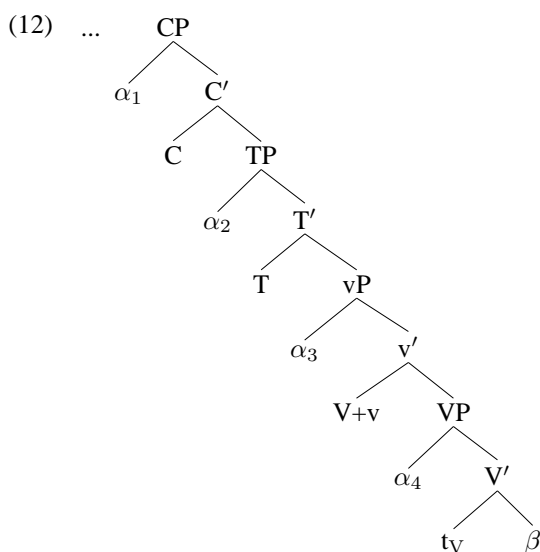
Assume that α_n in (12) is a specifier that is last-merged in its phase whereas β is a complement. The task then is to derive (in accordance with (5)) that α_n is a barrier whereas β is not (necessarily; see below) a barrier. Turning to (last-merged) speci-fiers first, it remains to be shown that no order of operations can yield a legitimate outcome if some item is to be extracted out of α_n in (12). Suppose first that edge feature insertion follows specifier feature discharge. This gives rise to the sequence in (13-a), which violates (11-b) – an edge feature is assigned to a phase head that has

---

[12] Note that this reasoning presupposes a non-recursive notion of edge, such that the edge β of a category α is not part of the edge of a phase Γ even if α is part of the edge of Γ. – At this point, it should be acknowledged that while non-recursive edges follow directly from the currently employed (and arguably maximally simple) notion of edge as consisting of all and only specifiers (perhaps only left-peripheral specifiers, and perhaps also adjuncts; see (3) in chapter 3), non-recursive edges might in fact not be an optimal solution if the PIC is not viewed as a primitive condition but derived from cyclic spell-out (cf. footnote 5 in chapter 2); see Richards (2010) for this point and a different approach to CED effects motivated by it.

discharged all its inherent structure-building features and has thereby become inert.[13]
Suppose next that the order of operations is reversed, i.e., edge feature insertion pre-
cedes specifier feature discharge. There are two possible scenarios. One possibility
is that the edge feature is assigned to the phase head, but does not end up on top of
the phase head's stack of structure-building features; see (13-b). This violates (11-c) –
inserted edge features must go to the top of a feature stack. Alternatively, edge feature
insertion precedes specifier feature discharge, but this time, the edge feature is indeed
assigned to the top of the stack, as in (13-c). Unfortunately, this does not help either,
because of (10-b): After the edge feature is inserted, it has to go away directly, given
Last Resort (attracting some category within the complement of the phase head to an
outer specifier). Therefore, such an edge feature cannot serve to move anything out of
α: When α enters the structure (as a result of [$\bullet\alpha\bullet$] discharge), the edge feature [$\bullet X\bullet$]
has come and gone already.

(13) *Why specifiers in (12) are barriers*:

a.
$$
\begin{array}{l}
\gamma:\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ \varnothing \\
\rightarrow\ \gamma:\ [\bullet X\bullet]
\end{array}
$$
$\rightsquigarrow$ violates (11-b)

b.
$$
\begin{array}{l}
\gamma:\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet\alpha\bullet]\ \succ\ [\bullet X\bullet] \\
\rightarrow\ \gamma:\ [\bullet X\bullet]
\end{array}
$$
$\rightsquigarrow$ violates (11-c)

c.
$$
\begin{array}{l}
\gamma:\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet X\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet\alpha\bullet]
\end{array}
$$
$\rightsquigarrow$ does not help because of (10-b)

To conclude, specifiers are barriers because of the PIC: There is no way to carry out
an intermediate movement step from a last-merged specifier to the specifier of the
minimal phase above it; so movement out of a last-merged specifier has to skip the
edge domain of the next phase, in violation of the PIC.

3.1.3   Complements

At this point, the question arises of why complements are not barriers (or, more pre-
cisely, do not have to be barriers – see below). In (12), β, VP, vP, TP, and (possi-
bly) CP are complements that are first-merged in their phases and have not yet dis-
charged the final structure-building feature of the respective phase head's feature list;
in other words, the operation that brings the complement into the structure is not the

---

[13] A remark on notation: α refers to a syntactic category (a phrase, possibly with rich internal structure);
in contrast, [$\bullet\alpha\bullet$] is a subcategorization feature selecting the category label of α. Thus, strictly speaking,
"alpha" is ambiguous between a syntactic category and its label. This simplification is harmless.

last structure-building operation carried out in the phase. Again, several a priori possible derivations need to be considered, but now, there has to be at least one derivation that succeeds. As it turns out, this derivation is straightforward. As shown in (14-a), no constraint is violated if edge feature insertion follows complement feature discharge but precedes specifier feature discharge. At the point where assignment of the edge feature to the phase head is called for (so as to move an item out of the complement β to an outer specifier of the phase), the complement feature $[\bullet\beta\bullet]$ must have been discharged, but the specifier-creating feature $[\bullet\alpha\bullet]$ is still present on the head. Therefore, edge feature insertion finds a stage of the derivation where it can apply non-vacuously without violating the ban on modifying inactive phase heads incorporated into the EFC. Of course, alternative derivations that mirror those in (13-b) and (13-c) are excluded for the same reasons as before: If edge feature insertion applies very early and precedes complement feature discharge, either the inserted edge feature fails to end up on top of the phase head's feature stack (as in (14-b)), or the edge feature is discharged again immediately (by merging some arbitrary category), without being able to attract anything out of the complement (which is not yet part of the structure); see (14-c).

(14)   *Why complements in (12) do not have to be barriers*:

     a.

$$
\begin{array}{l}
\gamma:\ [\bullet\beta\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet X\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ \varnothing
\end{array}
$$

                                $\rightsquigarrow$ violates nothing

     b.

$$
\begin{array}{l}
\gamma:\ [\bullet\beta\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet\beta\bullet]\ \succ\ [\bullet X\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet X\bullet]\ \succ\ [\bullet\alpha\bullet]
\end{array}
$$

                                $\rightsquigarrow$ violates (11-c)

     c.

$$
\begin{array}{l}
\gamma:\ [\bullet\beta\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet X\bullet]\ \succ\ [\bullet\beta\bullet]\ \succ\ [\bullet\alpha\bullet] \\
\rightarrow\ \gamma:\ [\bullet\beta\bullet]\ \succ\ [\bullet\alpha\bullet]
\end{array}
$$

                        $\rightsquigarrow$ does not help because of (10-b)

Thus, movement out of complements can respect the PIC: There is a stage in the derivation where the complement feature has already been discharged (so that subsequent edge feature insertion can attract an item within the complement), but the specifier feature has not yet been discharged.

3.1.4   The Intermediate Step Corollary

The present approach makes an interesting prediction concerning the exact nature of the landing site of intermediate movement steps: Intermediate movement steps to phase edges must take place *before* a (final) specifier is merged. Assuming that tucking in (in the sense of Richards (2001)) is not an option in natural languages (see below),

this implies that subsequent Merge of a specifier ends up in a higher position. Thus, we end up with the result that an intermediate movement step (triggered by an edge feature that is not inherently present on a phase head) ends up in a specifier that is asymmetrically c-commanded (and preceded) by a category whose presence is due to an inherent (e.g., subcategorization) feature of the same phase head. Let us call this property the *Intermediate Step Corollary*.

(15) *Intermediate Step Corollary*:
      Intermediate movement steps to specifiers of X (as required by the PIC) must take place before a final specifier is merged in XP.

I take this corollary of the revised EFC under the present set of assumptions to be somewhat unusual but entirely unproblematic.[14] The consequence of (15) for successive-cyclic movement is shown schematically in (16), where successive-cyclic movement of the complement of V (rather than out of the complement of V) first takes place to an inner specifier of V; then to an inner specifier of v (out of the complement VP); and then on to some specifier outside of the complement vP.[15]

(16)   DP$_2$ ...
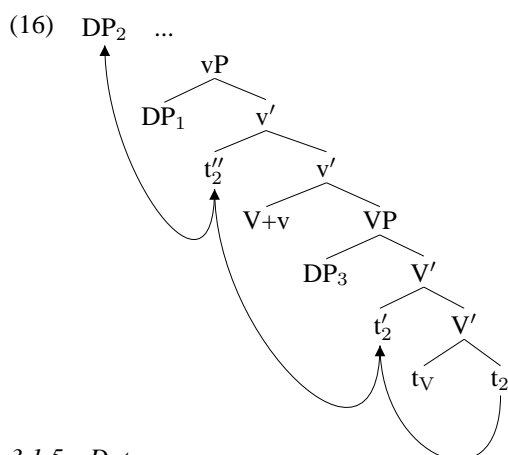
3.1.5   Data

This way, standard CED data are derived, like the Sentential Subject Constraint effect (cf. (69) in chapter 1) in English that is illustrated in (17) (assuming that the effect is indeed real, effective before clausal extraposition, and not an instance of something else, as argued by Koster (1978)); or the standard Subject Condition effect with DPs in

---

[14] What is more, in chapter 5 I will argue that the Intermediate Step Corollary actually holds the key to a new understanding of operator island (i.e., *wh*-island and topic island) effects that does not rely on a constraint like Relativized Minimality.

[15] Also recall from chapter 3 that the present approach, where all phrases are phases, is inherently incompatible with Anti-Locality requirements on movement (see Grohmann (2000; 2003), Abels (2003)).

(18). External argument CPs and DPs are merged in Specv as the last operation taking place in vP; an edge feature cannot be instantiated on v after the external argument has been merged (because the phase head is inert at this point); and subsequent movement of the *wh*-phrase in (17) and (18) out of the external argument therefore violates the PIC.

(17) a. $[_{DP_1}$ Who] did the reporters expect $[_{CP}$ that the principal would fire t_1] ?
 b. *$[_{DP_1}$ Who] was $[_{CP}$ that the principal would fire t_1] expected by the reporters ?
 c. *$[_{DP_1}$ Who] did $[_{CP}$ that Mary was going out with t_1] bother you ?

(18) a. *$[_{DP_2}$ Who(m)] has $[_{DP_1}$ a comment about t_2] annoyed you ?
 b. *$[_{PP_3}$ About whom] has $[_{DP_1}$ a comment t_3] annoyed you ?

The CED in its classical formulation covers subjects and adjuncts. The barrier status of adjuncts follows immediately if we assume that adjuncts are to be reanalyzed as last-merged specifiers of special functional projections (see Alexiadou (1997), Cinque (1999)). The reasoning is then identical to the one given before for subjects. Some standard English examples instantiating the Adjunct Condition part of the CED are given in (19) (see (95) in chapter 1).[16]

(19) a. *$[_{DP_1}$ Who] did you get jealous $[_{CP}$ because I talked to t_1] ?
 b. *$[_{PP_1}$ To whom] did they leave $[_{CP}$ before speaking t_1] ?
 c. *$[_{DP_1}$ Who] did they leave $[_{CP}$ before speaking to t_1] ?

Finally, given that dative ('affected') GOAL arguments are merged last in VP (see subsection 3.1.2 of chapter 3), in SpecV (in contrast to other GOAL arguments that are realized by PPs), they are predicted to invariably qualify as barriers, just like external arguments merged in Specv. This prediction seems correct for German; cf. (20-a) (external nominative DP as a barrier) and (20-b) (dative DP as a barrier); see Fanselow (1987; 2001a), Grewendorf (1989), and Müller (1995).[17]

---

[16] That said, it has sometimes been observed that adjuncts tend to be stricter islands than subjects, and less susceptible to cross-linguistic variation (see Frank (2002), Stepanov (2007), and chapter 2). The present analysis does not rule out the possibility that some other factor might additionally block extraction from adjuncts, or preclude what might exist as a way out for extraction from subjects in some languages after all (see section 6 below for some speculations).

[17] As remarked before, subject raising to SpecT is optional in German; so both a derivation in which the subject DP in (20-a) is in Specv, and one in which it is in SpecT, must be excluded. Also, as noted in chapter 2, Jurka (2008; 2010) presents an experimental study suggesting that there are CED effects with in situ subjects in Specv in German.

It has sometimes been argued that certain kinds of subject DPs are transparent for extraction in German; see, e.g., Haider (1983; 1993a) and Diesing (1992). As noted by Fanselow (2001a, 422), many of the putative

(20) a. *[$_{PP_1}$ Über wen] hat [$_{DP}$ ein Buch t_1] den Fritz   beeindruckt ?
          about whom has   a   book$_{nom}$ the Fritz$_{acc}$ impressed

     b. *[$_{PP_1}$ Über wen] hat der Verleger   [$_{DP}$ einem Buch t_1] einen
          about whom has the publisher$_{nom}$ a   book$_{dat}$ a
       Preis   gegeben ?
       price$_{acc}$ given

3.2   Agree

3.2.1   Extraction from Complements Reconsidered

The present analysis makes two related predictions for extraction from objects that look problematic. First, given what has been said so far, we expect that THEME (accusative) objects can avoid becoming a barrier in VP only if there is something else that is merged later in the same phase (e.g., a GOAL – dative – object). This prediction is not borne out; consider, e.g., the German examples in (21-a) (extraction from an accusative object in the presence of a higher dative object with a ditransitive verb) and (21-b) (extraction from an accusative object with a simple transitive verb). Second, as has been observed by Cattell (1976) and (from a slightly different point of view) Bach & Horn (1976), movement out of THEME (accusative) objects is known to depend

---

counter-examples in German involve passive or unaccusative constructions with the nominative DP in situ, in a complement position, as in (i-ab). See the next subsection for an analysis, and also (25) below.

(i) a. [$_{PP_1}$ Über wen] wurde [$_{DP}$ ein Buch t_1] gelesen ?
         about whom was   a book$_{nom}$ read

    b. [$_{PP_1}$ Über wen] ist [$_{DP}$ ein Buch t_1] erschienen ?
        about whom is   a book$_{nom}$ appeared

The contrasts between extractions from nominative DPs with individual-level and stage-level predicates noted in Diesing (1992), to the extent that they are real, can presumably for the most part be reinterpreted along the same lines, as reflecting a difference between VP-internal complements and specifiers of v; see (ii-ab) (my judgements; Diesing assigns a * to (ii-a)).

(ii) a. ??Was$_1$ sind [$_{DP}$ t_1 für Leguane]   intelligent ?
         what are   for iguanas$_{nom}$ intelligent

    b. Was$_1$ sind [$_{DP}$ t_1 für Leguane]   verfügbar ?
        what are   for iguanas$_{nom}$ available

(Here and in what follows, I presuppose that *was für* ('what for') phrases are DPs; however, the exact categorial status and internal organization of these arguments are not really important for any of the substantial points made in this chapter. Still, it can be noted that the kind of structure attributed to *was für* phrases by Leu (2008) is incompatible with the present approach because it relies on the availability of extraction from a last-merged specifier.)

Other apparent counter-examples involve PPs headed by *von* ('of'), as in (iii-a), or *zu* ('to'), as in (iii-b) (see Haider (1993a, 172-173) and references cited there). For these, a non-movement analysis seems systematically available; see the discussion around (35) in section 4. A similar reasoning applies in the case of DP split constructions such as (iii-c) (see Fanselow (1988), Riemsdijk (1989), Fanselow & Ćavar

on a tight relation between V and N (the formation of a "natural predicate"); see the contrast between (21-b) and (21-c), which so far also remains without an explanation. Thus, the fact that accusative DPs in German (and other languages) that are simultaneously first- and last-merged in VP are partly, but not fully transparent local domains is not accounted for under present assumptions.

(21) a. [$_{PP_1}$ Worüber] hat er   der Maria   [$_{DP}$ ein Buch t_1] gegeben ?
          about what has he$_{nom}$ the Maria$_{dat}$ a   book$_{acc}$ given

     b. [$_{PP_1}$ Worüber] hat der Fritz   [$_{DP}$ ein Buch t_1] gelesen ?
          about what has the Fritz$_{nom}$ a   book$_{acc}$ read

     c. *[$_{PP_1}$ Worüber] hat der Fritz   [$_{DP}$ ein Buch t_1] geklaut ?
          about what has the Fritz$_{nom}$ a   book$_{acc}$ stolen

3.2.2 *Abstract Incorporation as Head Movement*

There are various ways to account for the contrast in (21-bc). An approach that I assume to be essentially on the right track centers around the notion of "abstract incorporation" and can be traced back to Baker (1988); versions of this general approach have been adopted in Müller (1995), Sauerland (1995), Davies & Dubinsky (2003), Schmellentin (2006), and Ott (2011b), among others. The basic idea is as follows: Abstract incorporation of N into V (i.e., head movement of N to V at the level of Logical Form (LF)) must apply to remove an initial barrier status from an NP.[18] Abstract incorporation may take place only at LF, but it can be viewed as co-indexing under

---

(2002)), which systematically violate constraints on extraction in German and may not involve movement at all in this language (though see Ott (2011a) for a recent movement-based approach that derives the non-occurrence of certain – low – island effects with this construction by assuming that the original constituent containing items like NP and DP in (iii-c) is not labelled in the base, and hence (by assumption) not a barrier.)

(iii) a. [$_{PP_1}$ Von den Studenten] haben [$_{DP}$ viele (t_1)] die Prüfung nicht geschafft
          of   the students   have   many   the exam   not   made

     b. ?[$_{PP_1}$ Zu   diesem Problem] haben [$_{DP}$ einige Briefe (t_1)] den Sender erreicht
          to/concerning this   problem have   several letters$_{nom}$ the station reached

     c. [$_{NP}$ Briefe]   haben mich [$_{DP}$ keine   (t_1)] erreicht
          letters$_{nom}$ have me   none$_{nom}$ reached

Finally, the most convincing counter-examples to the generalization that subject DPs in Specv (or, optionally, SpecT) are barriers for extraction in German typically involve configurations where the subject DP is adjacent to V as a result of object scrambling. I take this phenomenon to be real. As I will show in section 5, this is an instantiation of a systematic melting effect, and entirely expected under present assumptions.

[18] Thus, head movement opens barriers. See Bobaljik & Wurmbrand (2003), Gallego & Uriagereka (2006), and den Dikken (2007; 2008) for recent implementations of the idea that head movement extends local domains and thereby opens up barriers.

minimality (cf. the Head Movement Constraint, HMC) in syntax.[19] Whether abstract incorporation can apply or not is partly determined by lexical factors: V-N:*read-book* is a natural predicate resulting from abstract incorporation, but V-N:*steal-book* is not (for most speakers); so abstract incorporation may take place in the former, but not in the latter case, removing the barrier in (21-b) but leaving it intact in (21-c).

This analysis arguably works well as such; however, there is a potential problem with the approach: The analysis is incompatible with DP-over-NP structures (see Abney (1987), Szabolcsi (1994); and Alexiadou, Haegeman & Stavrou (2007) for a more recent overview). The reason is that N cannot abstractly incorporate into V in the presence of an intervening D; such a movement step would be expected to give rise to a HMC (or minimality) violation (see subsection 3.12 of chapter 1; but also chapter 6, where the HMC will be called into question). I would therefore like to suggest a reanalysis of the abstract incorporation approach to extraction from nominal projections.

3.2.3 Abstract Incorporation as Agree

Suppose that abstract incorporation does not involve movement at any stage of the derivation. Rather, it is an instance of Agree, in the following sense: If V and N form a natural predicate, they share an abstract feature $[*p*]/[p]$ (where 'p' is mnemonic for 'natural predicate'). Agree requires c-command, but is less local: An intervening D is unproblematic. Given that abstract incorporation is reanalyzed as an instance of $[*p*]/[p]$ Agree, we can now assume that *read-book* involves a $[*p*]/[p]$ Agree relation (hence, a natural predicate), whereas *steal-book* does not; cf. (22).[20]

(22) a. $[_{VP} V_{[*p*]} [_{DP} D [_{NP} N_{[p]} PP]]]$ *read book*
     b. $[_{VP} V [_{DP} D [_{NP} N PP]]]$ *steal book*

Of course, this analysis implies that Agree operations can find a way to avoid PIC violations that does not exist for movement (from non-edge domains of a phase). More specifically, it seems that either the PIC is relaxed for Agree operations (as has been proposed by Bošković (2007), among others), or that Agree can be successive-cyclic,

---

[19] Alternatively, abstract incorporation may be viewed as overt, syntactic head movement that is accompanied by a pronunciation of the trace, as in Stechow (1992) and Fanselow (2001b).

[20] Whereas *read-book* is a natural predicate (permitting extraction from DP) and *take-picture* is also a natural predicate (permitting extraction from DP), *take-book* (for most speakers at least) is not (which blocks extraction from DP here). This effect is discussed, and derived, in at least some of the work on abstract incorporation cited above. In the present context, it may suggest an understanding of $[*p*]/[p]$ Agree according to which a natural predicate relation of V and N (which is eventually encoded in the lexicon) is captured by $[*p*]$ on V and $[p]$ on N; but it is not the case that some V would inherently be marked $[*p*]$, or some N would inherently bear $[p]$.

proceeding from one head to the next one higher up (see Legate (2005), among others). A relaxation of Agree relations along these lines is required independently, and under standard (i.e., somewhat less local) conceptions of what a phase is (see Chomsky (2000; 2001b)) for cases of long-distance agreement, at least for those cases of long-distance agreement that can be shown not to be parasitic on XP movement (as proposed by Polinsky & Potsdam (2001) for Tsez), and that cannot be reanalyzed in terms of extremely local domains after all (see Boeckx (2004)). Given that cases of long-distance agreement should not be taken to motivate a less restrictive concept of phase as such (as argued by Chomsky (2001b) for Agree applying to T and a nominative object in Icelandic, and by Bhatt (2005) for long-distance agreement in Hindi), exempting Agree from the PIC and assuming cyclic Agree look initially like the most plausible options. Of these two basic possibilities, the former one is also excluded if hypothesis (51) of chapter 3 is adopted, which would then effectively only leave cyclic Agree as an analytical option.[21]

Given that probe features on V can remove barrier status from an accusative object DP in VP, it must be ensured that this option does not arise for dative objects in SpecV – more generally, probe features on a phase head can never remove barrier status from a last-merged specifier (like an external argument in Specv). This follows if two assumptions are made. First, a probe feature cannot carry out Agree with (some item in) its specifier (see Chomsky (2001b; 2008)). This can be derived under the hypothesis that Agree requires c-command, and is thus restricted to the complement domain of a head.

Second, a probe feature cannot carry out Agree with (some item in) its complement after a specifier has been merged. To see why this must be excluded, consider the following scenario: Suppose that a phase head γ has been merged with some other category which becomes its complement, with one structure-building feature

---

[21] Note that under the present set of assumptions, Agree operations must be permitted to cross a phase boundary even for instances of local case assignment – e.g., accusative case assignment by v to DP in VP. This reasoning already suggests that it cannot be a case or agreement-related Agree relation between v and an object DP in VP that is responsible for the latter's transparency for extraction (notwithstanding the lexical variation visible in (21-b) vs. (21-c), which would remain unaccounted for): When an edge feature is needed to move an object DP (or something contained in that object DP) from (within) the complement position in VP to a SpecV position, v is not yet present – so the relevant probe feature must be located on V. Thus, a case-based approach might arguably work with instances of lexical case assignment by V, but it would not work with structural case assignment by v. – Curiously enough, lexical case-marking typically induces barrier status on the lexically case-marked DP, at least in a language like German. This does not yet follow in the present approach. In fact, additional assumptions would be necessary to ensure the mere *possibility* of a barrier status of lexically case-marked DPs. A stipulation that would yield the intended result would be that V can assign lexical case only after movement to v; note that this would be compatible with the SCC to be discussed below, in (23).

[•F•] remaining on γ in the stack of structure-building features, and one probe feature [∗F∗] remaining on γ in the separate stack of probe features. For instance, v has been merged with VP, and now has a subcategorization feature [•D•] left that will generate an external argument specifier, and a [∗acc∗] case feature (alternatively, a probe that captures a relevant bundle of ϕ-features) on v for Agree with an object DP in VP. Then, Agree(v,α in VP) (for accusative case assignment and ϕ-feature agreement) must precede Merge(DP$_{ext}$,v′) – if it did not, i.e., if the two operations were permitted to apply in reverse order, v would still be active (in the sense of the EFC in (11)) after the external argument DP has been merged, and could be provided with an edge feature attracting an item out of the external argument DP, to an outer specifier of v, after all. This would make it possible to avoid a PIC violation in the next step, and would thus undermine a derivation of the CED in (5) from the PIC. Therefore, it must be ensured that all probe features of a phase head are discharged before a specifier (with which the probe can never check a probe feature, given the first assumption) is introduced. Indeed, the ban against leaving probe features undischarged on a phase head follows without further ado from a restrictive version of the *Strict Cycle Condition* (see Chomsky (1973)), as it is given in (23).[22]

(23)   *Strict Cycle Condition* (SCC):
       Within the current domain Γ, a binary syntactic operation may not exclusively apply to positions α, β if α and β are both included within another domain Δ that is dominated by Γ.

Consequently, if a phase head γ has an undischarged probe feature [∗F∗] after it is merged with a category that becomes its specifier, no legitimate continuation of the derivation will be possible: If [∗F∗] on γ undergoes Agree with γ's specifier, the c-command requirement on Agree operations is violated; and if [∗F∗] undergoes Agree with (an item within) γ's complement, the SCC is violated (within the projection of γ that includes the specifier, such an Agree operation would exclusively apply to the

---

[22] The definition of the Strict Cycle Condition in (23) differs in subtle respects from the one given in (23) in chapter 1, (8) in chapter 2, and (1) in chapter 3. For one thing, it is confined to binary operations; for another, it recognizes all projections as cyclic domains, not just XPs. This latter assumption ensures that scenarios involving tucking in as they are envisaged in Richards (2001) can also be excluded by the SCC; recall the discussion of (16). As will become clear momentarily in the main text, assuming all projections to be cyclic domains is also important in the present context. Still, it should be clear that all previous evidence for the SCC encountered so far is still compatible with the new formulation. In particular, the conclusion that a post-movement insertion of intermediate traces as an instance of Chomsky's (1991) complex operation Form Chain violates the Strict Cycle Condition (see the discussion of (22) in chapter 1) is still valid.

lower projection of γ that only includes γ and its complement).[23]

3.2.4 Transparent Last-Merged Complements

In sum, the account of the (lexically determined) potential absence of barrier status with *last-merged* complements can be schematically depicted as in (24). Suppose that a phase head γ (e.g., V) is merged with α (e.g., DP) and has thereby discharged all its structure-building features. The first possibility then is that edge feature insertion follows complement feature discharge, and there is no probe feature on the phase head. In this case, edge feature insertion violates clause (b) of the EFC (and if no edge feature can be inserted, subsequent extraction from the last-merged complement violates the PIC); see (24-a). Alternatively, edge feature insertion follows complement feature discharge, with a probe feature still present on the phase head, as in (24-b). This derivation violates nothing.

(24) *Why last-merged complements do not have to be barriers*:

    a.

$$
\begin{array}{c}
\boxed{\gamma:\ [\bullet\alpha\bullet]} \\
\rightarrow \boxed{\gamma:\ \varnothing} \\
\rightarrow \boxed{\gamma:\ [\bullet X\bullet]}
\end{array}
\qquad \rightsquigarrow \text{violates (11-b)}
$$

    b.

$$
\begin{array}{c}
\boxed{\begin{array}{c}\gamma:\ [\bullet\alpha\bullet] \\ [*F*]\end{array}} \\
\rightarrow \gamma:\ [*F*] \\
\rightarrow \boxed{\begin{array}{c}\gamma:\ [\bullet X\bullet] \\ [*F*]\end{array}}
\end{array}
\qquad \rightsquigarrow \text{violates nothing}
$$

Crucially, to avoid a SCC violation (as it would occur with specifiers), the probe feature $[*F*]$ must be discharged before the structure-building edge feature $[\bullet X\bullet]$ is dis-

---

[23] At this point, a remark may be due concerning the assumption that structure-building and probe features show up on separate stacks of a head, which I have adopted throughout this chapter (cf. page 168). The alternative would be that the two feature types show up in a single hierarchy, in which case they would have to be interleaved: Probes can only be discharged after a complement has been merged, and before a specifier is merged, given the SCC; otherwise, the derivation would break down. However, there is a conceptual and an empirical problem with such a view. On the conceptual side, the order of the mixed classes of features on a single head would then have to be stipulated as a lexical property of the head even though it can be predicted from the interaction of the SCC and the fact that an Agree operation can only be carried out if there are two distinct categories involved (with one of them c-commanding the other). The two-stack approach avoids this redundancy. In addition, there is another, more pressing, empirical problem with the one-stack approach: A probe feature on a phase head that makes edge feature assignment to that phase head possible must be discharged *before* the derivationally assigned edge feature is discharged, for the reasons to be discussed immediately in the main text. But there is no way to derive this if there is only a single, mixed stack of features: An assigned edge feature would end up on top of the stack, above the probe feature, and the probe feature could then not be discharged before the specifier is introduced.

charged in (24-b) (and this is unproblematic given that the two features are on different stacks; see the last footnote).

The same reasoning as the one given here for last-merged complement DPs may also be relevant for projections of the clausal spine, which do not necessarily take specifiers. For instance, in German, there are clauses where VP, vP, TP, and CP all fail to be inherently equipped with structure-building features that generate a specifier. Still, as shown in (25), extraction from an object DP in VP may take place (here, the sole DP argument of an unaccusative verb stays in situ, and there is long-distance *wh*-extraction from the DP; see Grewendorf (1989) and Fanselow (2001a)).

(25)   Worüber$_2$ denkst du   [$_{CP}$ t$_2''''''$ dass [$_{TP}$ t$_2'''''$ [$_{vP}$ t$_2''''$ [$_{VP}$ t$_2'''$ [$_{DP}$ t$_2''$ ein [$_{NP}$ t$_2'$
        about what think   you   that
        Buch t$_2$]] t$_V$] erscheint-v] T]] ?
        book   appears

In (25), the projections of the clausal spine (CP, TP, vP, VP) can avoid barrier status if their heads undergo Agree operations with the minimally c-commanding head. Extending the concept of *status-government* introduced in Bech (1955/1957) in such a way that a clausal head status-governs the head of its verb complement, it is assumed in Sternefeld (1991b) and Müller (1995) that such a status-government systematically removes barrier status from clausal projections in German (and other languages). In the studies just cited, status-government is modelled in terms of abstract incorporation (co-indexing of heads); in the present approach, this translates into Agree relations encoding the concept of status-government. Consequently, clausal projections are not barriers, even if there is no regular specifier present (and the projection is thus last-merged). Exactly the same reasoning can be taken to apply for the structure of DPs (see, for instance, the D-N relation in (25)). Arguably, the difference between bridge verbs and non-bridge verbs can be modelled in the same way (but cf. chapter 6 below for a qualification): A bridge verb (like *denken* ('think') in the German example (25)) may undero Agree with the head of its clausal complement CP; a non-bridge verb (like *wissen* ('know'), which gives rise to ungrammaticality if it replaces *denken* in (25)) may fail to be able to undergo Agree with the head of its CP sister.[24]

---

[24] If a C-T Agree relation were to fail to obtain in certain languages if C has certain properties, complementizer-trace effects could arise as consequences of a TP barrier (i.e., the impossibility of inserting an edge feature on C for an item in the edge domain of TP). See Koster (1986) and Müller (1995) for such an approach to complementizer-trace effects in languages like English. However, a bit more would have to be said under these assumptions to make non-subject movement across a lexical complementizer *that* possible. The more general question that arises in this context is how to account for differences between various types of categories with respect to island violations (subject vs. object, and argument vs. adjunct). I have nothing insightful to say about this issue here; but cf. chapter 5 for some pertinent remarks.

3.2.5   Movement of Last-Merged Complements

At this point, the related question needs to be addressed of how the very first step in the course of successive-cyclic extraction not *from*, but *of* a complement can be brought about in cases where there is no second item that is merged in the same projection as a consequence of the phase head's inherent structure-building features; consider, e.g., the extremely local step from CompN to SpecN that would seem to be demanded by the PIC in (25).[25] The first thing to note is that the present set of assumptions leads us to expect that this first movement step of a last-merged complement is only possible if the head has a probe feature left that renders it active in the sense of the EFC, and thus permits edge feature insertion. Against this background, the following possibilities arise.

First, heads typically undergo various kinds of Agree relations with their complements, and this might account for the fact that (last-merged) complements are typically mobile. Still, in some cases, where such an Agree relation between a head and its (last-merged) complement might not hold, the complement might in fact be precluded from moving. In a way, this would amount to a version of a requirement of *lexical government* in the ECP.[26]

A second option might be to ensure the mobility of complements in general by assuming that there is *always* a probe feature that triggers an Agree operation between head and complement; the underlying motivation might be that the head-complement relation is privileged over other syntactic relations, and a head must carry out a special identification of its complement. The simplest technical implementation of this assumption may be to assume that the probe feature that is involved in obligatory head-complement Agree is just a copy of the subcategorization feature. On this view, if a head has a subcategorization feature [•F•] on top of its (possibly singleton) stack of structure-building features, it is also equipped with a matching [∗F∗] feature on its probe feature stack. Thus, if V has a top-most feature [•D•], it also has a probe feature [∗D∗]; discharge of [•D•] introduces the complement into the structure, and discharge of [∗D∗] (by Agree with the categorial feature of the complement, which can act as the goal even though it has also been previously used as the matching feature for structure-building) effects a special identification of the complement. The same pro-

---

[25] I am grateful to Hans van de Koot and Arnim von Stechow for discussion of this issue; also compare the remarks in footnote 21.

[26] However, the analogy would only hold if the ECP is assumed to be *conjunctive* rather than *disjunctive*. A disjunctive interpretation of the ECP used to be the standard one (see (5) in chapter 1, and Chomsky (1981; 1986a), and Lasnik & Saito (1984; 1992), among many others); however, a conjunctive version of the ECP is proposed by Stowell (1985, 27), Rizzi (1990, 32), and Hornstein & Weinberg (1995, 253 & 272ff); also see the references cited in the latter article.

cedure would then apply with N and P in (25): N bears [●P●], which is accompanied by [∗P∗].

As a third possibility, suppose that the edge of a phase not only contains specifiers, but also last-merged complements which, by definition, are topmost in their projections. If so, an edge feature need not be inserted on phase-heads for the movement of last-merged complements (but it still needs to be inserted for movement out of last-merged complements, and extremely local movement would still be required for complements in the presence of an inherent specifier of the phase head; see note 15).[27]

A fourth solution could be to adopt an even simpler concept of phase edge: On this view, an item α would be at the edge of a phase headed by γ iff α has been merged with γ. This would then imply that non-last-merged complements are also directly mobile, without requiring an extremely local movement step to a phase-internal specifier. As with the third option, movement from last-merged specifiers would continue to be blocked throughout, and movement from last-merged complements would depend on probe features on the phase head.[28]

All four options seem viable in principle, but they have different consequences. They also all raise potential problems. Consider the first two proposals first. If the general mobility *of* last-merged complements is ensured by assuming that there is still some probe features on the phase head left, the possibility to derive CED effects with movement *from* last-merged complements will be lost unless further assumptions are made: Illicit cases of extraction from (last-merged) complement DPs, illicit cases of extraction from (last-merged) complement CPs with non-bridge verbs, and related phenomena can now not be traced back to an inactive phase head because the DPs and CPs involved here can of course themselves undergo movement. However, later in the book (tentatively in the appendix to the present chapter, and then in a more assertive way towards the end of chapter 6) I will in fact argue that these types of analyses of extraction from complement DP and complement CP may have to be abandoned anyway, and may have to be replaced by a slightly different approach that postulates structural differences between the complement types. Under this assumption, the problem that I have just sketched disappears (especially for the first proposal).

---

[27] This option, while simple as such, requires a distinction between last-merged complements (i.e., items that are both last-merged and first-merged) and last-merged specifiers (i.e., items that are last-merged and non-first-merged), given the approach to English double object constructions in chapter 3: Last-merged right-peripheral complements are then part of the edge domain of a phase head; last-merged right-peripheral specifiers are not.

[28] Thanks to Stefan Keine (p.c.) for suggesting this alternative, and pursuing its consequences in some detail.

Let me next turn to potential problems with the third and fourth approaches, according to which a last-merged complement (but not an item in a last-merged complement's edge domain) is in the edge domain of the phase head. These approaches are directly compatible with the account of possible asymmetries between extraction from a last-merged complement and extraction of a last-merged complement in terms of the variable availability of probe features on the phase head. However, adopting such an approach will require a radically different approach to freezing effects from the one presented in the next section, where local movement of non-last-merged complements is required for the analysis of freezing to work.

Thus, we may ask whether there might be an approach that makes short movement of a last-merged complement to a local edge position systematically available and still captures variation in extraction from last-merged complements by tying it to the activity or inactivity of the phase head. Here is a suggestion.[29] Suppose that the requirement (11-b) in the Edge Feature Condition is replaced with (26).

(26) γ is active.
      A phase head γ is active if (a) or (b) hold.
      a. γ has not yet discharged all its structure-building or probe features.
      b. γ's sister bears an unchecked feature [F] (matching some [•F •]).

The addition in (26-b) is the important part of (26). It has the effect that activity of a phase head can be ensured not only by the phase head's intrinsic properties, but also by properties of its sister: A feature that keeps an XP first-merged with a phase head active also keeps the phase head active. The features in question are those that are looked for by some movement-inducing structure-building feature – e.g., [wh], [Σ], [top] (see chapter 3 and below).[30] From (26), the following consequences can be derived. First, only complements can keep a phase head active if it has no operation-inducing features left anymore (specifiers are not sisters of phase heads). Second, only those complements can keep a phase head active that need to be moved to the phase edge themselves. Thus, if a complement does not have to move itself, it will not be able to help an item in its edge domain; therefore, the above account of restrictions on extraction from last-merged complements can be maintained even though complements themselves are entirely mobile. (If both a last-merged complement and its specifier need to move, the phase head is still active; see the next section.) Finally,

---

[29] Similar assumptions are made in Assmann's (2010) analysis of parasitic gap constructions, for independent reasons.

[30] It is also worth pointing out that the approach to operator islands in terms of feature maraudage developed in the next chapter depends on a similar assumption about the activity of features targetted by movement-inducing features.

it needs to be guaranteed that an active complement cannot keep a phase head active once a last-merged specifier has been merged, much as in the case of probe features triggering Agree. This follows if it is assumed that edge-feature insertion and move-ment to the edge are carried out as quickly as possible (as seems natural); after the complement has been moved around the head, to a local specifier, the phase head does not have a sister with an unchecked feature anymore.

For the time being, I will conclude that all five approaches to the general mobility of last-merged complements envisaged here work, but there may be (different kinds of) compatibility problems with other analyses that may require (different kinds of) additional assumptions. In general, it seems that the first approach and the final ap-proach just sketched might work best. For the sake of concreteness, I will for now postulate that at least one of these two approaches is on the right track.[31]

That said, let me return to the main plot and sum up what has been said so far: The extended version of the present approach that takes into account probe features on phase heads makes some interesting, and largely correct, predictions. First, last-merged specifiers continue to be strict barriers. Second, last-merged complements are not barriers if the phase head has an additional probe feature for Agree with/into the complement; otherwise last-merged complements are barriers. However, there is also a third prediction that I have not yet mentioned. I address this in the following subsection.

3.2.6 Transparent Non-Last-Merged Specifiers and Complements

As it stands, the analysis predicts that non-last-merged specifiers (i.e., non-last-merged, non-first-merged items) and non-last-merged complements (i.e., non-last-merged, first-merged items) are not barriers, independently of whether the phase head has appropriate probe features. Radical as though it may be, this prediction actu-ally seems to be generally confirmed; see, e.g., the German examples in (27) (from Müller (1995), based on analogous Dutch data discussed in Koster (1987)). Here, a (non-argument) PP specifier of an unaccusative V is transparent for extraction (more specifically: topicalization) of an R-pronoun in the presence of a higher VP-internal specifier (the sole argument DP of the unaccusative verb); cf. (27-a). In contrast, the same (non-argument) PP specifier of an unergative V is a barrier for R-pronoun ex-traction (see (27-b)) because the nominative argument DP is merged in Specv rather than in SpecV, which implies that PP is last-merged in VP.[32]

---

[31] This deviates from Müller (2010b), where the third approach is ultimately adopted.

[32] The difference between unaccusative V and unergative V presupposed here is independently motivated by a variety of tests, an immediately obvious one being perfect auxiliary selection – *sein* ('be') vs. *haben* ('have'); see Grewendorf (1989).

(27) a. Da$_1$ ist [$_{VP}$ er [$_{V'}$ [$_{PP}$ t$_1$ mit] [$_{V'}$ [$_{PP}$ zum Doktor] gegangen]]] v
          there is   he   with   to the doctor   gone
          'He went to the doctor with it.'

      b. *Da$_1$ hat [$_{vP}$ er [$_{v'}$ [$_{VP}$ [$_{PP}$ t$_1$ mit] [$_{V'}$ das Rennen beendet]] v]]
          there has   he   with   the race   finished
          'He finished the race with it.'

Still, it is very well possible that this consequence may ultimately prove too far-reaching: For instance, certain accusative objects might be barriers despite the presence of a higher dative object. If so, this might be taken to suggest that the two objects do in fact show up in two distinct verbal projections, as has sometimes been argued.

4. Freezing

4.1 Freezing Effects

The approach developed so far predicts that last-merged specifiers that arise as a consequence of the lowest-ranked subcategorization feature on a phase head are barriers. At this point, the question arises of whether last-merged specifiers that come into existence because of a movement-inducing (inherent) edge feature on a phase head are also barriers – and if so, whether this can be derived within the general theory adopted here. As for the empirical part, it seems fair to conclude that there is disagreement in the literature concerning the question of whether moved items (that have reached their final position) are always barriers for extraction to a higher position (i.e., whether there is a general freezing effect; cf., e.g., Ross (1967), Wexler & Culicover (1980), Browning (1991), Collins (1994), and the references given in subsection 3.5 of chapter 2); or whether in at least some cases, a moved item may stay (or, indeed, become) transparent for extraction (see, e.g., Chomsky (1986a)). To give a few more specific examples: Whereas Cinque (1990) and Sternefeld (1991b) argue that SpecC is invariably a barrier (even if the item moved to this position is transparent in situ), Chomsky (1986a) and Lasnik & Saito (1992) assume that SpecC is transparent for extraction (even if the item in SpecC is a barrier in situ). Similarly, Browning (1991), Collins (1994), Chomsky (1995), Rezáč (2004), Broekhuis (2005), and Boeckx & Grohmann (2007) (among many others) postulate that SpecT is a barrier; but Chomsky (2008), Gallego & Uriagereka (2006) (see chapter 2), and Neeleman & van de Koot (2010) suggest that this is not (or not always) the case. As for SpecC, Chomsky (1986a) bases his assumption that SpecC is transparent on the Spanish data in (28-ab) (from unpublished work by Esther Torrego), where it looks as though movement from a *wh*-moved subject is possible even though the subject is a barrier in situ.[33]

---

[33] Here and in what follows, (*) characterizes an example whose grammaticality status is disputed. In the context of the present discussion, (*) in (28), (29), and (30) stands for (relative, in (29)) wellformedness.

(28) a. *Esta es la autora [$_{PP_1}$ de la que] [$_{DP_2}$ varias traducciones t_1] han
             this is the author   by whom   various translations   have
             ganado premios internacionales
             won   awards international

     b. (*)[$_{PP_1}$ De que autora] no sabes   [$_{CP}$ [$_{DP_2}$ que   traducciones t_1] han
             by what author   not you know   which translations   have
             ganado t_2 premios internacionales] ?
             won   awards international

Comparable judgments for analogous cases of extraction from SpecC in English are
provided by Lasnik & Saito (1992); see (29) ((29-a) = (ii) from footnote 42 in chapter
3).

(29) a. ??(*)Who$_1$ do you wonder [$_{DP_2}$ which picture of t_1] Mary bought t_2 ?
     b. ??(*)Who$_1$ do you wonder [$_{DP_2}$ which picture of t_1] t_2 is on sale ?

A set of examples used to argue for a transparency of SpecT in English given in Chom-
sky (2008) is (30-ab).

(30) a. *[$_{PP_1}$ Of which car] did [$_{DP_2}$ the driver t_1] [$_{vP}$ t_1 cause a scandal] ?
     b. (*)[$_{PP_1}$ Of which car] is [$_{DP_2}$ the driver t_1] likely [$_{TP}$ t'_2 to t_2 cause a scan-
         dal] ?

### 4.2 Predictions of the Analysis So Far

Notwithstanding the empirical issue of what the correct empirical generalization is,
the question can be raised what the approach developed so far predicts with respect to
the presence or absence of CED effects in the context of (regular, non-intermediate)
movement to specifier positions. As it stands, the prediction is clear enough: If nothing
else is said, it follows from the present system that freezing effects can be avoided
with complement movement (as in (29-a)), but not with movement of a specifier that
is last-merged in its base position. Here is why complement movement is so far not
predicted to create a freezing effect for extraction out of the complement: Suppose
that α is a complement of V, and in α's edge position is a category β. Both α and β
are needed by categories γ, δ that have the appropriate movement-inducing structure-
building features that recognize α and β (alternatively, some item γ is equipped with
both features); γ and δ are still part of the numeration on the VP cycle. Since inserting
edge features to the phase head V at this point will thus have an effect on outcome
(in the sense made precise in chapter 3; cf. (11-a) above), it is legitimate (if the phase
head is not yet inert at the stage where assignment of an edge feature must take place).
Consequently, β can move out of α to SpecV, and α can also move to SpecV; the two
movement steps can apply in any order, the only requirement being that the second
movement goes beyond the landing site of the first movement (because of the SCC,

which precludes tucking in).[34] From this point on, the two items move hand in hand (but separately), to higher specifiers, until they reach their respective ultimate landing sites (see Heck (2004) for extensive discussion of derivations of this type).

For last-merged specifiers, such a way of avoiding a barrier effect does not seem to exist, though. When such a specifier has been base-merged with a phase head, the phase head becomes inert, and no edge feature can be established on the phase head that might attract a specifier out of the last-merged specifier. Still, a closer look reveals that there is a bit more to be said about this case. Consider the abstract structure in (31), where ZP is a last-merged specifier and WP is in Z's edge domain. An edge feature cannot be instantiated on X for extraction of WP to SpecX; XP is then merged with Y, and Y will not be able to see WP in its complement XP (given non-recursive edges).

(31)
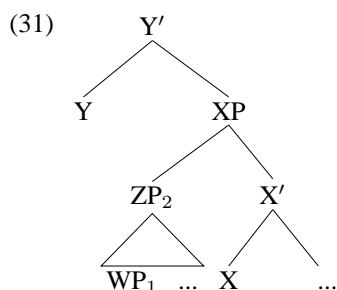
In this configuration, assuming that Y is still active, an edge feature can be inserted on Y, and movement to SpecY can be triggered, for ZP (which is still accessible) but not for WP (which is not accessible to Y anymore). Thus, in the next step, ZP may move to SpecY, as in (32).

(32)

---
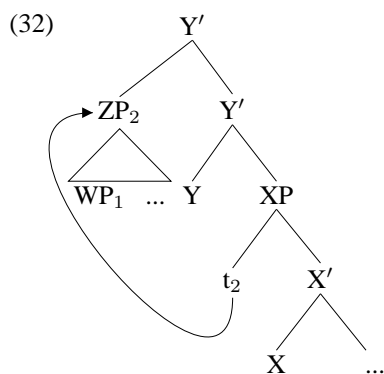
[34] Given two of the approaches to mobility of last-merged complements outlined above, α may not even have to undergo the first movement step if it is last-merged.

However, at this point it seems that WP_1 should be in the PIC-accessible domain of the phase head Y again, as a consequence of successful piggybacking. If so, an edge feature on Y could now be inserted for WP_1 movement (assuming, as before, that Y is active as such because it has another structure-building feature left that generates a specifier), and we would expect a freezing effect to be absent with movement of last-merged specifiers, too (much as suggested in Chomsky (1986a) and Lasnik & Saito (1992)). Still, there are various pecularities connected to such a derivation. Most notably, it would be unique in that the phase head Y has to be able to "change its mind" concerning the question of whether an edge feature can be inserted for an item that it shares a structure with. Here and henceforth, I will assume that this is not possible: If a phase head γ and an XP α are part of the same structure, and γ cannot receive an edge feature for α (according to the Edge Feature Condition), then this decision is irreversible throughout the remainder of the derivation. Under this assumption, we end up with the result that the present approach predicts freezing effects with last-merged specifiers but not with complements (and non-last-merged specifiers).

I do not take these predictions made by the present approach to be empirically correct (cf. chapters 1 & 2); therefore, additional assumptions are called for. The empirical problem is that there are many well-established, entirely uncontroversial freezing effects with extraction from complements. For instance, a topicalized VP is always a barrier blocking any kind of extraction ouf of VP in German; see (33) (from Müller (1998)): (33-a) illustrates VP topicalization; (33-b) illustrates ordinary *wh*-extraction from VP; and (33-c) shows that the two operations cannot be combined. Similarly, as observed by Postal (1972), a topicalized PP is a barrier for further extraction from PP in English; see (34): Postposition stranding becomes impossible under PP topicalization. And there are many more freezing effects of this type.

(33) a. Ich denke [$_{CP}$ [$_{VP}$ das Buch gelesen]$_2$ hat keiner t$_2$]
         I   think   the book read   has no-one

    b. [$_{DP}$ Was]$_1$ denkst du [$_{CP}$ t$'_1$ hat keiner [$_{VP}$ t$_1$ gelesen]$_2$] ?
         what   think you   has no-one   read

    c. *[$_{DP}$ Was]$_1$ denkst du [$_{CP}$ [$_{VP}$ t$_1$ gelesen]$_2$ hat keiner t$_2$] ?
         what   think you   gelesen   has no-one

(34) a. Who$_1$ do you think that he will talk [$_{PP_2}$ to t$_1$] ?
    b. *Who$_1$ do you think that [$_{PP_2}$ to t$_1$] he will talk t$_2$?

Furthermore, closer scrutiny suggests that the putative exceptions to freezing in (28), (29), and (30) may to a significant extent be misanalyzed, and may not involve extraction from a moved item after all (the derivations involving extraction indicated above would then be ungrammatical). Thus, the exceptional examples tend to involve *of*-phrases (or their equivalents in other languages), which are known to be independently available as optional arguments in many cases. For instance, this is the only possibility

in a German example like (35), which would otherwise have to involve long-distance scrambling from a finite clause in German (indicated here by "(*t_1)") – an operation that is known to be strongly excluded, with no variation whatsoever involved (see Ross (1967), Müller & Sternefeld (1993), Bayer & Kornfilt (1994), Grewendorf & Sabel (1994), and Fanselow (2001a), among others).

(35) dass ich [$_{PP_1}$ von Peter] glaube [$_{CP}$ dass [$_{DP_2}$ eine neue CD (*t_1)] erschienen
    that I   of Peter believe   that   a   new CD   appeared
    ist]
    is

A reanalysis of what may at first sight look like an instance of *of*-phrase movement from a moved item (with exceptional properties from the point of view of locality theory) as resulting from external Merge of an optional argument has been argued for with respect to constructions such as those in (28) and (30) by Koster (1987, 196f.), Cinque (1990, 47), Sternefeld (1991b, 121), Müller (1995, 397f.), Barbiers (2002, 54), and Gallego (2007, 349) (the list is probably not exhaustive). I assume that this reanalysis is essentially correct (and that, in addition, data like those in (29) can be classified as ill formed). If so, (36) can be maintained as a valid empirical generalization (also cf. Rizzi (2006)).[35]

(36) *Freezing Generalization*:
    A trace t must not be included in a moved XP (i.e., an XP that binds a trace) if the antecedent of t c-commands XP.

(36) is formulated in such a way that it does not block traces in moved items (with the antecedent outside of the moved item) as such: Remnant movement configurations, where the antecedent eventually shows up in a lower position than the trace included in a moved XP, are permitted; remnant movement systematically exhibits an anti-freezing effect.[36] Typical examples involving remnant VP topicalization in German

---

[35] Barbiers (2002, 49) discusses a Dutch construction that looks like a good candidate for a genuine counter-example to (36); see (i), where *wh*-extraction from PP strands P in the matrix clause.

(i) Waar$_1$ had jij dan [$_{PP_2}$ t_1 mee] gedacht [$_{CP}$ dat je de vis t_2 sou   moeten snijden] ?
    where had you then   with thought   that you the fish   would must   cut
    'What had you thought to be forced to cut the fish with?'

The account to be developed in this section will have nothing to say about the wellformedness if (i).

[36] A side remark: Nothing in this chapter forces a decision with respect to the question of whether movement leaves a trace, a full copy, or nothing at all. None of the primitive constraints adopted here (like the PIC, the EFC, or the SCC) refer to traces or copies; and where such items are referred to (as in the Freezing Generalization (36)), this is just for expository convenience. Thus, the approach is compatible with the conclusion in subsection 7.3 of chapter 3 that movement might simply leave nothing at all behind.

are given in (37) (cf. Thiersch (1985), den Besten & Webelhuth (1990), Grewendorf & Sabel (1994), and Müller (1998)): A VP from which some item (here, DP_1) has been scrambled subsequently undergoes topicalization, thereby producing an unbound trace.[37]

(37) a. $[_{VP_2}$ t_1 Zu lesen] hat $[_{DP_1}$ es] keiner t_2 versucht
             to   read   has it_{acc} no-one$_{nom}$ tried

     b. $[_{VP_2}$ t_1 Gelesen] hat der Fritz $[_{DP_1}$ das Buch] nicht
               read   has the Fritz$_{nom}$ the book$_{acc}$ not

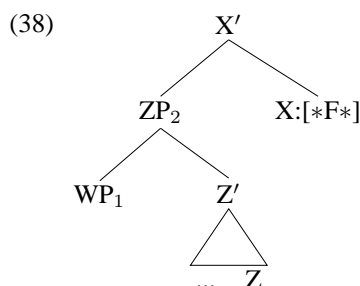### 4.3   Deriving the Freezing Generalization

As a first step towards deriving (36), let me make explicit some assumptions about movement-inducing features. First, given that there are only two feature stacks associated with lexical items (viz., a stack of probe features and a stack of structure-building features), [●F●] features that trigger internal Merge are on the same stack as [●F●] features that trigger external Merge. However, the former must be lower on the hierarchy – they become accessible only after all subcategorization features of a lexical item have been discharged. Next, I assume, as before (see chapters 1 and 3), that movement-inducing features include [●wh●] (on C or some related functional head of the left periphery, for *wh*-movement); [●top●] (on C or some related functional head of the left periphery, for topicalization); [●Σ●] (on V, for highly local scrambling to SpecV, or on v, for scrambling to Specv; see Müller (1998), Grewendorf & Sabel (1999) and Sauerland (1999), among others);[38] and [●D●] (on T, for EPP-movement to SpecT). Against this background, I would like to suggest that multiple edge feature insertion follows a feature hierarchy going back to Williams (1974).

For concreteness, consider a derivation that has reached the stage indicated in (38), with WP a specifier of Z and ZP a complement of X (e.g., WP could be a DP of VP, and VP (= ZP) could be a complement of v, showing up to the left of v in German; but (38) is more general, and the linearization options indicated here, although modelled on the concrete case just mentioned, are irrelevant for the discussion that follows). Suppose that both WP_1 and ZP_2 are phrases that will eventually (in the course of the derivation) be needed by movement-inducing features [●F●] on lexical items that, at this point, are

---

[37] This analysis implies that restructuring (or 'coherent') infinitives as in (37-a) are bare VPs. See, e.g., Wurmbrand (2001). Nothing depends on this issue in the present context.

[38] To simplify matters, I assume here (as I did in chapter 3) that moved unstressed pronouns (i.e., 'object-shifted pronouns') and scrambled items both occupy a Specv position in German, triggered by [●Σ●] on v. Since the two movement types behave differently in a number of respects (e.g., with respect to obligatoriness, order preservation, and interaction with optional subject raising triggered by an optional EPP feature on T in German), this may perhaps ultimately not prove correct. However, this question does not affect the issues currently under consideration.

still part of the numeration. Then, assuming that last-merged complements are not in the edge domain of a phase (see section 3.2.5), X has to be provided with edge features that attract *both* WP$_1$ and ZP$_2$ to SpecX. This presupposes that X in (38) either has an undischarged structure-building feature or an undischarged probe feature left. Let us assume the latter (as is indicated in (38): [∗F∗]).[39]

(38)

```
                    X′
                   /  \
                  /    \
              ZP₂      X:[∗F∗]
             /  \
            /    \
         WP₁     Z′
                 /\
                /  \
               /____\
              ...   Z
```

An important assumption that I would like to put forward at this point is that the assignment of the required edge features to X does not proceed in an unconstrained way; rather, there is a fixed order of edge feature insertion which follows a feature hierarchy that mirrors the typical hierarchy in phrase structures: [•Σ•] ≫ [•D•] ≫ [•top•] ≫ [•wh•] (cf. the reverse order SpecC(wh) > SpecC(top) > SpecT > Specv, SpecV in clausal trees). Thus, if WP$_1$ in (38) is required by a lexical item (in the numeration) with a structure-building feature [•F$_i$•], and ZP$_2$ in (38) is required by a lexical item (in the numeration) with a structure-building feature [•F$_j$•], and [•F$_i$•] ≫ [•F$_j$•], then the corresponding edge feature for F$_i$ must be established first on X, and F$_j$ is assigned to X later (both in accordance with the EFC). Arguably, the simplest way to execute this idea is to give up the assumption that edge features assigned to phase heads in accordance with the EFC are category-neutral; thus, in what follows I assume that the edge feature assigned to a phase head identifies the category on which it is based. Still, category-sensitive edge features should presumably be kept distinct from movement-inducing features that are inherent to lexical items, so as to ensure that, e.g., there is no *wh*-movement from a target *wh*-position associated with an interrogative C.[40] A suitable notational convention might be to render the category-

---

[39] See footnote 42 below on whether things might be slightly different if there were a structure-building feature on X instead. Also note that the present discussion presupposes the first approach to the mobility of last-merged complements developed in section 3.2.5. If one adopts the fifth approach based on the extended concept of phase head activity in (26), [∗F∗] would in fact not need to be present; but the considerations that follow in the main text will be otherwise identical.

[40] But see Reis & Rosengren (1992) for a possible exception in German that I will come back to in chapter 5. In addition, it is arguably the case that accounts of effects like the ones at hand would still be available if

sensitive edge features that are inserted in the course of the derivation in accordance with the EFC as $[\bullet X_{wh}\bullet]$, $[\bullet X_{top}\bullet]$, $[\bullet X_{\Sigma}\bullet]$, etc. Suppose further that there is an additional condition stating that an edge feature cannot be inserted for a lower category WP_1 after the insertion of an edge feature for a higher, more inclusive category ZP_2 has been triggered. The two conditions are summarized in (39).[41]

(39) *Restrictions on multiple edge feature assignment*:
A categorially marked edge feature $[\bullet F_j\bullet]$ cannot be assigned to a phase head γ for a category β if (a) or (b) hold.

    a. An edge feature $[\bullet F_i\bullet]$ also needs to be assigned to γ, and $[\bullet F_j\bullet]$ corresponds to a structurally higher position than $[\bullet F_i\bullet]$.

    b. An edge feature $[\bullet F_k\bullet]$ has already been assigned to γ for a category α, and α includes β.

Thus, (39-a) predicts that multiple assignment of categorially marked edge features $[\bullet F_i\bullet]$, $[\bullet F_j\bullet]$ to a single phase head respects the hierarchy $[\bullet F_i\bullet] \gg [\bullet F_j\bullet]$, where $[\bullet F_j\bullet]$ is intrinsically associated with a higher position in the clausal structure than $[\bullet F_i\bullet]$.

Returning to (38), assume first that WP_1 is an item that qualifies as an F_i-element (that will eventually be attracted by a lexical item bearing $[\bullet F_i\bullet]$, which so far is still part of the numeration); that ZP_2 is an item that qualifies as an F_j-element (that will eventually be attracted by a lexical item bearing $[\bullet F_j\bullet]$); and that $[\bullet F_i\bullet] \gg [\bullet F_j\bullet]$. The derivation of (38) can then be legitimately continued as in (40): Based on X's activity (due to the probe feature $[*F*]$, which is not yet discharged), the edge feature $[\bullet F_i\bullet]$ is first added to X; see (40-a). Assuming that feature checking obeys an Earliness requirement (see Pesetsky (1989; 2000), Pesetsky & Torrego (2001), Chomsky (2001b)), $[*F*]$ is then discharged immediately in the next step, undergoing Agree with some goal in WP; see (40-b). Next, $[\bullet F_j\bullet]$ is assigned to X, ending up on top of the stack of structure-building features; see (40-c). This step is unproblematic because X is still active at this point, having received an edge feature in the last but one step (discharging $[\bullet F_i\bullet]$ would have made this insertion impossible). In the fourth step, $[\bullet F_j\bullet]$ is discharged, moving the F_j-element ZP_2 to a specifier of X (an instance of

---

such a distinction is not made.

[41] See Grewendorf (2003; 2004) and Abels (2008) for analyses of remnant movement asymmetries that incorporate a condition like (39-a) (and that either are, or could easily be made, compatible with the present approach); and Williams (1974; 2003) and Sternefeld (1992) for the origins of the condition, which apparently was informally referred to as the 'Williams cycle' in Chomsky's 1974 MIT class lectures. Also note that (39-b) would follow from a minimal extension of the Strict Cycle Condition in (23). Thus, it is not inconceivable that (39-a) and (39-b) will eventually both be derivable from a more general concept of cyclicity. For reasons of space and coherence, I will not pursue these issues here, though.

string-vacuous movement that is needed nevertheless because it is the only way to satisfy the PIC on the cycle defined by the next phase head, given that ZP_2 is not part of the edge of X in the presence of an additional specifier – recall the discussion on pp. 183-186); see (40-d). Finally, $[\bullet F_i \bullet]$ is discharged by moving the F_i-element WP_1 out of ZP_2, to an outer specifier. As a consequence, the phase head has lost all its operation-inducing features, and the phase is completed; see (40-e).

(40) $[_{X'} [_{ZP_2} WP_1 [_{Z'} ... Z]] X:[*F*]]$

    a. $[_{X'} [_{ZP_2} WP_1 [_{Z'} ... Z]] X:[*F*], [\bullet F_i \bullet]]$

    b. $[_{X'} [_{ZP_2} WP_1 [_{Z'} ... Z]] X:[\bullet F_i \bullet]]$

    c. $[_{X'} [_{ZP_2} WP_1 [_{Z'} ... Z]] X:[\bullet F_j \bullet] \succ [\bullet F_i \bullet]]$

    d. $[_{X'} [_{ZP_2} WP_1 [_{Z'} ... Z]] [_{X'} t_2 X:[\bullet F_i \bullet]]]$

    e. $[_{XP} WP_1 [_{X'} [_{ZP_2} t_1 [_{Z'} ... Z]] [_{X'} t_2 X]]]$

This derivation respects all conditions, including the two restrictions on multiple edge feature assignment in (39-a) and (39-b). The continuation of this derivation is such that WP_1 and ZP_2 move together through the next cycles (with edge feature assignment for WP_1, viz., $[\bullet F_i \bullet]$, preceding edge feature assignment for ZP_2, viz., $[\bullet F_j \bullet]$ on each phase head), until the ultimate landing sites associated with the lexical items are reached that bear the pertinent structure-building features as an inherent property. Given that $[\bullet F_i \bullet]$ corresponds to a lower landing site in the phrase structure tree than $[\bullet F_j \bullet]$, derivations of this type will create licit remnant movement configurations such as those in (37) (where, e.g., $[\bullet F_i \bullet]$ is $[\bullet X_\Sigma \bullet]$, and $[\bullet F_j \bullet]$ is $[\bullet X_{top} \bullet]$, the F_i-element WP_1 is DP_1, and the F_j-element ZP_2 is VP_2).

    In contrast to this, consider the ill-formed continuation of (38) in (41). Suppose that the only relevant difference is that this time, ZP_2 is the F_i-element (i.e., the item that must be moved to the specifier of the phase so as to provide an accessible element for subsequent movement to the specifier of a lexical item inherently bearing $[\bullet F_i \bullet]$), and WP_1 is the F_j-element. In this case, a dilemma arises because (39-a) and (39-b) cannot both be fulfilled. (39-a) demands that $[\bullet F_i \bullet]$ is inserted as the next step in (41), as in (40); see (41-a) (if $[\bullet F_j \bullet]$ is assigned first, the derivation breaks down immediately). Probe feature discharge in (41-b) is unproblematic. However, any continuation of (41-b) will now run into problems. It is clear that $[\bullet F_j \bullet]$ would have to be added next, and that this feature would have to go the top of the feature stack on X (given the EFC), as in (41-c). However, given that $[\bullet F_j \bullet]$ corresponds to WP_1 in this case (which is the F_j-element), and WP_1 is included in ZP_2 at this point, such an assignment of an edge feature invariably violates (39-b). Consequently, $[\bullet F_j \bullet]$ cannot be assigned to X, and WP_1 cannot move out of ZP_2 on the XP cycle. Furthermore, given that edges are defined non-recursively (cf. footnote 12), such that the edge β of an edge α of a phase Γ is not part of the edge of Γ; and given that a phase head cannot change its mind about the accessibility of an item that has co-occurred with it in the same structure

(and has been declared as inaccessible then – recall the discussion of (32)) – given these two assumptions, the PIC ensures that WP_1 will not be able to move out of ZP_2 at any subsequent step of the derivation either.

(41) $[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{*}F{*}]\ ]$

    a. $[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{*}F{*}],\ [{\bullet}F_i{\bullet}]\ ]$

    b. $[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{\bullet}F_i{\bullet}]\ ]$

    c. *$[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{\bullet}F_j{\bullet}] \succ [{\bullet}F_i{\bullet}]\ ]$

Thus, freezing effects like those in (30) are accounted for.[42] Here, $[{\bullet}F_i{\bullet}]$ is $[{\bullet}X_D{\bullet}]$ (i.e., the edge feature for the EPP feature of T), and $[{\bullet}F_j{\bullet}]$ is $[{\bullet}X_{wh}{\bullet}]$ (the edge feature for a $[{\bullet}wh{\bullet}]$ feature on interrogative C); the F_i-element ZP_2 is DP_2, and the F_j-element WP_1 is PP_1 (PP_1 must have undergone edge-feature driven movement to the edge of DP_2 first because of the PIC). Similar reasonings apply to all instances of freezing effects where two different movement types are involved (e.g., (28-a) in Spanish, if raising is involved here). For freezing effects where two identical structure-building features occur (as in (28-b) in Spanish – with the the structure as indicated, i.e., without an optional matrix argument interpretation – and (29) in English), a minimal extension of what has been suggested so far is required: It suffices to assume that

---

[42] However, closer scrutiny reveals that there may still be a potential loophole, and a bit more might have to be said about other kinds of examples involving freezing. The preceding reasoning is based on the assumption that X does not have any structure-building features left in (38). However, suppose that it did have some such feature left, as shown in (i).

(i)
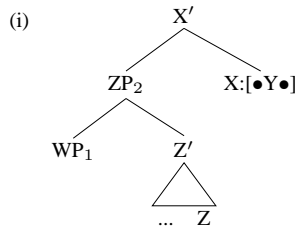
Now consider the following derivation which differs minimally, but crucially, from the one shown in (41). As before, $[{\bullet}F_i{\bullet}]$ (corresponding to the lower target position) is assigned first, but it is then *discharged immediately* in the next step, triggering movement of ZP_2 to a specifier (this is possible because X is still active after this discharge, because of $[{\bullet}Y{\bullet}]$). Then, $[{\bullet}F_j{\bullet}]$ is added (as in (i-c)), and discharged again. Finally, $[{\bullet}Y{\bullet}]$ is discharged.

(ii) $[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{\bullet}Y{\bullet}]\ ]$

    a. $[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{\bullet}F_i{\bullet}] \succ [{\bullet}Y{\bullet}]\ ]$

    b. *$[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{\bullet}Y{\bullet}]\ ]$

    c. $[_{X'}\ [_{ZP_2}\ WP_1\ [_{Z'}\ ...\ Z\ ]]\ X{:}[{\bullet}F_j{\bullet}] \succ [{\bullet}Y{\bullet}]\ ]$

In order to rule out this derivation (which evades (39-b) if it is assumed that no record is kept of formerly

there is an addition to (39-a) such that in cases of identical structure-building features, the higher, more inclusive category must be targetted first; compare the modification of (39) in (42) with the original formulation.

(42) *Restrictions on multiple edge feature assignment* (extended):

A categorially marked edge feature $[\bullet F_j \bullet]$ cannot be assigned to a phase head γ for a category β if (a) or (b) hold.

    a. An edge feature $[\bullet F_i \bullet]$ also needs to be assigned to γ, and $[\bullet F_j \bullet]$ corresponds to a structurally higher position than $[\bullet F_i \bullet]$.

       If two instances of $[\bullet F_i \bullet]$ need to be assigned to γ, edge feature assignment for the higher category must take place first.

    b. An edge feature $[\bullet F_k \bullet]$ has already been assigned to γ for a category α, and α includes β.

Insertion of an identical feature for a second category dominated by the first one is then always impossible, because of (42-b).[43]

    Under these assumptions, the present approach does not simply draw the line between freezing effects with bound traces, and an absence of freezing effects with unbound traces (i.e., remnant movement). Remnant movement constructions of the type in (43-ab) (= (62-a), (62-b) in chapter 1), which were grouped under the label of Unambiguous Domination effects, are also correctly excluded. These examples differ minimally from those in (37) in that VP_2 and DP_1 are moved because of the same structure-building feature (viz., $[\bullet \Sigma \bullet]$). This consequence is particularly relevant because whereas these kinds of effects could be shown to be derivable from the (G)MLC (as F-over-F condition effects), the account of (G)MLC effects in terms of the PIC developed in chapter 3 did not actually have anything to say about them, and necessitated the postulation of a Residual F-over-F Condition (RFOF; see (49) in chapter 3). The RFOF can now be dispensed with.

(43) a. *dass $[_\alpha$ t$_1$ zu lesen $]_3$ $[_{DP}$ das Buch $]_1$ keiner t$_3$ versucht hat
          that   to read   the book$_{acc}$ no-one   tried   has

    b. *dass der   Frank $[_\alpha$ t$_1$ t$_2$ zu reparieren $]_3$ dem Matthias$_1$ den Drucker$_2$ t$_3$
       that ART Frank   to fix   ART Matthias$_{dat}$ the printer$_{acc}$

---

assigned features on phase heads), it can be assumed that edge features required by the PIC have to be inserted in toto before they can be discharged; thus the step in (ii-b) is illegitimate (as indicated here). Thanks to Silke Fischer (p.c.) for pointing out the problem and sketching the solution.

[43] On the other hand, multiple assignment of categorially identical edge features as such is unproblematic, as long as the two items for which the features are inserted do not stand in a dominance relation. This is required for cases of multiple *wh*-movement, or multiple scrambling that is not extremely local, etc.; also cf. section 6.2.1 in chapter 5.

versprochen hat
promised   has

At this point, let me come back to the question to what extent the analyses of possible asymmetries between extraction of a last-merged complement and extraction from a last-merged complement discussed in section 3.2.5 are compatible with the present approach to freezing. The first thing to note is that the third and fourth approaches (according to which a last-merged complement is in the edge domain of a phase head) would require massive extra assumptions. The reason is that the approach to freezing based on restrictions on edge feature insertion as in (42) can only work if *both* ZP_2 and WP_1 are required to move in a structure like (38) even in cases where ZP_2 is a last-merged complement. As for the first, second, and fifth approaches (according to which a last-merged must move to a specifier position in the same phase, and an edge feature must be present), they are fully compatible with the approach to freezing effects developed here.

To sum up this section, it has turned out that whereas the present approach (which centers around the assumption that edge feature insertion is possible only as long as a head can in principle drive operations, and has not yet become inert) fully succeeds in deriving CED effects with externally (last-) merged specifiers from the PIC, a bit more must be said to ensure that all internally (last-) merged specifiers are barriers, too (unless certain kinds of remnant movement are involved). To achieve this result, I have adopted the two restrictions on multiple edge feature insertion in (42), both of which seem natural; moreover, both restrictions can arguably be viewed as conservative extensions of the Strict Cycle Condition (see footnote 41).

Interestingly, the present approach may have to depend on additional assumptions to derive freezing, but it automatically (and in the simple, non-extended form sketched in sections 2-3) predicts a reverse phenomenon that I call *melting*.

5.   Melting

There is a suprising further effect that is predicted under the system of assumptions laid out in sections 2-3: A specifier α of a phase Γ that is normally (i.e., if it is last-merged in its projection) an island for extraction should "melt", i.e., cease to be a barrier when some β becomes an outer specifier of Γ by movement to a higher position within the same phase. To the best of my knowledge, melting has not yet been discussed as a natural class of phenomena in the literature; but the effect seems real enough.

A melting effect can be observed with highly local movement (viz., scrambling) in front of base-merged argument DPs in languages like German and Czech.

5.1 Melting Effects with Scrambling in German

Consider first melting effects with scrambling in German. In (44-ab), it is shown that one and the same subject DP is a barrier for *was für* ('what for') split if it precedes an object DP, but becomes transparent for extraction of *was* if the object DP undergoes local scrambling in front of the subject.[44]

(44) a. *Was$_1$ haben [$_{DP_3}$ t$_1$ für Bücher] [$_{DP_2}$ den Fritz]   beeindruckt ?
      what have   for books$_{nom}$ the Fritz$_{acc}$ impressed

    b. Was$_1$ haben [$_{DP_2}$ den Fritz] [$_{DP_3}$ t$_1$ für Bücher] t$_2$ beeindruckt ?
       what have   the Fritz$_{acc}$ for books$_{nom}$ impressed

The same kind of melting effect with subject DPs created by local object scrambling to a position in front of the subject shows up with extraction of a PP; see (45-a) (with subject and object both in situ) vs. (45-b) (with local object scrambling).

(45) a. *[$_{PP_1}$ Über wen] hat [$_{DP_3}$ ein Buch t$_1$] [$_{DP_2}$ den Fritz]   beeindruckt ?
      about whom has   a book$_{nom}$ the Fritz$_{acc}$ impressed

    b. [$_{PP_1}$ Über wen] hat [$_{DP_2}$ den Fritz] [$_{DP_3}$ ein Buch t$_1$] t$_2$ beeindruckt ?
      about whom has   the Fritz$_{acc}$ a book$_{nom}$ impressed

The empirical evidence seems fairly uncontroversial. The data have been checked with a number of (linguist) native speakers; and, as mentioned above, examples of the type in (44-b) and (45-b) have figured prominently in approaches that strive to classify subject DPs as transparent for extraction in German even in constructions which are neither unaccusative nor passive (see footnote 17).[45]

---

[44] As shown by Grewendorf (1989) on the basis of a number of tests, a psych verb like *beeindrucken* ('impress') takes a regular external argument in German; i.e., the nominative DP here is never VP-internal. (As a matter of fact, one of the ten or so tests employed in that work is based on extraction from DP, as in (44-a); see Grewendorf (1989, 182).) However, it may be worth pointing out that the pattern in (44) would be identical if *Bücher* ('books') were to be replaced with *Leute* ('people'), and *beeindruckt* ('impressed') with the agentive, non-psych verb *getroffen* ('met').

[45] Against the background of Müller (2010b), Jurka (2010, 98-103) reports an experimental study (based on a questionnaire distributed among non-linguist speakers) of melting effects in German. This investigation could not corroborate a significant melting effect; only a subset of the speakers systematically produced the melting pattern in a way that would be statistically relevant. As noted in footnote 37 of chapter 2, I believe there is every reason to be suspicious of these kinds of statistical results in the domain of highly complex syntactic structures, especially if the subjects are non-linguists. In the case at hand, there do indeed appear to be various problematic aspects with the empirical material that was tested: First, all the sentences involved *was für* split, none PP extraction; but the latter operation is arguably much more representative of extraction than *was für* split, which is known to yield more murky (or subtle) judgements as a tendency, and which also differs from PP extraction from DPs in other ways (see Müller (1995, 66-67)). Second, it seems that the empirical material tested by Jurka is not of exactly the same type as the material presented in this chapter in order to exemplify the melting effect: The melting example in Jurka (2010, 100) has the scrambled object separated from the in-situ subject by both a temporal adverb and (a complex) locative adverbial phrase;

The analysis of melting effects is straightforward. First, recall that I have assumed that scrambling is feature-driven: It is an operation triggered by designated structure-building features [•Σ•] on V or v; only the latter option is relevant in cases of scrambling to a position in front of agentive subject arguments in situ, i.e., Specv. Second, suppose that v still has a movement-inducing feature [•Σ•] left on the stack of structure-building features *after* discharge of its final subcategorization feature [•D•] (which has triggered external Merge of the subject DP). Thus, the phase head v is still active at this point of the derivation, and an edge feature [•X•] (or, given the reasoning in the previous section, [•X$_{wh}$•]) can be inserted in accordance with the EFC before the phase is complete. Third and finally, this implies that the phase head v can now attract an item out of the external argument in the inner Specv position to an outer specifier, and thereby ensure that this item is accessible for further movement as part of the edge of the phase (given the PIC). As in the derivations discussed in section 3 (cf. (16) in particular), it follows from the phase heads' last-in/first-out property that the item extracted out of the subject must land in a specifier position that is lower than the position of the scrambled object (but, of course, within the same XP), as an instance of the Intermediate Step Corollary (15). The relevant part of the derivation (focussing on the vP domain) is shown for a melting example like (45-b) in (46). Here DP$_2$ is the scrambled object argument, and DP$_3$ is the subject argument in its base position.

---

this configuration does not exist in the non-melting context. This mismatch may have been required by the experimental design but clearly represents a confounding factor in that parsing difficulties are obviously increased in the melting condition; to properly judge the relevant examples, this factor would have to be actively suppressed, which is something that the experimental design does not allow for (and which non-linguists are presumably not very good at anyway). Finally, as Jurka himself acknowledges, OS word order (as required for the melting effect to emerge) vs. SO word order is in itself a potentially confounding factor that was not controlled in the experiment (and that could only be controlled by massively complicating the experimental design and providing natural contexts *exclusively* favouring one or the other order, which, as Jurka notes, is inherently extremely difficult, perhaps even impossible). Thus, Jurka (2010, 103) concludes that "there might still be some reality to" the melting effect, which might be experimentally detected "under improved methodological conditions". Throughout the remainder of this book I will continue to assume that the effect is indeed very real.

(46) PP$_1$ ...   vP
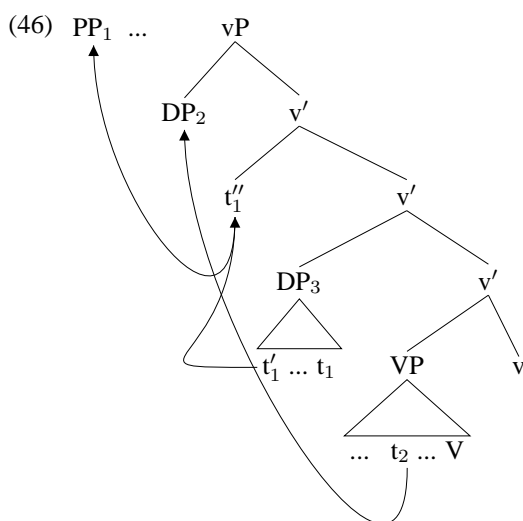
(47) illustrates schematically how the stack of structure-building features associated with a phase head γ in a melting derivation changes, and is eventually reduced to zero. As before, $[\bullet\beta\bullet]$ creates the complement of the phase head, and $[\bullet\alpha\bullet]$ creates the specifier. However, by assumption, the phase head also has a movement-inducing feature $[\bullet\delta\bullet]$ (the scrambling feature $[\bullet\Sigma\bullet]$, in the case just discussed – also recall from the previous section that features that trigger internal Merge are initially ranked below features that trigger external Merge on a lexical item), so that an edge feature $[\bullet X\bullet]$ (i.e., $[\bullet X_{wh}\bullet]$, in the case at hand) can be assigned to the phase head that extracts an item out of the specifier without inducing a CED effect.

(47) *How the melting effect is derived*:

$$
\begin{array}{l}
\gamma\text{: } [\bullet\beta\bullet] \;\succ\; [\bullet\alpha\bullet] \;\succ\; [\bullet\delta\bullet] \\
\rightarrow \gamma\text{: } [\bullet\alpha\bullet] \;\succ\; [\bullet\delta\bullet] \\
\rightarrow \gamma\text{: } [\bullet\delta\bullet] \\
\rightarrow \gamma\text{: } [\bullet X\bullet] \;\succ\; [\bullet\delta\bullet] \\
\rightarrow \gamma\text{: } [\bullet\delta\bullet] \\
\rightarrow \gamma\text{: } \varnothing
\end{array}
$$

$\rightsquigarrow$ violates nothing

The important difference to typical CED effects with subject DPs is that in (47), the feature that introduces the subject DP (viz., $[\bullet\alpha\bullet]$) is not the lowest-ranked feature on the stack of structure-building features; so the phase head is still active once this feature is discharged, and can be targeted by edge feature insertion, which may then extract some item out of the subject DP's edge domain.

Given this approach, we expect that the same effect should also arise in double object constructions: A nominative DP externally merged in Spec$v$ should undergo

melting and become transparent for extraction if a dative DP (indirect object) is scrambled in front of it, even if the accusative DP (direct object) stays in situ, in the VP. As shown in (48-a) vs. (48-b), this prediction is borne out.

(48) a. ?*Was$_1$ haben [$_{DP_3}$ t$_1$ für Leute]   [$_{DP_2}$ dem Fritz]   [$_{DP_4}$ Bücher]
          what have   for people$_{nom}$ the Fritz$_{dat}$ books$_{acc}$
          geschenkt ?
          given

      b.   Was$_1$ haben [$_{DP_2}$ dem Fritz]   [$_{DP_3}$ t$_1$ für Leute]   t$_2$ [$_{DP_4}$ Bücher]
          what have   the Fritz$_{dat}$ for people$_{nom}$ books$_{acc}$
          geschenkt ?
          given

(48-b) is particularly interesting because it shows that the melting effect is not an *adjacency* effect – as a result of indirect object scrambling to Specv in (48-b), the subject and the verb do not become adjacent because the direct object still intervenes. Thus, V adjacency may sometimes be a side effect of melting, but it is irrelevant as such.

Finally, the present analysis also predicts that local (VP-internal) scrambling of the direct object to a position in front of the indirect object should give rise to melting of the indirect object; as shown in (49-a) (with the direct object in situ) vs. (49-b) (with a locally scrambled direct object), this is the case.

(49) a. *Was$_1$ hat er   [$_{DP_3}$ t$_1$ für Leuten] [$_{DP_2}$ den Fritz]   vorgestellt ?
          what has he$_{nom}$ for people$_{dat}$ the Fritz$_{acc}$ introduced

      b.   Was$_1$ hat er   [$_{DP_2}$ den Fritz] [$_{DP_3}$ t$_1$ für Leuten] t$_2$ vorgestellt ?
          what has he$_{nom}$ the Fritz$_{acc}$ for people$_{dat}$ introduced

To sum up so far, it is correctly predicted that local scrambling to a position in front of a specifier that would otherwise be last-merged in its phase creates a melting effect in German: The specifier becomes transparent for extraction.

5.2 Melting Effects with Scrambling in Czech

The same phenomenon shows up in Czech; it can be seen very clearly in DP split constructions (see Kučerová (2006) and Fanselow & Lenertová (2011) for recent discussion), but also with other instances of movement from DP. The analysis that suggests itself for this effect is essentially identical to that given for German (assuming that Czech scrambling of the type shown here targets Specv); the different position of the verb is irrelevant.[46]

---

[46] Data and original judgements in this subsection are due to Petr Biskup and Denisa Lenertová; the judgements have been confirmed by Pavel Caha and Ivona Kučerová.

Consider DP split constructions first. (50-ab) show that whereas the subject DP *žádná stará* ('the old (one)', with a missing noun) is a barrier for extraction of the adjective *stará*, it becomes transparent when the accusative object *Petra* ('Peter') is locally scrambled in front of it; (50-b) may not be fully well formed, but it is perceived as a drastic improvement over (50-a).[47]

(50) a. *Stará$_1$ neuděřila [$_{DP_3}$ žádná t$_1$] Petra$_2$
          old$_{nom}$ hit   no$_{nom}$ Petr$_{acc}$
          'No old one hit Petr.'
     b. (?)Stará$_1$ neuděřila Petra$_2$ [$_{DP_3}$ žádná t$_1$] t$_2$
          old$_{nom}$ hit   Petr$_{acc}$ no$_{nom}$
          'No old one hit Petr.'

The same kind of melting effect with splits taking place from subject DPs can be seen in (51). In (51-a), *holka* ('girl') is moved from a subject DP, stranding *žádná* ('no'), while both subject and object are in their canonical positions. In (51-b), local object scrambling makes the subject transparent for extraction; the result has the same status as instances of extraction from the object DP itself; see (51-c).[48]

(51) a. *[$_{NP_1}$ Holka] neuděřila [$_{DP_3}$ žádná t$_1$] Petra$_2$
          girl$_{nom}$ hit   no$_{nom}$ Petr$_{acc}$
          'No girl hit Petr.'
     b. [$_{NP_1}$ Holka] neuděřila Petra$_2$ [$_{DP_3}$ žádná t$_1$] t$_2$
          girl$_{nom}$ hit   Petr$_{acc}$ no$_{nom}$
          'No girl hit Petr.'
     c. [$_{NP_1}$ Holku] neuděřil Petr$_2$ [$_{DP_3}$ žádnou t$_1$]
          girl$_{acc}$ hit   Petr$_{nom}$ no$_{acc}$
          'Petr hit no girl.'

Turning next to cases of PP extraction from DP (as an instance of topicalization), the examples in (52) show the same kind of subject DP melting that we have just seen with DP split constructions. (52-c) involves extraction from an object DP, which is possible; (52-a) illustrates the fact that extraction from a subject DP is not possible if both subject and object occupy their base positions; and (52-b) is again a case of

---

[47] I assume here that DP split constructions in Czech are derived by regular movement of the displaced item, not by some combination of copy movement plus deletion (as, e.g., in Fanselow & Ćavar (2001)), or by base-generation. The very fact that CED effects can be noted here would seem to lend strong support to this view (in contrast to what we have seen with DP split constructions in German; see footnote 17).

[48] Based on work by Beck (1997) and Pesetsky (2000), Kučerová (2006) identifies intervention and anti-intervention effects in DP split constructions in Czech. All the examples in this subsection are constructed in such a way (e.g., by avoiding certain verb placements) that this potentially interfering issue does not arise.

melting: The subject DP ceases to be a barrier for extraction when the object DP is scrambled in front of it.

(52) a. *[$_{PP_1}$ O   starých autech] oslovila   [$_{DP_3}$ kniha t$_1$] Petra$_2$
              about old   cars   fascinated   book$_{nom}$ Petr$_{acc}$
         'A book about old cars fascinated Petr.'

   b. (?)[$_{PP_1}$ O   starých autech] oslovila   Petra$_2$ [$_{DP_3}$ kniha t$_1$] t$_2$
              about old   cars   fascinated Petr$_{acc}$ book$_{nom}$
         'A book about old cars fascinated Petr.'

   c.   [$_{PP_1}$ O   starých autech] četl Petr$_2$ [$_{DP_3}$ knihu t$_1$]
              about old   cars   read Petr$_{nom}$ book$_{acc}$
         'Petr read a book about old cars.'

Finally, the same pattern emerges with *wh*-movement from DP. As is well known, such extraction may violate the Left Branch Condition in various Slavic languages; I will have nothing to say about this here. Extraction of *čí* ('whose') from an object DP is possible (see (53-cd), with variable positioning of the verb); in contrast, extraction of *čí* from a subject DP is impossible (see (53-a)) – unless, that is, the object is scrambled across the subject, as in (53-b).[49]

(53) a. *Čí$_1$ udeřila [$_{DP_3}$ t$_1$ sestra]   Petra$_2$ t$_2$?
         whose hit   sister$_{nom}$ Petr$_{acc}$
         'Whose sister hit Petr?'

   b. (*)Čí$_1$ udeřila Petra$_2$ [$_{DP_3}$ t$_1$ sestra]   t$_2$?
         whose hit   Petr$_{acc}$ sister$_{nom}$
         'Whose sister hit Petr?

   c.   Čí$_1$ udeřil Petr   [$_{DP_3}$ t$_1$ sestru] ?
         whose hit   Petr$_{nom}$ sister$_{acc}$
         'Whose sister did Petr hit?'

   d.   Čí$_1$ Petr   udeřil [$_{DP_3}$ t$_1$ sestru] ?
         whose Petr$_{nom}$ hit   sister$_{acc}$
         'Whose sister did Petr hit?'

Thus, the Czech data illustrate the same pattern as the German data in the previous section: Local scrambling creates a melting effect. Interestingly, the Czech examples differ from the German examples with respect to the position of the verb. In all the cases listed here (except for (53-d)), the verb precedes both subject and object.[50] Since

---

[49] It should be noted, however, that the contrast may not be as crystal-clear as those discussed before, as indicated by (*) in (53-b).

[50] Of course, the question arises which position V occupies in these cases. It seems clear that V has undergone movement here, but what exactly the nature of the landing site is does not seem to be important for present purposes, so I will leave this issue unresolved.

extraction from the subject is only possible when the object is scrambled, and as a result comes to separate the verb and the subject, the conclusion arrived at on the basis of data like (48-b) in German is reinforced: The melting effect (as it can be seen with subjects in German) cannot be reduced to an adjacency effect.

5.3   Further Issues

The present approach differs from other theories of locality with a focus on CED effects in that it envisages the possibility that some XP α may or may not be a barrier in two syntactic contexts that do not differ with respect to the relation between α and the surrounding heads of the clausal projection. Thus, I have assumed so far that DP_3 in (44), (45), (48), and (49) in German occupies exactly the same position in the (a)-examples and in the (b)-examples; and the same goes for (the subject) DP_3 in the Czech examples (50), (51), (52), and (53). Nevertheless, DP_3 qualifies as a barrier if it is last-merged in vP, but becomes transparent by local scrambling of some item to an immediately preceding position in vP (the melting effect). If the structural position of DP_3 itself is identical throughout, the differences with respect to extraction options do not seem to be accountable for under standard approaches to the CED for principled reasons. Consider, for instance, the classical approach to CED effects in a pre-minimalist setting: XP is a barrier iff it is not L-marked by an appropriate head (as in Chomsky (1986a), Cinque (1990), and much related work; see footnote 10 of chapter 2 for a definition). If the relations between the potentially relevant heads (C, T, v, V) and DP_3 are identical in all the German examples in (44), (45), (48), and (49), DP_3 should either uniformly qualify as a barrier, or it should uniformly be transparent; and the same goes for the Czech examples in the previous subsection. Similar conclusions apply in the case of more recent, minimalist reconstructions of CED effects (cf. the literature discussed in chapter 2). Therefore, if it can be shown that DP_3 occupies identical positions in the (a)-examples and in the (b)-examples, this can be taken as direct support for the present analysis.

Indeed, all available evidence points in this direction. To give an example (and focussing on German), topicalization of V plus object DP (stranding a subject DP) typically yields well-formed results in German, whereas topicalization of V plus subject DP (stranding an object DP) does not. Thus, the generalization suggests itself that VP can undergo topicalization whereas vP cannot undergo topicalization (at least not as easily; see Haider (1993a) for a few, systematically classifiable counter-examples). For present purposes, we may ignore the question of what accounts for this generalization (and how the putative exceptions are treated); it may suffice to note that the examples in (54) instantiate this contrast.

(54) a. *$[_{vP_2}$ $[_{DP_3}$ Ein Buch] t_1 beeindruckt] hat ihn$_1$ nicht t_2
              a   book$_{nom}$ impressed   has him$_{acc}$ not

    b. *[$_{VP_2}$ [$_{DP_3}$ Ein Buch]   [$_{DP}$ den Fritz]   beeindruckt] hat wirklich nicht t_2
              a   book$_{nom}$ the Fritz$_{acc}$ impressed   has really   not

    c.   [$_{VP_2}$ [$_{DP_3}$ Ein Buch] gelesen] hat er$_1$ nicht t_2
              a   book$_{acc}$ read   has he$_{nom}$ not

As shown in (55), this contrast is maintained in those cases where *wh*-extraction from DP$_3$ has taken place. Here, *wh*-extraction from DP$_3$ (subject in (55-a), object in (55-b)) to the embedded SpecC position is followed by long-distance topicalization of V+DP$_3$. This involves a remnant movement configuration. However, the configuration is legitimate as such (abstracting away from a weak *wh*-island effect that cannot be avoided; see Fanselow (1987) and chapter 5); in particular, no constraints on remnant movement are violated (see Müller (1998)). Therefore, the fact that (55-a) is ungrammatical strongly suggests that DP$_3$ does not show up in a lower, VP-internal position even if *wh*-extraction has taken place, and a melting effect due to object fronting cannot be excluded.[51]

(55) a. *[$_{VP_2}$ [$_{DP_3}$ Ein Buch t_1] t_4 beeindruckt] weiß ich nicht [$_{CP}$ [$_{PP_1}$ über
                    a   book$_{nom}$ impressed   know I now   about
      wen] ihn$_4$ t_2 hat]
      whom him$_{acc}$ has

    b. ??[$_{VP_2}$ [$_{DP_3}$ Ein Buch t_1] gelesen] weiß ich nicht [$_{CP}$ [$_{PP_1}$ über wen] er$_4$
                    a   book   read   know I not   about whom he
      t_2 hat]
       has

I thus take it to be uncontroversial that subject DPs in German (i.e., nominative DPs in constructions that are neither unaccusative nor passive) never show up VP-internally.[52]

    The remaining alternative is that the subject DPs (labelled DP$_3$) in (44), (45), and (48) are in Specv in the well-formed cases, and in SpecT in the ill-formed cases; and that the theory of locality is then somehow sensitive to this difference. However, such

---

[51] As a matter of fact, the illformedness of (54-a) might already suggest the same thing (depending on further assumptions).

[52] Note that this is the main reason for excluding a version of the present analysis according to which scrambling does not involve movement, but is rather due to an operation of subcategorization feature reversal in the lexicon, as envisaged by Haider (1988) and Fanselow (2001a) for German. On such a view, a "scrambled" object would be externally merged above the subject, and this would then also account for the melting effect. I do not adopt such an analysis for various reasons here (having to do with what still strikes me as significant evidence for scrambling as movement in German, despite the arguments to the contrary given in Fanselow (2001a), on which also see Assmann & Heck (2011)), independently of the issue just discussed in the text. However, these considerations notwithstanding, it is at least worth bearing in mind that there is no principled incompatibility of a scrambling-as-base generation approach with the analysis of CED effects suggested in the present chapter.

an account does not seem viable either. The first thing to note is that whereas there is good evidence that the subject DP is within vP in examples like (44-b) (given that scrambling in German cannot target a domain beyond vP), there is no evidence whatsoever that the subject DP has undergone optional raising to SpecT in ungrammatical (non-melting) examples like (44-a); indeed, from a theory-internal point of view, it is hard to see what could force (string-vacuous) subject raising to SpecT in (cases like) (44-a) while (at least optionally) blocking it in (cases like) (44-b).

Second, Webelhuth (1992), Diesing (1992) and others have proposed that certain items (like the particles *ja*, *doch*, *denn*, *etwa*) demarcate (what is in current terminology) the vP edge. Suppose that this is the case (for base generation; see below, and notwithstanding the critical remarks in chapter 3, subsection 3.4). It can then be noted that the addition of a particle preceding the subject DP does not improve extraction from the subject DP in the examples that do not involve local object scrambling; on the other hand, the melting examples stay well formed. See (56-a) vs. (56-bc), and (56-d) vs. (56-ef). This state of affairs implies that the barrier status of subject DPs in German discussed so far uniformly involves Specv, which confirms the reality of the melting effect. It also implies that particles like *denn* and *wohl* either do not enter the syntactic derivation via structure-building features, or fail to be part of the vP proper. I will assume the former because the examples in (56-cf), with object scrambling to a position in front of the particle, are well formed; given that scrambling is confined to the vP/VP domain in German, this implies that the particles in question may demarcate the vP edge before movement, but do not have to show up at the left phonological border of vP after scrambling has applied.[53]

(56) a. *Was$_1$ haben denn [$_{DP_3}$ t$_1$ für Bücher] [$_{DP_2}$ den Fritz]   beeindruckt ?
        what have PRT   for books$_{nom}$ the Fritz$_{acc}$ impressed

    b. Was$_1$ haben denn [$_{DP_2}$ den Fritz] [$_{DP_3}$ t$_1$ für Bücher] t$_2$ beeindruckt ?
        what have PRT   the Fritz$_{acc}$ for books$_{nom}$ impressed

    c. Was$_1$ haben [$_{DP_2}$ den Fritz]   denn [$_{DP_3}$ t$_1$ für Bücher]   t$_2$ beeindruckt ?
        what have   the Fritz$_{acc}$ PRT   for books$_{nom}$ impressed

    d. *[$_{PP_1}$ Über wen] hat wohl [$_{DP_3}$ ein Buch t$_1$] [$_{DP_2}$ den Fritz]
        about whom has PRT   a book$_{nom}$ the Fritz$_{acc}$

---

[53] The data reported here are based on an informal survey among eight native speaker linguists. The judgements were nearly unanimous: Adding a particle in front of the subject DP does not improve grammaticality. Diesing (1992, 32) has contradictory judgements for data of the type in (56-a) and (56-d). I do not follow her judgements here; but it might still be worth noting that if one were to adopt Diesing's (1992) view, this would in fact follow in the present approach as a melting effect under the assumption that particles like *denn* and *wohl* enter the syntactic derivation as a result of designated structure-building features on v, rather than by some other (adjunction-like) operation, as assumed here.

       beeindruckt ?
       impressed

    e. $[_{PP_1}$ Über wen] hat wohl $[_{DP_2}$ den Fritz] $[_{DP_3}$ ein Buch t_1] t_2
           about whom has PRT   the Fritz$_{acc}$ a book$_{nom}$
           beeindruckt ?
           impressed

    f. $[_{PP_1}$ Über wen] hat $[_{DP_2}$ den Fritz]   wohl $[_{DP_3}$ ein Buch t_1] t_2
           about whom has   the Fritz$_{acc}$ PRT   a book$_{nom}$
           beeindruckt ?
           impressed

Consider finally what happens when these kinds of particles are added in a position to the right of the subject DP. On the one hand, it does not come as a surprise that judgements for (57-ac) stay as they are for the corresponding examples in (56-ad) (the subject DP is an island to begin with, independently of a freezing effect, and, assuming that reflexive melting is not available, there is no melting configuration because the object has not passed the subject). On the other hand, an order with the particle following both the subject (from which extraction takes place) and the object also leads to ungrammaticality; see (57-bd). This follows as a freezing effect: Given that the particles demarcate the left edge of vP after external Merge, movement of the subject must be involved here (to an outer specifier of v, since optional raising to SpecT would exclude subsequent object scrambling to a position in front of it); this creates a freezing effect which cannot be undone by object fronting (given the assumptions laid out in the previous section).

(57) a.   *Was$_1$ haben $[_{DP_3}$ t_1 für Bücher] denn t_3 $[_{DP_2}$ den Fritz] beeindruckt ?
         what have   for books$_{nom}$ PRT   the Fritz$_{acc}$ impressed

    b.   *Was$_1$ haben $[_{DP_2}$ den Fritz] $[_{DP_3}$ t_1 für Bücher] denn t_3 t_2
         what have   the Fritz$_{acc}$ for books$_{nom}$ PRT
         beeindruckt ?
         impressed

    c.   *$[_{PP_1}$ Über wen] hat $[_{DP_3}$ ein Buch t_1] wohl t_3 $[_{DP_2}$ den Fritz]
           about whom has   a book$_{nom}$ PRT   the Fritz$_{acc}$
           beeindruckt ?
           impressed

    d.   *$[_{PP_1}$ Über wen] hat $[_{DP_2}$ den Fritz] $[_{DP_3}$ ein Buch t_1] wohl t_3 t_2
           about whom has   the Fritz$_{acc}$ a book$_{nom}$ PRT
           beeindruckt ?
           impressed

To end this section, let me turn to a derivation that may at first sight seem problematic for the present approach because it looks as though it might indicate a loophole offering melting derivations for non-melting (i.e., ungrammatical) constructions. Consider

the following, alternative derivation of a non-melting example like (44-a):

(58) *Was$_1$ haben [$_{TP}$ [$_{DP_3}$ t$_1$ für Bücher] [$_{vP}$ [$_{DP_2}$ den Fritz] [$_{v'}$ t$_1'$ [$_{v'}$ t$_3$ [$_{v'}$ [$_{VP}$ t$_2$
      what have   for books$_{nom}$ the Fritz$_{acc}$
      beeindruckt] v]]]] T] ?
      impressed

Here, the object DP$_2$ is scrambled to an outer specifier of vP, in front of the subject's base position. Before the object undergoes scrambling, the subject would therefore be predicted to be transparent for extraction: An edge feature can be assigned to v because the phase head is still active at this point, bearing the [•Σ•] feature for local object scrambling; and the *wh*-phrase *was* may thus move to an outer specifier before the object is scrambled. From this position, it may then move to SpecC in accordance with the PIC (via SpecT, a further intermediate step that is not indicated in (58)). All this is as in genuine melting examples like (44-b). However, there is a crucial difference between (58) and (44-b): In the latter example, the subject DP stays in situ; but in the former, it undergoes optional raising to SpecT (triggered by an EPP feature that can optionally be present on T).

Given the assumptions made so far, there is nothing wrong with the derivation in (58), but it must of course be excluded in some way. Various possible solutions suggest themselves; I will briefly sketch two. First, note that (58) instantiates a type of Duke-of-York derivation (see Pullum (1979)) based on the interaction of optional movements that is well known as a potential problem for theories of movement, independently of the issues that are presently under consideration. The order A∼B is first changed into the order B∼A, which is subsequently changed into an order A∼B again, with no interaction with other material (C) involved. As argued by Ross (1967) and Haider (1993a) (among many others), such cases of *globally string-vacuous optional movement*, which reproduces the very same order that was present initially at the end of the derivation, need to be blocked in any theory that envisages optional movement; otherwise sentences could have infinite numbers of legitimate derivations.[54] Duke-of-York derivations resulting in globally string-vacuous movement of the type in (58) could be prohibited by adopting an economy constraint on optional features in numerations: The insertion of optional features in numerations is blocked if these features do not have an effect on outcome.[55]

---

[54] Note that whereas globally string-vacuous movement would be excluded, this would not hold for locally string-vacuous movement, as it in fact had to be assumed to be legitimate above. Locally string-vacuous movement differs from globally string-vacuous movement in that it eventually (at the end of the derivation) leads to new orders of items.

[55] See Müller (1998) for such a constraint, which also derives various kinds of improper movement effects that were subsumed under the Principle of Unambiguous Binding (PUB) in Müller & Sternefeld (1993).

An alternative solution to the problem of excluding (58) would be to assume that subjects are in fact not part of the edge domain of vP in their base position; only a derived specifier would count as part of the edge (see Müller (2007)). Then, movement of the subject DP_3 in (58) would have to proceed via a further specifier of vP, and this would give rise to the dilemma for freezing derivations sketched in section 4 – both the subject DP_3 and the *wh*-element to be extracted from it then need to move to an outer specifier of v, triggered by [●D●] in one case, and by [●wh●] in the other, with the edge feature for the more inclusive category inserted first (and, as a consequence, insertion for the dominated category not available anymore). The two solutions differ slightly in their predictions, but both seem viable in principle.[56]

There may well be other derivations that may at first sight look like they provide loopholes for extraction, but it seems to me that there is no reason to assume that they cannot be excluded a similar way, by invoking independently motivated constraints.[57] In general, then, I take the melting effect to be well established. Most of the examples I have discussed involve the vP domain, and this is certainly not accidental. The vP domain has exactly the properties that are favourable to the occurrence of melting: It is a domain that hosts an externally merged specifier (the external argument), as well as possible further internally merged specifiers (scrambled items that stay in this position for the rest of the derivation). The VP domain is similar in this respect in double object

---

[56] An object DP may cross adverbial material (adjuncts) or particles but end up below the subject. As for adverbs, it could be assumed that they show up in a position c-commanded by v in this case, either in a VP specifier, or in a designated functional projection intervening between vP and VP – recall that the barrier status of adjuncts is derived if these items are last-merged specifiers of special functional projections (see Alexiadou (1997), Cinque (1999)). As for particles, which I have provisionally assumed to demarcate the vP edge (following standard practice), the order $DP_{nom}{\sim}DP_{acc}{\sim}$PRT, which is possible as such, would have to be derived by movement. The subject DP remains a barrier in this construction:

(i) *Was_1 haben [$_{DP_3}$ t_1 für Bücher]   [$_{DP_2}$ den Fritz]   denn t_3 t_2 beeindruckt ?
      what have   for books$_{nom}$ the Fritz$_{acc}$ PRT   impressed

Here, the second analysis (in terms of a narrower concept of edge) might offer a somewhat simpler account than the first analysis (that relies on a ban on globally string-vacuous movement): Movement in (i) is not globally string-vacuous, but it instantiates a freezing effect if DP_3 obligatorily moves to (an outer) Specv. Then again, the approach in terms of obligatory intermediate subject (more generally, specifier) movement to an outer specifier raises questions of compatibility with the approach to superiority effects given in chapter 3, where it was crucial that a base-generated specifier may produce a balanced phase that blocks edge feature insertion for movement of a lower competing item; see in particular the discussion of superiority effects in double object constructions in English, where both *wh*-phrases are externally merged in the same phase.

[57] To name just one example: As observed by Winfried Lechner (p.c.), a melting effect fails to show up in cases like (i-a), where v could be assumed to be equipped with two [●Σ●] features, and both PPs undergo scrambling to Specv, with PP_1 being extracted out of PP_2 after the latter has undergone scrambling (and v still active at this point, due to the remaining [●Σ●] feature).

contexts, and a melting effect can be observed here, too. Other domains lacking this dual role (e.g., CP and TP, which only have derived specifiers, except perhaps for items like expletives that extraction cannot take place from for obvious reasons) are therefore not expected to show melting effects. Still, it is likely that melting effects can be found in other domains as well (e.g., NP and PP); but I will have to leave that for future research.

6. Outlook and Conclusion

6.1 Outlook

It has sometimes been claimed, for certain movement types or certain languages, that subjects are not necessarily barriers for extraction, or that, more generally, the CED does not hold; see Haider (1983; 1993a) on German, Frantz (1980) on Blackfoot, Bickel (2004) on Belhare, Lasnik & Saito (1992), Takahashi (1994), Rackowski & Richards (2005) on Japanese (for null operator movement), Zeller (2005) on Zulu, and Stepanov (2007) on Navajo, Turkish, Palauan, Hungarian, and Russian.[58] As noted above, at least for German, it seems to me that many of the putative counter-examples to the generalization that subjects in Specv are islands can be rejected by showing that they either involve VP-internal arguments or no extraction at all (or melting); and it does not strike me as unlikely that similar explanations can be given for many apparent counter-examples in other languages. However, at present I take it to be an open question whether such a way out is available for all constructions, and all languages where apparent exceptions from the CED have been observed.

   One case that is instructive in this context is the behaviour of expletive constructions in English. In English expletive *there* constructions, the associate DP is not inherently a barrier; see (59-ab) (see Moro (1997) and Lasnik & Park (2003)). In

---

(i)   a. *dass sie   [$_{vP}$ [$_{PP_1}$ über Bakunin] [$_{v'}$ [$_{PP_2}$ von einem Vortrag t$_1$] [$_{v'}$ [$_{DP_3}$ eine
           that she$_{nom}$ about Bakunin   of   a   talk   a
           Aufzeichnung t$_2$] machte]]]
           recording$_{acc}$ made

       b. *dass sie   [$_{vP}$ [$_{PP_1}$ über Bakunin] [$_{v'}$ [$_{DP_3}$ eine Aufzeichnung [$_{PP_2}$ von einem Vortrag t$_1$]]
           that she$_{nom}$ about Bakunin   a   recording$_{acc}$ of   a   talk
           machte]]
           made

However, notwithstanding the fact that extraction of a PP from a PP is ungrammatical for independent reasons in German (see (i-b), with PP$_2$ in situ), (i-a) instantiates scrambling from a scrambled item, which is a configuration of just the type shown to be generally impossible in section 4 above.

[58] Recall from footnote 13 of chapter 2 that Rackowski & Richards (2005, 585) assume that CED effects with subjects can be avoided if a language can a establish an Agree relation between v and a specifier, in violation of the c-command requirement on Agree relations assumed in Chomsky (2001b; 2008) and much related work.

contrast, DP is a barrier for extraction in locative inversion constructions, as in (59-c).

(59) a.   Which wall$_1$ do you think there$_3$ was [$_{vP}$ t$_3$ [$_{v'}$ v [$_{DP_2}$ a picture of t$_1$]]] ?
      b.   Which candidate$_1$ were there$_3$ [$_{DP_2}$ posters of t$_1$] all over town ?
      c.   *Who$_1$ do you think [$_{PP_3}$ on this wall] hung [$_{DP_2}$ a picture of t$_1$] t$_3$?

Does the wellformedness of (59-ab) raise a problem for the view that last-merged specifiers of vP are islands? Given the present system of assumptions, the contrast in (59) would seem to suggest that DP$_2$ in (59-c) is last-merged in its projection whereas DP$_2$ in (59-ab) is not. I take it to be uncontroversial that DP$_2$ is indeed last-merged in its projection in (59-c), with PP$_3$ initially occupying a lower position and then undergoing raising to an outer specifier position (perhaps SpecT). As for (59-ab), the absence of a CED effect follows immediately if one assumes that *there* is a subject and DP$_2$ is its predicate (see Williams (1994; 2006)), with the two items merged in the same projection (vP), as argued in Hazout (2004) and Hartmann (2005). If so, DP$_2$ in (59-ab) is in fact not last-merged in its projection because *there* is merged later, in the same projection, as indicated (subsequently, *there* undergoes raising to SpecT). Therefore, DP$_2$ does not have to be an island for extraction.

Another case that might shed some light on this issue is that of a German expletive pronoun *es* optionally accompanying a finite subject clause from which extraction has taken place. Fanselow & Mahajan (2000) claim that (60-a) (extraction in the presence of *es*) is just as good as (60-b) (extraction from a subject clause without *es* being present); interestingly, for some speakers, the presence of *es* actually *improves* the example. (That said, for what is perhaps the majority of speakers, both examples are equally ungrammatical, as one would expect under the approach developed above; also see Stechow (2000).)

(60) a. #Wen$_1$ ärgert es dich   [$_{CP}$ dass sie t$_1$ liebt] ?
          whom$_{acc}$ annoys it you$_{acc}$ that she$_{nom}$ loves
      b. #Wen$_1$ ärgert dich   [$_{CP}$ dass sie t$_1$ liebt] ?
          whom$_{acc}$ annoys you$_{acc}$ that she$_{nom}$ loves

Suppose, at least for the sake of illustration, that the amelioration effect triggered by the presence of *es* is real. In the present system, this could then be interpreted as a kind of "pseudo-melting" effect, similar to the situation in *there*-constructions: If one assumes (i) that the (phrase headed by the) expletive does not stand in a dominance relation with the subject clause (contra Ross (1967) and much later work), and (ii) that the expletive is merged after the finite subject clause, but within the same projection, then some item in the left edge of the subject clause will be able to undergo edge-feature driven movement to an outer vP specifier (while the subject clause is still in situ), and a CED effect can be avoided. Such an analysis in terms of [•expl•] features (which presumably would by definition have to be ordered between standard

subcategorization features and movement-inducing features in features stacks of lexical items) might then possibly be extended to cases like (60-b) – in other words, zero expletives might underlie an absence of CED effects.[59] Still, as remarked above, it is far from clear whether either one or both of the examples in (60) should indeed be categorized as grammatical in the first place (and if they are not, giving up assumption (i) or assumption (ii) will suffice to maintain the CED effect).

The problems just noted notwithstanding, it is worth pointing out that this approach might also contribute to a solution of a well-known problem that shows up with many theories of locality in a domain where judgements are much less controversial, viz., *wh*-extraction (or relativzation, or topicalization) from infinitives in German: In contrast to finite clauses, subject (and object) infinitives are transparent for extraction in German (cf. Haider (1983; 1993a), Sternefeld (1985), and Grewendorf (1989)). A pair of examples that shows this effect is (61-bc), with the subject infinitive in the pre-subject (base) and in an extraposed position. In (61-a), the subject infinitive is accompanied by an expletive *es*, and extraction is also possible.[60]

(61) a. Was$_1$ hat es sich nicht gehört [$_{CP}$ t$'_1$ PRO t$_1$ zu beanstanden] ?
          what$_{acc}$ has it not been proper   to object to

      b. Was$_1$ hat [$_{CP}$ t$'_1$ PRO t$_1$ zu beanstanden] sich nicht gehört ?
          what$_{acc}$ has   to object to   not been proper

      c. Was$_1$ hat sich nicht gehört [$_{CP}$ t$'_1$ PRO t$_1$ zu beanstanden] ?
          what$_{acc}$ has not been proper   to object to

Assuming that evidence for a zero expletive alternating with *es* can be found in this case, the transparency of all kinds of subject infinitives in German could be treated as another instance of a pseudo-melting effect.

Needless to say, all these remarks are somewhat speculative, and the late-merged items creating pseudo-melting effects postulated by these analyses would eventually have to be substantiated by independent evidence. However, the main purpose of this subsection is not to justify actual analyses, but rather to show a direction in which one might profitably look upon encountering exceptions to the CED with external arguments in Specv that are transparent for extraction (or related configurations), in a given language: In those cases, it might well be that there is a (possibly non-overt) item that is merged later, in the same projection. This item would permit the insertion

---

[59] The analysis sketched here presupposes that CP extraposition from a Specv (or, for that matter, VP-internal complement) position does not create an island, as an instance of freezing. A freezing effect can be avoided with extraction from a CP if CP extraposition targets a high position, making it an instance of remnant movement; see Müller (1998).

[60] Note that (61-a) must involve extraposition, as is typical in the presence of expletives that accompany clauses.

of an edge feature for some item in the specifier of an otherwise opaque XP, thereby creating a pseudo-melting effect.[61]

That said, in the appendix, I will pursue the issue even one step further, and investigate the possibility of deriving reflexes of successive-cyclic movement as instances of pseudo-melting.

6.2   Core Assumptions Reconsidered

I have shown that the version of the Condition on Extraction Domain (CED) in (5) (according to which some item is a barrier if the operation that has merged it in a phrase is the final operation taking place in that phrase) follows as a theorem in a derivational theory of syntax that incorporates the Phase Impenetrability Condition (PIC) and the four assumptions in (62).

(62) a.   All syntactic operations are driven by features of lexical items.
      b.   These features are ordered on lexical items.
      c.   All phrases are phases.
      d.   Edge features that trigger intermediate movement steps can be added only as long as the phase head is still active.

As with the analysis of (G)MLC effects in chapter 3, the question arises to what extent these four main assumptions underlying the present approach to CED effects are crucial; after all, it might be that abandoning one or more of these assumptions does not necessarily imply that no (relevant) version of the CED can be derived. In fact, closer inspection reveals that the four assumptions are not equally important.

Consider first assumption (62-c). As noted in section 2.3 of chapter 3, I think there is solid independent evidence (both empirical and conceptual) for assuming that all phrases are phases, in the sense of the PIC. Still, if one assumes that there are fewer phases, this may diminish some of the empirical impact of the CED then derivable, but it in no way undermines the theorem status of the CED as such. For concreteness, suppose that only vP and CP are phases; then the (last-merged) specifiers of vP and CP are derived as islands. If DP is also a phase, then DP specifiers come out as islands, too; and so on. This does not cover all the observable CED effects, but it covers most of the effects discussed in the present chapter, and thus represents an interesting result.[62]

---

[61] This reasoning is comparable to Bittner & Hale's (1996) treatment of accusative case assignment in terms of abstract pseudo-coarguments.

[62] Since phases thus help to account for CED effects, under any definition (narrow or broad), the present chapter can be viewed as an argument in support of the general concept of a phase. It certainly calls into question the claim that "locality doesn't offer any argument for phases" (cf. Boeckx & Grohmann (2007, 214)).

Turning next to assumption (62-a), I would again like to contend that whereas adopting it is well motivated, given an incremental, minimalist approach to structure-building where the Inclusiveness Condition (although minimally violable with intermediate movement steps) ensures that all operations are ultimately driven by the properties of lexical items; but it can easily be replaced with alternative assumptions without threatening to undermine the overall approach. The main task of assumption (62-a) is to make it possible to decide, at any given step of the derivation, whether a phase head is still active, or has become inert. In the present approach, the (in-)activity of a phase head can be determined by inspecting whether it has any (probe or structure-building) features left to discharge; but if the (in-)activity of a phase head can be determined in some other way (e.g., by carrying out semantic interpretation cyclically, or by inspecting subarrays), then this will also be sufficient for deriving CED effects from the PIC.

Similar considerations apply in the case of assumption (62-b). The hypothesis that operation-inducing features on a lexical item are ordered in stacks can be taken to be independently motivated, and it permits a simple account of the last-in/first-out property of edge feature insertion, but replacing it with an alternative assumption that does not rely on feature hierarchies would leave most of the reasonings above intact (potentially giving rise to analyses of the type discussed in footnote 52).[63] Again, what is crucial is that there is some notion of a phase head becoming inert at some stage of the derivation.

Let me finally turn to assumption (62-d): Edge features that trigger intermediate movement steps can be added only as long as the phase head is still active. Replacing *after* with *before* in the EFC (as documented in the different definitions in (4)/(3) and (11)) forms the gist of the present chapter. At least at first sight, this would seem to be the sole indispensable part of the derivation of CED effects from the PIC. However, even here we can distinguish between the underlying idea and its technical implementation, and replace the latter within alternative systems of movement and structure-building. What is needed is some way of classifying intermediate movement steps as costly; some operation ζ is required to effect them, and ζ's application is restricted to contexts in which the phase head is still active. In the present approach, ζ is the insertion of an edge feature; but ζ could in principle be any other means triggering intermediate movement steps.[64] This is then the irreducible core of the present pro-

---

[63] Of course, dispensing with assumption (62-a) implies dispensing with assumption (62-b) as well, whereas the opposite does not hold.

[64] Here are two possible options that do without edge features (others are conceivable): First, on the basis of the non-feature driven approach developed in Heck & Müller (2000; 2003) (see chapter 3), ζ could be assumed to be movement to an intermediate position that violates Last Resort in order to satisfy Phase

posal: Operation ζ (however it is defined) is responsible for inducing movement to (intermediate) phase edges, and ζ can only apply at any given step of the derivation if the phase head providing the phase edge qualifies as still "active" (however this is defined). Under this assumption, the CED effects follow from the PIC, in line with Chomsky's informal proposal that "what yields the subject-island effect, it appears, is search that goes to deeply into a phase already passed" (see Chomsky (2008, 154)).

Last but not least, I would like to highlight a general point concerning the organization of grammar. Deriving CED effects from the PIC in the way laid out in this chapter is only possible in a derivational approach to syntax that recognizes *before* and *after* as basic theoretical concepts, and in which there is a total order of all operations in a derivation (and thus no room for operations applying simultaneously in a single phrase marker). Therefore, to the extent that the account can be characterized as successful, it provides an argument for a strictly derivational organization of grammar (and thereby enforces the conclusion reached at the end of the preceding chapter).

7. Appendix: Movement-Related Morphology

7.1 Background

In general, there are three possible types of analyses of (apparent) correlations of syntactic and morphological phenomena. First, one may claim that the morphological phenomenon makes the syntactic phenomenon possible. Second, one may claim that the syntactic phenomenon makes the morphological phenomenon possible. And third, one may claim that closer inspection reveals that there is no interesting (and theoretically relevant) interaction of the syntactic phenomenon and the morphological phenomenon after all.

Consider, for instance, the issue of a correlation between the syntactic phenomenon of V-to-T movement and the morphological phenomenon of rich verb inflection. On the one hand, it is often argued that rich verb inflection makes V-to-T movement possible (see Roberts (1993), Vikner (1997; 2001b), Holmberg & Platzack (1995), and Rohrbacher (1999), among others, with different concepts of what it

---

Balance; and the assumption replacing the EFC in (11) might be that a phase automatically counts as balanced if its phase head is inert. An inert phase head would thus make a Last Resort violation incurred by an intermediate movement step fatal, thereby yielding a PIC violation on the subsequent cycle with last-merged specifiers. Second, against the background of Chomsky (1993), ζ may be understood as movement to intermediate positions that violates the Fewest Steps condition but is required for convergence of the derivation; the assumption replacing the EFC in (11) might then be that convergence holds by definition if the head of the landing site is not syntactically active anymore. This would imply the selection of a derivation that minimizes the number of movement steps in CED contexts, which violates the PIC. (To be sure, invoking Fewest Steps like this would not be compatible with the meta-requirements on constraints adopted throughout this monograph.)

means for an inflectional system to be "rich"). On the other hand, it has also been proposed that the correlation goes in the other direction, such that V-to-T movement makes rich verb inflection possible (see Bobaljik (2002)). Finally, it has been suggested that there is no (synchronically relevant) correlation between V-to-T movement and rich verb inflection (see Alexiadou & Fanselow (2000) and Vikner (2001a)).

As a second case in point (one that is perhaps even better known), let us look at the question of a correlation between (subject) pro-drop (which, for present purposes, is simply to be understood as the omission of a referential argument pronoun; see chapter 2 for some further constraints) and rich verb inflection. The hypothesis that rich verb inflection makes pro-drop possible can be viewed as being part of the linguistic folklore by now; see, e.g., Rizzi (1986), Jaeggli & Safir (1989), Müller (2006), and Neeleman & Szendrői (2006) for various kinds of approaches that may differ radically in their understanding of what "rich" verb inflection should be taken to be, but that all agree on the general direction of the implicational relation. In line with this, to the best of my knowledge, there are no proposals holding that pro-drop makes rich verb inflection possible (though Chomsky's (1981) approach, where ungoverned PRO is responsible for pro-drop and rule R is assumed to optionally apply in the syntax in pro-drop languages might be viewed as going in this direction). However, in principle this should qualify as a perfectly legitimate hypothesis (of course, one would have to say something about the occurrence of rich verb inflection in non-pro drop contexts). The third possibility, viz., that there is no (synchronically relevant) correlation between pro-drop and rich verb inflection, has also been pursued (see Grimshaw & Samek-Lodovici (1998)). In addition, it has been claimed that (at least certain) instances of pro-drop are not to be treated as syntactic phenomena as such at all (see Adger (2003) and Holmberg (2005)).

A third example concerns the issue of a correlation between scrambling and rich nominal case inflection. Again, the standard assumption is that rich nominal case inflection makes scrambling possible (see, e.g., Haider (1988) and Fanselow (1992); also Chomsky (1965, 126)); however, one can also make a case that the opposite holds true – viz., that scrambling makes rich nominal case inflection possible (see Müller (2002), where the occurrence of nominal case inflection with non-scrambled DPs is accounted for by invoking McCarthy's (1999) concept of *sympathy*). Finally, it might also be the case that there is no (synchronically relevant) correlation between scrambling and rich nominal case inflection (see Müller (1995)).

Thus, in the three empirical domains just considered (V-to-T movement, pro-drop, and scrambling), all three logically possible approaches concerning a potential correlation with the occurrence of specific morphological exponents can be (and, for the most part, have been) pursued; still, it seems fair to conclude that in all three domains, the view that pieces of morphology make syntactic operations possible is the standard one. In view of this, one may take it to be initially surprising that

the situation is reversed with the (possible) correlation of successive-cyclic move-
ment and specific items of movement-related morphology. Here, the standard view
clearly is that the syntactic operation is a precondition for the morphological opera-
tion to occur. More precisely, it is generally assumed that successive-cyclic movement
makes movement-related morphology possible: Morphological pieces act as *reflexes*
of successive-cylic movement (see McCloskey (1979; 2002), Clements et al. (1983),
Collins (1993; 1994), Chung (1994; 1998), Cole & Hermon (2000), Fanselow & Ćavar
(2001), Sabel (2000), Bouma et al. (2001), Dalrymple (2001, ch. 14), Lahne (2009),
and many others). In contrast, as far as I am aware, the opposite assumption – that
movement-related morphology makes successive-cyclic movement possible – has not
really been pursued.[65]

The goal of this appendix is to show that the approach to CED effects devel-
oped in this chapter may offer a new and interesting perspective on the correlation
of movement-related morphology and successive-cyclic movement, where the former
is in fact a precondition for the latter to apply. More specifically, the discussion in
subsection 6.1 above has suggested a concept of *pseudo-melting* as a way to avoid
island effects. This concept can be extended to the domain of movement-related mor-
phology: Movement-related morphology can be viewed as effecting pseudo-melting,
in the sense that a category which would otherwise act as an island becomes transpar-
ent for movement. On this view, movement-related morphology is not a reflex; rather,
it makes (certin kinds of) successive-cyclic movement possible. I hasten to add that
so far, there is little independent empirical evidence for such a view of movement-
related morphology that would show that it is superior to the standard view (and this
is why the topic of movement-related morphology is confined to an appendix). How-
ever, on the conceptual side, the new approach might offer an account as to why this
phenomenon exists in the first place (unlike most standard approaches).

Let me begin with a terminological clarification. So far, I have not been explicit
as regards the notion of pseudo-melting. Now I will be. Suppose that a *melting* effect
describes the situation where internal Merge of some (typically overt) *non-expletive*
XP above a specifier α within the same phase makes α transparent for extraction. In
contrast, a *pseudo-melting* effect can be said to arise when external Merge of some
(possibly covert) *expletive* XP above a specifier α within the same phase makes α
transparent for extraction. Thus, on this view, *all* the relevant well-formed data in
subsection 6.1 are actually instances of pseudo-melting (including *there*-constructions

---

[65] At least not for the case of A-bar movement; see footnote 69 below. – Incidentally, the third a priori pos-
sible position, viz., that there is no (synchronically relevant) correlation between successive-cyclic move-
ment and movement-related morphology also has not yet been adopted; but see Lahne (2009), who argues
that virtually all pieces of movement-related morphology have independent occurrences in non-movement
contexts.

in English).[66]

7.2 Pseudo-Melting in Last-Merged Complements?

Recall that extraction from last-merged complements prima facie poses a problem. This problem was solved by postulating probe features (encoding natural predicates or status-government) which can render a phase head active even *after* the last (and only) structure-building (subcategorization) feature has been discharged. However, this made it necessary to assume that Agree can evade the PIC (an assumption that is arguably independently motivated, given long-distance agreement).

Here is an alternative to this view. Perhaps a verb like *read*, as a lexical property, is accompanied by an empty pseudo-melting expletive (at least in the context of *book*), whereas a verb like *steal* is not (see subsection 3.2.2). Similarly, it might be the case that a verb like *think*, as a lexical property, subcategorizes for an empty pseudo-melting expletive in addition to its CP complement, whereas verb like *know* does not.

Let us explore some of the consequences of this assumption. First, such an approach would be at variance with proposals according to which there is *more structure* with non-bridge (e.g., factive) verbs than there is with bridge verbs; see the additional "operator of some sort" in Manzini (1992, 115), and also Roberts & Roussou (2002).[67]

Second, both bridge verbs like *glauben* ('believe') and *denken* ('think'), and non-bridge verbs like *wissen* ('know') can co-occur with an overt expletive *es* in German. Assuming that such an overt expletive does not improve extraction (but see the remarks in section 6 above), it must be postulated that *es* either dominates its "associate" CP in the base, or is merged in a different projection. For the sake of concreteness, let us assume the former.

Third, and perhaps most importantly, a further consequence would be that in order to maintain the barrier status of subjects in Specv and SpecT (and of indirect objects in SpecV), an additional restriction on pseudo-melting expletives is necessary. This restriction can be tentatively formulated as in (63).

(63) *Restriction on subcategorization features*:

A subcategorization feature that does not go back to a θ-grid can only be dom-

---

[66] In chapters 6 & 7, I will address cases where a phase head bears a priori unexpected operation-inducing features (and thereby makes it possible to circumvent an otherwise expected CED effect); however, these features do not effect Merge of some XP (rather, they trigger verb-second movement of an in-situ V and successive-cyclic deletion, respectively). Although the patterns discussed there are similar to the one addressed in the main text here, the exact mechanics of the analyses are not; therefore, I will reserve the term pseudo-melting for cases of expletive Merge.

[67] However, also cf. de Cuba (2007), who argues that clauses embedded by factive verbs may in fact exhibit less structure than clauses embdded by bridge verbs.

inated by at most one subcategorization feature on a feature stack of a lexical item.

I leave open the question of what (63) might follow from. The consequences for the occurrence of expletives are these: (i) (63) (trivially) permits a subcategorization feature for VP on v ([●V●]), for vP on T ([●v●]), and for TP on C ([●T●]), none of which go back to θ-grids. (ii) (63) permits subcategorization features for expletives if there is only one other XP in the same phase that has been merged earlier (as in English *there* constructions – see (59) –, or in pseudo-melting analyses of transparent last-merged complements). (iii) (63) does not permit subcategorization features for expletives with subjects (i.e., external arguments base-merged in Specv) and indirect objects (which are base-merged in SpecV).[68]

Suppose, at least for the sake of the argument, that such an approach to the transparency of last-merged complements is viable. This then offers an fresh view on the occurrence of movement-related morphology.

7.3   Movement-Related Morphology

7.3.1   Hypothesis

The hypothesis I would like to pursue now is that movement-related morphology is not a reflex of successive-cyclic movement; rather, movement-related morphology is introduced as a last resort so as to avoid an island violation with last-merged XPs (typically complements) where probe features (that would keep the phase head active, and accessible as a target for edge feature insertion) are not available. To check the implications of this assumption, I will now consider three instances of movement-related morphology that have been discussed in the literature, viz., complementizer selection in Modern Irish, *MeN* deletion in Malay, and complementizer selection in Kinande. I address the three phenomena in turn, giving most attention to the first one in order to clarify basic questions and concepts.[69]

7.3.2   Complementizer Selection in Modern Irish

Data   Complementizers in Modern Irish vary in form, depending on whether or not movement has taken place via SpecC (see McCloskey (1979; 2002)). The basic ob-

---

[68] Note that (63) is therefore incompatible with the approach to putative subject clause transparency in German sketched above.

[69] I will have nothing to say here about the interaction of participle agreement and A-movement in Germanic and Romance languages; see Svenonius (2001), Holmberg (2002), D'Alessandro & Roberts (2008) and Richards (2009) for recent discussion. As alluded to in footnote 65, some of these approaches do in fact assume that it is the presence of designated pieces of inflectional morphology (or of the features that are eventually realized by certain morphological exponents) that makes movement possible (and necessary), for instance, by creating a phase that would otherwise not be present.

servations are these. The regular form of declarative C is *go*. However, if SpecC is
targetted by movement, C takes the form *aL*; this is thus an instance of movement-
related morphology. In addition, if a long-distance dependency is expressed without
movement (which McCloskey argues to be an option), by a resumptive pronoun in situ,
C takes the form *aN*. In what follows, I will ignore this third possibility, focussing on
the first and second one.[70] Relevant data are given in (64-a) (with no A-bar movement)
and (64-bc) (with *wh*-movement and relativization via empty operator movement, re-
spectively).[71]

(64) a.   Creidim gu-r   inis sé bréag
        I-believe C:*go*-PAST tell he lie
        'I believe that he told a lie.'

     b.   Céacu ceann$_1$ a   dhíol tú t_1 ?
        which one   C:*aL* sold you
        'Which one did you sell?'

     c.   an t-ainm OP$_1$ a   hinnseadh dúinn a   bhí t_1 ar an áit
        the name   C:*aL* was told to us C:*aL* was   on the place
        'the name that we were told was on the place'

The gist of the analysis in McCloskey (2002, 203) is this: C which bears both an
Op-feature and an EPP-feature is realized as *aL*; C which bears only an EPP-feature is
realized as *aN*; and C which bears neither an Op-feature nor an EPP-feature is realized
as *go*.

Sketch of an Analysis   Let us see what one would have to assume in an approach that
recognizes movement-related morphology in Modern Irish as an instance of a pseudo-
melting effect. Suppose that VP and vP are inherently transparent in Modern Irish,
either because they are not last-merged in their projections, or because they are status-
governed last-merged complements (or because extraction is permitted by a pseudo-
melting expletive that is necessarily realized as zero); however, suppose furthermore
that TP is *not* inherently transparent in Modern Irish: It is the last-merged complement
of C, and C does not have a probe feature for T that might keep it active after TP is
merged. TP should then qualify as a barrier in Modern Irish. To permit extraction,
a subcategorization feature for an expletive item can be inserted on C as a last resort
strategy. This creates a pseudo-melting effect: TP becomes transparent because an
edge feature can now be instantiated on C that attracts an item in the edge domain
of T, to the inner specifier of C (in line with the Intermediate Step Corollary). The

---

[70] Thus, I will also abstract away from mixed patterns arising from combinations of the second and the
third strategies ("chain hybridization"); see Assmann, Heck, Hein, Keine & Müller (2010).

[71] Throughout, I presuppose that the second possibility does indeed involve movement; but see Adger &
Ramchand (2005) for a reanalysis in terms of Agree.

expletive that is required by the newly added structure-building feature on the phase head is subsequently merged in the C domain. This concludes the syntactic part of the analysis. Now there is an additional expletive that shows up as an outer specifier of C. The remaining question is how this leads to the new morphological realization *aL*.

The key to a solution is that the expletive is a functional category. Let us assume that functional categories are mere bundles of morpho-syntactic features in syntax; they are subject to post-syntactic morphological realization (see Halle & Marantz (1993) and much other work in Distributed Morphology). This assumption, as such, does not yet suffice. As it stands, the left edge of the CP domain looks as in (65) (where δ is the item that has carried out the intermediate movement step to SpecC).

(65) $[_{CP}$ EXPL $[_{C'}$ δ_1 $[_{C'}$ C $[_{TP}$ t_1 ... $]]]]$

The problem is that EXPL and C must somehow be realized as *aL*, whereas C alone is realized as *go*. Various possibilities arise. One is to assume that C-realization is context-dependent – if there is an EXPL in C's edge, morphological C realization proceeds differently. Another possibility arises if we take into account that δ_1 is an item that eventually leaves the CP; at the point when C (or the left edge of CP) is realized, δ is not part of the structure anymore.[72] Thus, before morphological realization, EXPL and C are adjacent. Given adjacency, there might be *fusion* of the two categories, leading to a primitive node that is spelled out as a unit (see Halle & Marantz (1993) and Halle (1997) for this kind of approach). Alternatively, we may assume that post-syntactic morphological spell-out can affect *subtrees*, not just *nodes* (see Ackema & Neeleman (2003; 2004), Neeleman & Szendrői (2006), Caha (2008), Starke (2009), and Taraldsen (2008), among others). On either view, as a result, the expletive in the outer SpecC position can only be spelled out together with C (see Lahne (2007)). The only additional assumption that needs to be made is that bare C is realized as *go*, whereas EXPL+C (as a fused node, or as a subtree) is realized as *aL*. Thus, in a nutshell, this analysis holds that TP is a barrier in Modern Irish; that an expletive is added to C to remove barrier status and permit successive-cyclic extraction (an instance of pseudo-melting); and that the pseudo-melting expletive is the locus of movement-related morphology.[73]

This approach seems to me to be tenable, and potentially revealing because it pro-

---

[72] Given the hypothesis made at the end of chapter 3, movement does not leave traces; t_1 is present in (65) purely to simplify exposition.

[73] A side remark: The movement-related morphological exponents in Irish are verbal particles that may in fact ultimately be better analyzed as vP-, or Agr$_O$P-items (cf. Sells (1984), Bouma et al. (2001), Lahne (2009) for discussion). In that case, it would have to be VP or vP (rather than TP) that is not inherently transparent.

vides an answer to the question of why it should be that movement-related morphology occurs in the first place. However, closer inspection reveals that it also raises a number of technical questions, most notably the following two.

- How can the syntactic last resort effect be implemented technically?

- How does post-syntactic morphological spell-out in the C domain work technically?

In what follows, I address these two questions in turn.

Movement-Related Morphology as a Last Resort Operation in Syntax   The first observation to be made here is that under present assumptions intermediate steps of successive-cyclic movement by themselves are already last resort (or repair) operations:[74] The Edge Feature Condition states that an edge feature can only be inserted if it has an effect on outcome (i.e., if it provides the only way to yield a balanced phase), and if the phase head is still active. Crucially, free edge feature insertion is blocked on this view (as noted before, the present proposal departs from Chomsky (2007; 2008) in this respect). Recall from footnote 10 of chapter 3 that there is an underlying motivation for considering intermediate edge feature insertion a last resort operation: The insertion of edge features in the derivation violates the Inclusiveness Condition adopted in Chomsky (1995; 2000; 2001b; 2008) because material that is not part of the numeration is introduced in the course of the derivation.

As a first approximation, suppose that we postulate that (in some languages, and with some kinds of phase heads – e.g., with C in Modern Irish), a subcategorization feature for an expletive [●expl●] can be inserted in the course of the derivation (in minimal violation of the Inclusiveness Condition) if this is the only way to avoid a PIC violation with movement.[75]

---

[74] Note that I distinguish here between "last resort" as a concept (essentially, that of a repair), and Last Resort as a constraint (that demands all syntactic operations to involve discharge of structure-building or probe features). This terminological conflation may be considered somewhat unfortunate (especially in view of the fact that there is a tight relation between concept and constraint), but it should generally be harmless.

[75] Arguably, this repair operation can best be reconstructed, and thereby formally made precise, in optimality-theoretic terms. One possibility is the following (see Heck & Müller (2003; 2007) for the general framework): The Feature Condition and Last Resort are inviolable and outrank an *Empty Output Condition* (EOC), whose violation by an optimal output encodes the breakdown of the derivation (i.e., ungrammaticality). The EOC in turn outranks the Inclusive Condition. Note that this implies that there are actually three violations of the Inclusiveness Condition in cases of pseudo-melting as a last resort. First, as in all cases of intermediate movement steps, edge feature insertion violates Inclusiveness. Second, insertion of the subcategorization feature that triggers expletive insertion violates Inclusiveness. And third, actual external Merge of the expletive also violates Inclusiveness yet once more since the expletive cannot

However, this simple view raises a problem. As it stands, the system seems to require look-ahead: If a language has the option of saving an otherwise illegitimate extraction from a last-merged complement by expletive insertion (i.e., pseudo-melting), it has to make use of it: If it does not, no edge feature can be inserted for an item in the last-merged complement, and a PIC violation will subsequently arise. However, the derivation may continue for a while before this can actually be seen (e.g., subsequent phase heads also fail to insert an edge feature, and the PIC problem only arises at what might be the very last step, when the phase head of the ultimate landing site becomes active). Thus, given the PIC-based analysis, an Inclusiveness violation incurred by expletive insertion may be required early on but turns out to be legitimate only much later in the derivation.[76] From a more general point of view, basically the same problem arises with edge feature insertion as such. The solution that I proposed in chapter 3 is to restrict edge feature insertion to those contexts where a *balanced phase* cannot otherwise be produced; the underlying rationale is that a phase that is not balanced will invariably lead to a violation of the PIC later in the derivation. This way, non-local information is locally anticipated. In the case at hand, the same solution suggests itself for the look-ahead problem with expletive (feature) insertion: It is not forced non-locally, by the PIC (eventually, the Feature Condition and Last Resort, see the last footnote), but locally, by the need to have a balanced phase. Thus, we can postulate that a subcategorization feature for an expletive can be inserted in the course of the derivation (in minimal violation of the Inclusiveness Condition) if this is the only way to produce a balance phase (rather than to satisfy the PIC, as in the earlier version). As before, the requirement has to be relativized to certain kinds of heads, and to certain languages (so as to ensure that movement-related morphology is not available everywhere, and may help circumvent CED effects, in all languages).

One might hope that this is all that needs to be assumed about the syntactic repair mechanism of pseudo-melting via expletive insertion. Unfortunately, this is not the

---

be part of the numeration (see Hornstein (2001)). From an optimality-theoretic perspective (see Prince & Smolensky (2004), Smolensky & Legendre (2006)), the Inclusiveness Condition is a regular DEP constraint. We can then distinguish between two kinds of Inclusiveness violations: Edge feature insertion and expletive-subcategorization feature insertion violate DEP(feature), but Merge of the expletive violates DEP(category). In other words, there are two constraints that must all be ranked below the EOC in order for last resort pseudo-melting (hence movement-related morphology) to arise: Inclusiveness([●X●]) and Inclusiveness(XP). Furthermore, these Inclusiveness constraints have to be relativized to certain syntactic heads (C in the case at hand). Thus, eventually the relevant constraints are *contextual faithfulness* constraints (see Beckmann (1998) on positional faithfulness in phonology, and Woolford (2007) for a recent application of contextual faithfulness in syntax) of the following type: Inclusiveness([●X●])/C and Inclusiveness(XP)/C.

[76] This problem is well known in local derivational approaches to syntax, most notably in the domain of resumptive pronouns, but also in strictly derivational approaches to reflexive binding; see Fischer (2004; 2006).

case. Two further issues need to be addressed. First, the additional structure-building feature [•expl•] that creates pseudo-melting must show up *below* the regular (inherent) complement feature on a lexical item. This is just what had to be excluded with inserted edge features to ensure the derivation of CED effects; recall that the EFC in (11) states that [•X•] ends up *on top* of γ's list of structure-building features (a pushdown automaton property). The second issue concerns the nature of the subcategorization feature for the expletive: What exactly does [•expl•] stand for? How do we know what kind of feature has to show up here?

As it turns out, there is a solution that covers both problems simultaneously. Suppose that the [•expl•] feature arises as a result of *copying* the complement subcategorization feature ([•T•], in the case at hand): The higher copy then triggers complement Merge, and the lower feature triggers expletive insertion.[77] In general, "expletives" are minimal feature bundles grouped around various kinds of category labels (see Grimshaw (1997)), and the present analysis conforms to this view. Note that the copying option does not exist for edge features: Edge feature are specific features that either do not encode any further (category or other) information ([•X•]), or (given the remarks in subsection 4.3 above) encode information that is unrelated to the complement subcategorization feature's category label ([•X$_{wh}$•], for instance). Finally, one might think that it has to be specifically ensured that the expletive subcategorization feature resulting form last resort-driven coyping of a complement sucategorization feature can never attract a category by internal Merge. However, no additional stipulation (such as blocking the expletive feature from carrying out internal Merge) seems necessary here: Suppose that such an expletive subcategorization feature would effect internal Merge of some item in the complement (rather than external Merge of an expletive). Then, the insertion of the expletive would not lead to a balanced phase (that is otherwise unbalanced); it would therefore be precluded.

The relevant part of derivation that instantiates the pseudo-melting (hence, the occurrence of movement-related morphology) by expletive insertion is shown in (66).[78]

---

[77] The copies must be minimally different to ensure this order; we can assume that the lower one qualifies as defective in some sense.

[78] Note that a bit more must then be said to ensure that last-merged specifiers continue to be strict barriers. The sequence of feature-insertion and feature-deletion operations in (66) starts with a subcategorization feature [•β•] that may in principle either be one which merges a complement, or in fact one which introduces a (last-merged) specifier; and expletive insertion must be blocked in the latter case. Thus, the problem is that (63) can distinguish between the two contexts in the numeration, but not once the derivation has moved on – however, the feature copying mechanism responsible for expletive insertion, by assumption, takes place in the derivation. Various options arise. For present purposes, the simplest (though certainly not optimal) solution may be to postulate that the derivation knows whether a structure-building feature was initially topmost on a stack or not; e.g., by assigning features designated diacritics that encode their initial status (as

(66) *Pseudo-melting and movement-related morphology*:

	γ: $[\bullet\beta\bullet]$
→ (copying)	γ: $[\bullet\beta\bullet]$ ≻ $[\bullet\beta\bullet]$
→ (complement Merge)	γ: $[\bullet\beta\bullet]$
→ (edge feature insertion)	γ: $[\bullet X\bullet]$ ≻ $[\bullet\beta\bullet]$
→ (Move from compl.)	γ: $[\bullet\beta\bullet]$
→ (expl. insertion)	γ: $\emptyset$

This, then, is the gist of the syntactic part of the present analysis of movement-related morphology. There is one more complication, though: The approach sketched so far works well for languages which have movement-related morphology in all intervening domains, but not in the C domain associated with the final landing site. There are various languages of this type (despite McCloskey's (2002) claims to the contrary). Cf., e.g, *u* as a C realization in Wolof (see Torrence (2005)), and Heck & Müller (2003) for more general remarks. However, Modern Irish is *not* like this: Movement-related morphology also shows up in the final landing site. This poses a problem for the analysis: If C has a [●wh●] (or [●rel●]) feature left at the point where TP is merged, TP cannot be a barrier for extracting an item out of it. Various options arise to reconcile the occurrence of movement-related morphology on the head of the final landing site in Modern Irish with the present approach. For one thing, it might be that the final landing site of movement is actually *above* CP, in a separate projection. For another, it might be that what looks like the moved item is in fact base-generated; the moved item is an empty operator (this is actually McCloskey's (2002) own approach). Under this view, it might simply be that the empty operator triggers the same kind of default spell-out as the pseudo-melting expletive. Finally, one might speculate that, e.g., C has a [∗wh∗] probe feature but no [●wh●] feature; the probe feature alone may be classified as "too weak" to keep a head active once the complement has been merged.[79]

Having answered the question of how movement-related morphology may arise as a repair phenomenon in syntax, let me now turn to the second question posed on page 223: How does morphological spell-out work?

Morphological Spell-Out   The fundamental question that needs to be addressed is this: How can it be possible that the expletive (EXPL) that is postulated in the syntactic account just sketched, and the C head are morphologically realized as a single unit?

---

a complement-creating or specifier-creating feature).

[79] Perhaps the [∗wh∗] would have to be assumed to differ from other probe features in this respect. Alternatively, probe features in general might be insufficient to maintain activity of a phase head; this would then require an analysis of transparent last-merged complements as always involving pseudo-melting.

There are two basic options that suggest themselves. One is to assume a post-syntactic spell-out of subtrees; the other one is to postulate post-syntactic fusion of EXPL and C before spell-out of single nodes.

Spell-out of subtrees would for the most part work without problems for Modern Irish: At the point where morphological realization of EXPL+C is to take place, the originally intervening δ has already gone on to the next phase, and EXPL and C have become adjacent. This is shown by the two steps of the abstract derivation in (67).[80]

(67) a.   [$_{CP}$ EXPL [$_{C'}$ δ_1 [$_{C'}$ C [$_{TP}$ t$_1$ ...]]]] $\rightarrow$ ... $\rightarrow$
      b.   ... δ ... [$_{CP}$ EXPL [$_{C'}$ t$'_1$ [$_{C'}$ C [$_{TP}$ t$_1$ ...]]]]

However, there is a potential complication. As noted, movement-related morphology in Modern Irish also shows up in the final landing site. For this last step, it might be that actual dislocation of the expletive to the position of C is required (otherwise, the moved item would intervene; but recall the various options discussed in the preceding paragraph *Movement-Related Morphology as a Last Resort Operation in Syntax*). Even though this problem strikes me as solvable, I will adopt the second option in what follows, and assume that there is post-syntactic fusion of EXPL and C. Consider the (partial) derivation in (68), where (68-a) and (68-b) are outputs that are generated in the syntax, and (68-c) is achieved by postsyntactic fusion of EXPL and C.

(68) a.   [$_{CP}$ EXPL [$_{C'}$ δ_1 [$_{C'}$ C [$_{TP}$ t$_1$ ...]]]] $\rightarrow$ ... $\rightarrow$
      b.   ... δ ... [$_{CP}$ EXPL [$_{C'}$ t$'_1$ [$_{C'}$ C [$_{TP}$ t$_1$ ...]]]] $\rightarrow$
      c.   ... δ ... [$_{CP}$ − [$_{C'}$ t$'_1$ [$_{C'}$ EXPL+C [$_{TP}$

Morphological realization of EXPL+C in Modern Irish can then be assumed to take place in the following way. Given fusion, the feature specifications of EXPL and C are combined into a single feature specification that is targetted only once by spell-out. Since the feature specifications of EXPL and C are different from the original C specification, it is likely that different morphological exponents will realize the two specifications: This accounts for the movement-related morphological effect. A simplified analysis that illustrates the basic working of variable morphological realization of C in contexts with and without movement in Icelandic is given in (69).

(69) a.   C bears the features F_1, F_2.
      b.   EXPL bears the feature F_3.
      c.   EXPL+C bears the features F_1, F_2, F_3.
      d.   *go* $\leftrightarrow$ F_1, F_2.
      e.   *aL* $\leftrightarrow$ F_1, F_2, F_3.

---

[80] Recall that traces are given here just for convenience; in the actual derivations, there is simply nothing in those places where the traces occur.

Suppose that the realization of bundles of morpho-syntactic features by morphological exponents follows the Subset Principle, a version of which is given in (70).[81]

(70) *Subset Principle*:
   A vocabulary item *V* is inserted into a functional morpheme *M* iff (i) and (ii) hold:

   (i)   The morpho-syntactic features of *V* are a subset of the morpho-syntactic features of *M*.
   (ii) *V* is the most specific vocabulary item among those that satisfy (i).

The notion of specificity can (for present purposes) be understood as in (71).

(71) *Specificity of Vocabulary Items*:
   A vocabulary item V_i is more specific than a vocabulary item V_j iff V_i has more (relevant) morpho-syntactic features than V_j.

As a consequence, *go* is the most specific compatible exponent for C alone, and it is therefore chosen via the Subset Principle if EXPL is not present. Note that *aL* is more specific, but not compatible because it has an additional feature which implies that it cannot realize a subset of the syntactic context, i.e., the specification in C. In contrast, *aL* is the most specific compatible exponent for EXPL+C (because it *does* realize a subset of the morpho-syntactic features of the syntactic context now), and it is therefore chosen via the Subset Principle if EXPL has undergone fusion with C.

   So far, so good. The analysis works well enough technically as it stands, but it does not yet capture a generalization proposed by Lahne (2009, 60) on the basis of a cross-linguistic study of pieces of movement-related morphology. Lahne observes that, across languages, movement-related morphology typically involves a *retreat to the general case* (in the sense of Halle & Marantz (1993)): A less specific (default) exponent is inserted as the reflex of movement. In contrast, on the view embodied in (69), pieces of movement-related morphology are more specific (EXPL+C has more features than just bare C).[82]

---

[81] See Kiparsky (1973), Lumsden (1992), Halle (1997), and Williams (1997), among many others, for versions of this constraint, sometimes under different names. The present version most closely follows Halle (1997).

[82] Lahne's (2009) own approach works as follows. A basic assumption is that it is the edge feature itself that gives rise to movement-related morphology. However, edge features that play a role in the course of successive-cyclic movement are not assumed to be primitives; rather, an *edge feature property* is assigned to some existing feature(s) on the phase head, and combines with the feature(s) that it targets, thereby creating a full-fledged edge feature. By triggering movement, the 'fused' edge feature is deleted, and with it a part of the original feature specification of the phase head. The feature that is deleted (as part of the edge feature) is not available for post-syntactic morphological realization anymore; thus we are dealing with an instance of

Let us assume that Lahne's generalization is correct. This may then be taken to suggest looking for an alternative. Here is one: Suppose that the Subset Principle is replaced by the *Superset Principle* (see Starke (2006; 2009), Caha (2007; 2008), and Taraldsen (2008)). In Superset Principle-based approaches to morphological realization, default exponents are those with more features (rather than those with fewer features). An informal statement of the Superset Principle is given in (72).

(72) *Superset Principle*:
    The phonological exponent of a vocabulary item is inserted into a node if the item matches all or a superset of the grammatical features specified in the node. Insertion does not take place if the vocabulary item does not contain all features present in the node. Where several items meet the conditions for insertion, the item containing fewer features unspecified in the node must be chosen.

The working of the Superset Principle can be illustrated with inflection of *be* in English (the discussion follows Caha (2007)). (73) is the paradigm that needs to be derived.

(73)

	Singular	Plural
1	am	are
2	are	are
3	is	are

As in other morphological approaches, syncretism is derived by decomposing standard grammatical categories into combinations of more primitive features – in the present context, by decomposing person and number features as in (74).

(74) *Feature decomposition*:
    participant, addressee, group, present

The three morphological exponents in (73) can then be assumed to be associated with the features shown in (75).

(75) a. /am/ ↔ [pres,part]
    b. /is/ ↔ [pres]
    c. /are/ ↔ [pres,part,addr,group]

Given the further assumption that the morpho-syntactic feature bundles for 1., 2., and 3. person, and for singular and plural, already differ in the syntax, the Superset Principle makes the correct predictions. The competition and its outcome are shown in

---

feature impoverishment in the syntax. Consequently, a less specific, default exponent is chosen by spell-out, as in other cases of impoverishment.

(76).

(76)	Features	Ø	Group
	part	am ~~are~~	are
	part,addr	are	are
	Ø	is ~~am~~ ~~are~~	are

By assumption, 3. person and singular are not marked at all in the syntax. Plural is marked by the feature [group], 1. person is marked by the feature [part(icipant)], and 2. person is marked by the features [part] and [addr(essee)]. Given the Superset Principle, /are/ fits everywhere (because it is maximally specified) but is blocked in 1. and 3. person singular contexts because more specific markers are available that also realize a superset of the syntactic specification; and /am/ fits in 1. and 3. person singular contexts but is blocked in the latter because a more specific marker is available here.

Thus, it is a property of Superset Principle-based approaches that the default exponent is overspecified; that the exponents that block its insertion in certain environments have fewer features; and that the feature specification of a morphological exponent does not have to be coherent, or capture a natural class of instantiations of a grammatical category. All this is in sharp contrast to Subset Principle-based approaches. With this in mind, let us now go back to movement-related morphology.

It turns out that Lahne's (2009) generalization can be accomodated in the present system if the Superset Principle governs morphological realization: Fusion of EXPL and C (yielding the set $\{F_1,F_2,F_3\}$) blocks the regular exponent for C (specified as, say, $\{F_1,F_2\}$, which is a perfect match for C alone) which now does not realize a superset of the feature specification of EXPL+C anymore, and forces insertion of the less specific default exponent (which does realize a superset of EXPL+C – minimally, by being specified as $\{F_1,F_2,F_3\}$ – but is blocked in bare C contexts as not being specific enough). The only thing that remains to be done at this point to give a complete analysis of the Modern Irish pattern of movement-related morphology in the C domain is to give some content to the abstract features F_1, F_2, and F_3 that feature in the analysis. However, since this issue is (a) secondary to my main objectives here, and would (b) require an in-depth study of the complementizer and verb inflectional system of Modern Irish, I will refrain from attempting such a task here, and rather turn to a second instance of movement-related morphology, viz., *MeN* deletion in Malay.

7.3.3 *MeN-Deletion in Malay*

Bahasa Melayu is a language with optional *wh*-in situ and *wh*-movement strategies. Moreover, *wh*-movement may be partial (in the sense that it may stop at any intermdiate SpecC position); see the examples in (77). Cole & Hermon (2000, 110) observe that overt movement of a *wh*-argument DP or DP operator requires the omission of the

meN-prefix on the verb in Bahasa Melayu.[83]

(77) a. Ali mem-beritahu kamu tadi [Fatimah mem-baca apa$_1$] ?
          Ali meN-told   you just   Fatimah meN-read what

      b. Apa$_1$ (yang) Ali (*mem)-beritahu kamu tadi [(yang) Fatimah
          what   that Ali meN-told   you just   that Fatimah
          (*mem)-baca t$_1$] ?
            meN-read

      c. Ali (mem)-beritahu kamu tadi [apa$_1$ (yang) Fatimah (*mem)-baca t$_1$] ?
          Ali (meN)-told   you just   what that Fatimah   meN-read

Here is a sketch of a pseudo-melting analysis. Suppose that a functional projection
AspP intervenes between vP and TP in Malay. Suppose further that Asp does not
status-govern vP (more generally, there is no probe feature on Asp that would be left
after Merge of vP), and that by itself, Asp does not take a specifier. Given the possibil-
ity of Inclusiveness violations on the Asp cycle in Bahasa Melayu, expletive insertion
is possible in SpecAsp if this is the only way to balance the AspP phase. This creates
a transparency of vP: A pseudo-melting effect arises. As for the morphological part
of the analysis, assume that *meN* is the standard realization of Asp. Fusion of EXPL
and Asp in movement contexts then makes *meN* insertion impossible (via the Superset
Principle). Consequently, a zero default marker is inserted (as in Lahne (2009)); but
the default marker is radically overspecified (given the Superset Principle).[84] A sim-
plified analysis of the occurrence of movement-related morphology in Bahasa Melayu
is given in (78) (as before, I will remain silent as to the possible denotations of F_1, F_2,
and F_3.)

(78) a. Asp bears the features F_1, F_2.
      b. EXPL bears the feature F_3.
      c. EXPL+Asp bears the features F_1, F_2, F_3.
      d. *meN* $\leftrightarrow$ F_1, F_2.
      e. Ø $\leftrightarrow$ F_1, F_2, F_3

7.3.4 Complementizer Selection in Kinande

As a last case study, let me briefly address the occurrence of movement-related mor-
phology in the C domain of the Bantu language Kinande, as described in Schneider-
Zioga (2005); cf. (79). The relevant observations are the following. First, C and EXPL
may be subject to separate spell-out: In (79-c), C is realized as *nga*, and EXPL as

---

[83] It therefore looks as though *wh*-in situ does not involve overt movement of some (possibly abstract) item;
but cf. Fanselow & Ćavar (2001) for a reanalysis.

[84] Alternatively, there may simply be no exponent available for insertion.

kyo. Second, there is *wh-agreement*, in the sense that the pieces of movement-related morphology differ, depending on properties of the moved item.

(79) a. [$_{CP}$ Ekihi$_1$ [$_C$ kyo] Kambale alangira t$_1$] ?
               what   that Kambale saw
           'What did Kambale see?'

      b. [$_{CP}$ Iyondi$_1$ [$_C$ yo] Kambale alangira t$_1$] ?
               who   that Kambale saw
           'Who did Kambale see?

      c. [$_{CP}$ Ekihi$_1$ [$_C$ kyo] Marya alengekanaya [$_{CP}$ t$_1$ nga-kyo Kambale
               what   that Mary thinks   C-that Kambale
           asoma t$_1$]] ?
           read
           'What does Mary think that Kambale read?'

A pseudo-melting analysis of these data could look as follows. First, to account for (79-c), it can be assumed that EXPL does not necessarily attach to C by *fusion*; an operation like *merger*, *lowering* or *local dislocation* (that leaves the two feature bundles intact) is also an option (see Embick & Noyer (2001)). The occurrence of two separate realizations as such can be seen as independent motivation for the approach in terms of an EXPL category as the source of movement-related morphology adopted here. Second, the agreement effect suggests that properties of the moved item may also influence morphological realization. This is to be expected if movement leaves traces after all – if so, (a) a moved item's trace may potentially retain morpho-syntactic features, and (b) this trace may also participate in fusion (with only EXPL, in the case at hand). In a trace-less approach like the one I am adopting here, the *wh*-agreement effect requires the postulation of an additional effect of feature manipulation exerted by the intermediately-moved *wh*-phrase on the C head that it is attracted by.[85]

7.4 *Conclusion*

This may suffice as an illustration of how instances of movement-related morphology in various kinds of languages can be addressed in terms of pseudo-melting. As noted above, I take the main virtue of the present analysis to be that it offers an explanation of why this phenomenon should exist in the first place. It can also be noted that the present analysis (a) suggests itself as a natural extension of the approach to CED effects developed in this chapter; and (b) offers a new perspective on the phenomenon of morphological "reflexes" of successive-cyclic movement that brings the phenomenon more in line with standard approaches to other cases of correlations of

---

[85] An alternative would be to assume that different base positions of the moved items are relevant; see Lahne (2009) on the intricate *wh*-agreement patterns in Chamorro.

syntactic and morphological properties. Needless to say, the new approach also raises a lot of new questions. Furthermore, strong empirical evidence that would favour it over the more established approach that treats movement-related morphology as a pure reflex of movement so far seems to be outstanding.[86]

[86] For instance, the present approach would be strongly supported by the discovery of evidence showing that movement-related morphology in some domain of some language is absent with categories that can independently be shown to be transparent, and shows up with categories that can independently be shown to be opaque as such.

Chapter 5

Operator Island Effects

1. Introduction

In chapter 1, I made the assumption that the MLC and Relativized Minimality should not *both* be assumed as primitive locality constraints in a theory of grammar; the constraints are repeated here (in versions that highlight the similarity) as (1) and (2).

(1) *Minimal Link Condition* (MLC):
　　In a structure α ... [... β ... [... γ ...] ...] ..., movement to a position δ in the minimal domain of α cannot affect γ if β c-commands γ and *β and γ bear the same feature required by α.*

(2) *Relativized Minimality* (derivational version):
　　In a structure α ... [... β ... [... γ ...] ...] ..., movement to a position δ in the minimal domain of α cannot affect γ if β c-commands γ and *β and δ are of the same structural type*.

Despite the main difference that the MLC talks about features where Relativized Minimality talks about positions, the two requirements are conceptually similar (as intervention constraints), and they may lead to redundancies empirically. In view of this, in chapter 1 I concluded that Relativized Minimality should be abandoned in favour of the MLC, not least of all because the MLC can be generalized to the (G)MLC, by incorporating the F-over-F Principle. Still, in chapter 2 I argued that the (G)MLC is empirically problematic, and also raises conceptual questions from the point of view of a strictly derivational approach to syntax (because it requires a significant amount of search space); the conclusion then was that the (G)MLC should ultimately be abandoned too, with its main effects following from the interaction of the PIC and the (revised) Edge Feature Condition (chapter 3). Clearly, the conceptual arguments against the (G)MLC carry over unchanged to Relativized Minimality – this constraint needs even more search space than the (G)MLC.

　　Against this background, the question arises of how typical Relativized Minimality effects can be derived under present assumptions. The goal of this chapter is to inves-

tigate operator island effects (*wh*-island effects and topic island effects) on the basis of the PIC-approach to locality constraints developed in the two preceding chapters.[1]

The conclusion for operator islands will be that the Intermediate Step Corollary of chapter 4 makes a new approach to A-bar intervention effects possible (in which the notion of intervention plays no role). The analysis is based on the concept of feature maraudage introduced in Georgi, Heck & Müller (2009), and I argue that it turns out to be empirically superior to Relativized Minimality in the domain of operator island effects in German (which show a kind of principled variability that is unexpected under Relativized Minimality).

2.   Background

2.1   Operator Islands without Intervention

Consider again some sentences illustrating the *wh*-island effect, as they were discussed before (cf. the examples in (3) and (104) from chapter 1).

(3) a. ?*[$_{DP_1}$ Which book] do you wonder [$_{CP}$ [$_{PP_2}$ to whom] John gave t_1 t_2] ?
     b. *How$_1$ does she know [$_{CP}$ [$_{DP_2}$ which car] Mary fixed t_2 t_1] ?

Relativized Minimality accounts for the illformedness of (3-ab) because movement of XP_1 crosses an intervening XP_2 that occupies a position of the same structural type (A-bar). As noted in chapter 1 (page 61), the (G)MLC can account for *wh*-island effects if it is assumed that the *wh*-item in the embedded SpecC position (i.e., the position that creates the island) still has a feature [wh] that may block attraction of the lower *wh*-phrase by the matrix interrogative C head even though this feature has been checked in the final landing site.

However, closer inspection reveals that the (G)MLC can in fact not account for *wh*-island effects anymore if assumptions are made that have become standard in more recent minimalist work, and that also figure prominently in chapters 3 and 4 above:[2] First, there is no reason to limit the number of specifiers of a head to one. And second, given the PIC (even with a less local concept of what a phase is than the one adopted in chapters 3 and 4), successive-cyclic movement must systematically be able to target specifiers of phase heads, even if there is already another specifier around. Thus, even

---

[1] I will have nothing to say about super-raising constructions. I will also remain silent on certain kinds of interventions effects with quantified items and negation, for which semantics-based accounts might suggest themselves; see the references listed in footnote 44 of chapter 2. Finally, I will make no attempt to derive Head Movement Constraint (HMC) effects – as a matter of fact, the analysis of a restriction on movement from embedded verb-second clauses in chapter 6 presupposes that the HMC is not valid as such.

[2] Strictly speaking, similar conclusions apply if Relativized Minimality is combined with these assumptions.

the simplest cases of successive-cyclic *wh*-movement, such as (4-ab) (in English and German, respectively, with the option of an in-situ subject in the latter case), must proceed via an additional specifier of v.

(4) a. (I wonder) [$_{CP}$ what$_1$ C [$_{TP}$ she$_2$ T [$_{vP}$ t$'_1$ [$_{v'}$ t$_2$ said-v [$_{VP}$ t$_V$ t$_1$]]]]]

   b. (Ich frage mich) [$_{CP}$ was$_1$ C [$_{TP}$ [$_{vP}$ t$'_1$ [$_{v'}$ Fritz [$_{VP}$ t$_1$ t$_V$] denkt-v]] T]]
       I wonder   what   Fritz   thinks

Under a multiple specifier/PIC approach, there is no obvious reason why an intermediate movement step of a *wh*-phrase to an outer specifier of an interrogative C (with another *wh*-phrase acting as an inner specifier) should be blocked: Informally speaking, the former *wh*-phrase that is on its way out of the CP does not want to see a [●wh●] feature (yet); it just wants to see what all items want to see that leave the clause, viz., an edge feature (under the assumptions made here). This becomes particularly clear under the approach to successive-cyclic movement adopted in chapter 3 above: Edge feature insertion on an interrogative C head that itself attracts a *wh*-phrase should be entirely unproblematic because, with one interrogative C$_{[●wh●]}$ waiting in the numeration, insertion of an edge feature on the other interrogative C$_{[●wh●]}$ head that is already part of the structure is the only way to satisfy phase balance; hence it should be permitted by the Edge Feature Condition. Thus, given successive-cyclic movement via phase edges, an intermediate movement step to an outer specifier (under standard assumptions, not those of chapter 4) of a CP with a *wh*-phrase in SpecC is not excluded. This, in turn, means that the *wh*-phrase that has reached its ultimate target position in the embedded CP does not intervene between the eventual landing site (or the next phase edge) and the other (intermediately moved) *wh*-phrase. More generally, the (G)MLC cannot derive *wh*-island effects anymore.

The situation is similar under the approach to (G)MLC effects developed in chapter 3: As such, there is nothing that would preclude insertion of an edge feature for the second *wh*-phrase, which moves this *wh*-phrase to the edge domain of the embedded interrogative C, where it is accessible for further movement operations.

In view of this, the only solution, it seems, would be to invoke an ad hoc stipulation that simply states that an edge feature cannot be inserted on C$_{[●wh●]}$. This would then make it possible to derive *wh*-island effects under both the (G)MLC account (because the *wh*-phrase in SpecC will then invariably be closer to the ultimate landing site) and the PIC-based account laid out in chapter 3 (because the second *wh*-phrase will then not be able to reach the edge domain of the embedded CP, and a PIC violation will ensue on the next – i.e., the VP – cycle). However, fortunately it turns out that adopting such a defeatist position would be premature.

The claim that I would like to defend in what follows is that the concept of feature *maraudage* that is argued for as a general theoretical option in Georgi et al. (2009) derives certain operator islands (viz., *wh*-islands and topic islands) and their selec-

tive nature, without recourse to a constraint like Relativized Minimality (see Rizzi (1990)) or the Minimal Link Condition (see Chomsky (2001b)) – but a precondition for this analysis to work is that assumptions about intermediate steps of successive-cyclic movement are made that correspond to those of chapter 4.

More specifically, the proposal is this: *Wh*-islands and topic islands do not instantiate an intervention effect; neither does the operator block an escape hatch (as in Chomsky (1977; 1986a)).[3] Rather, the item that is to be long-distance moved marauds the set of A-bar-related features on C *before* C can attract a *wh*-phrase or topic that would erect the island. The option for such a maraudage does not arise under the standard assumption that successive-cyclic movement steps target the *outer* specifiers of phase heads; but it does arise under an approach in which intermediate steps of successive-cyclic movement must not target the outermost specifier of C (if there is more than one). This is just what the approach to CED effects developed in chapter 4 forced us to assume (cf. the *Intermediate Step Corollary*); to the extent that this approach then permits a simple, non-intervention-based account of operator island effects that is not available under the standard approach, it is independently confirmed.

As far as I am aware, there are only two other structural (and syntactic) analyses of operator islands that do not rely on the idea of intervention, or the blocking of an escape hatch. As an excursus, let me briefly indicate the workings of these two approaches before presenting the new approach in terms of maraudage.

2.2 A Tree Adjoining Grammar Approach

The first such analysis is the Tree Adjoining Grammar approach to *wh*-islands developed in Kroch (1989) and Frank (2002).[4] Here, the basic premise is that all long-distance dependencies must be brought about by (counter-cyclic) insertion of so-called auxiliary trees that pump up the local tree structure generated thus far ('elementary trees'); e.g., (5-a) is derived by inserting (5-b) (where *think* subcategorizes for a C′ category) into the C′ node of (5-c), which only has local movement to the minimal SpecC position.

(5) a.   Who$_1$ do you think she likes t$_1$?

---

[3] Note incidentally that *wh*-islands are derived from (a version of) the CED in Chomsky (1986a): Simplifying somewhat, TP is considered a potential barrier (a 'blocking category'; cf. footnote 10 in chapter 2) that transmits its status as a potential barrier to the next higher XP, i.e., CP, where barrier status is then realized. (The reason is that TP is not L-marked by a lexical category; note that an embedded interrogative CP *is* L-marked by the matrix verb, so it does not qualify as a barrier in and of itself.) On this view, the only way to escape from TP is to move to SpecC: A blocking category can transmit barrier status for some item only if that item is included in it. Of course, concepts like 'blocking category' (TP) and 'barrier by inheritance' (CP, in the case at hand) are highly dubious under minimalist assumptions.

[4] Also see Unger (2010, 71) for a computational-minimalist approach incorporating similar ideas.

b. $[_{C'}$ do you think C']   (auxiliary tree)

c. $[_{CP}$ who$_1$ $[_{C'}$ C she likes t$_1$]]   (elementary tree)

On this view, a *wh*-island construction like (6-a) would require inserting ('adjoining') a recursive auxiliary tree like (6-b) into the elementary tree (6-c); and the problem is that a non-multiple-*wh*-movement language like English does not permit trees of the type in (6-c) that would be required to feed a *wh*-island construction.

(6)   a.   ?*What book$_1$ did John ask whom you had given t$_1$?

b. $[_{C'}$ did John ask C']

c.   *$[_{CP}$ what book$_1$ $[_{C'}$ whom$_2$ $[_{C'}$ C you had given t$_2$ t$_1$]]]

2.3   An Optimality-Theoretic Approach

The other approach that does without intervention or blocking of an escape hatch is the optimality-theoretic approach to *wh*-islands in Legendre, Smolensky & Wilson (1998). The approach relies on violable and ranked constraints; in brief, it works as follows. All instances of extraction from a CP (even a complement CP) are assumed to violate locality constraints (most phrases are barriers), but such constraint violation is possible if it is (a) minimal, and (b) forced by higher-ranked constraints. By assumption, the constraint that forces *wh*-movement to the target *wh*-scope position (the analogue of the Feature Condition in the present approach) cannot play this role: It is ranked lower than the locality constraint that an item (more precisely, an adjunct) violates when it undergoes extraction from an object CP. However, a third constraint that is highest-ranked (in effect: inviolable) demands that (lexically determined) selection requirements are respected. One such requirement is that a V that selects an interrogative CP finds a *wh*-item in the embedded SpecC (or C) position. With this in mind, consider the contrast in (7-ab).

(7)   a.   How$_1$ do you think $[_{CP}$ Mary fixed the car t$_1$] ?

b.   *How$_1$ does she know $[_{CP}$ $[_{DP_2}$ which car] Mary fixed t$_2$ t$_1$] ?

By assumption, either there are no intermediate traces, or intermediate traces can be established in the embedded SpecC positions of both sentences (which we expect given the option of multiple specifiers; see above). Therefore extraction of the adjunct violates locality constraints on movement in exactly the same way in (7-a) and (7-b) – thus, there is no intervention effect induced by *which car* in (7-b). Rather, the reason why (7-b) is ungrammatical is that it is blocked by a competing candidate with a better constraint profile: (8) violates the constraint that would normally trigger *wh*-movement to the matrix clause, but since this constraint is ranked lower than the locality constraint violated with extraction from all CPs, (8) is the optimal candidate.

(8) #She knows $[_{CP}$ how$_1$ $[_{DP}$ which car$_2$] Mary fixed t$_2$ t$_1$]

(8) does not look like a well-formed English sentence, but Legendre et al. (1998) assume (based on the copy theory of movement) that *wh*-in situ may involve multiple *wh*-movement with selective PF realization of copies, such that one of the traces (rather than its antecedent) is PF-realized in English (also see Pesetsky (2000), Grewendorf (2001), and footnote 25 of chapter 3).

Why, then, is (7-a) possible after all, and not blocked as suboptimal by a competing candidate that has the *wh*-phrase staying in the embedded clause throughout the derivation, just like (8) does? The answer is that the matrix verb in (7-a) selects a declarative CP complement, and not an interrogative CP complement; but if the *wh*-phrase *how* stays in the embedded clause, the embedded clause will have to be interpreted as a *wh*-clause. This would violate the highest-ranked selection requirement. Consequently, long-distance movement in (7-a) is the only option, and optimal, even though it violates the locality constraint that is always violated with (adjunct) extraction from CP (and whose violation is fatal in (7-b)).

This approach fully exploits the hypothesis that constraints may be violable, and thereby manages to do without the notion of intervention in an account of *wh*-islands. I take this analysis to be simple and elegant, and I think it would certainly be worth to pursue the general line of reasoning further. However, in the present context, we may note that Legendre et al.'s (1998) approach is radically incompatible with the local, derivational framework adopted throughout this monograph, where only a small portion of syntactic structure is available for constraint evaluation at any given step; and that it is far from clear whether the analysis can be extended to topic islands in a straightforward way. Be this as it may: The analysis that I will develop below will turn out to be quite different from the one just sketched; apart from the lack of a concept of intervention, there is very little that the two approaches have in common.

3. Assumptions

3.1 The Intermediate Step Corollary

The new analysis of operator islands rests on two assumptions. The first assumption is that intermediate movement steps end up in non-outermost specifiers (if there is another specifier). As extensively argued in chapter 4, this follows from the approach to CED effects based on the interaction of the revised Edge Feature Condition and the PIC. In a nutshell, intermediate movement to a phase edge requires the insertion of edge features, and given the Edge Feature Condition in (11) of chapter 4, the phase had must have a structure-building feature (that introduces a specifier) left at this point.[5]

---

[5] I abstract away here from probe features that suffice to keep a phase head active for complements (but not for specifiers). In operator island contexts, this option is irrelevant.

Given the last-in/first-out property of feature discharge, this implies that the newly inserted edge feature is then discharged first, and the original structure-building feature is discharged last. I have referred to this consequence of the theory as the Intermediate Step Corollary; it is repeated from (15) of chapter 4 in (9).

(9) *Intermediate Step Corollary*:
    Intermediate movement steps to specifiers of X (as required by the PIC) must take place before a final specifier is merged in XP.

An instance of successive-cyclic movement of a DP through the extended verbal domain then looks as in (10) (= (16) of chapter 4).

(10) DP_2 ...
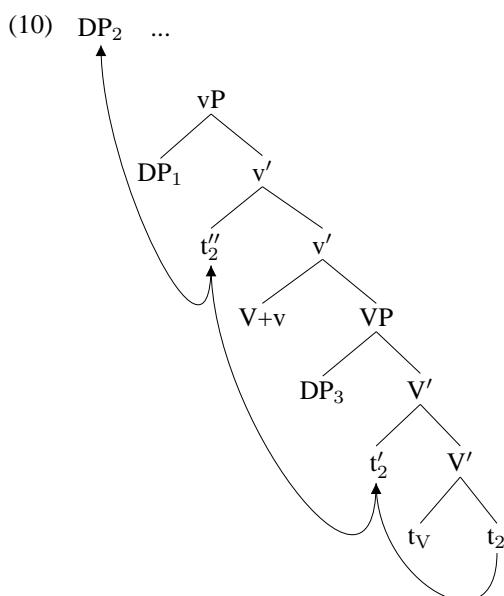
3.2   Maraudage

The second assumption concerns maraudage. Georgi, Heck & Müller (2009) propose the concept of feature maraudage in (11).

(11) *Maraudage*:
    Certain goal features of Agree or Merge operations (among them [person] and [animate]) are checked if the structural conditions for checking are met.

In Georgi, Heck & Müller (2009), syntactic contexts are considered where a head ζ serves two items α and β (i.e., both α and β need to carry out a feature-based operation with ζ), and the item that first enters into a relation with ζ (say, α) takes more out of the set of features than ζ originally provides for it – it "marauds" ζ's feature

set and discharges features that were designed for the other item (β).[6] When the remaining item β finally enters the structure, ζ has lost the features it would need for β. As shown in Georgi et al. (2009), such an instance of maraudage may have various effects, depending on the kinds of features that are involved, and on the specific syntactic domain in which maraudage occurs. Evidence is presented from the areas of agreement, case assignment, and movement.

Based on Béjar & Řezáč (2009), it is argued that certain facts in the domain of *ergative displacement in Basque* are best viewed as instantiations of secondary repair operations that restore the feature set for an external argument DP after an internal argument DP has marauded v's ϕ-feature set for the external argument. With respect to case, it is shown that the phenomenon of *global case splits* (see Silverstein (1976)) can be reconciled with a derivational bottom-up syntax without look-ahead or backtracking (and very small accessible domains at each step of the derivation) if it is assumed that a differentially case-marked internal argument DP marauds the ϕ-feature set of v, which leads to severe restrictions on the type of external argument DP that v can subsequently combine with.[7] Finally, as regards *movement*, maraudage offers a new approach to operator island effects: Given the Intermediate Step Corollary, the item that is supposed to undergo successive-cyclic extraction from a CP arrives in the em-

---

[6] Chomsky's (2001b) Maximize Matching Effects is arguably similar in nature. Also compare Abels' (2003) account of anti-locality effects (incompatible though it is with the present approach), which presupposes that goal features must be checked whenever a configuration that permits checking exists, even if that leads to certain options not being available that would otherwise be; and Anagnostopoulou (2005), Adger & Harbour (2007) and Richards (2008) on the Person Case Constraint.

[7] In global case splits, an internal argument DP is differentially case-marked if both the internal argument DP and the external argument DP are atypical from the perspective of Hale/Silverstein scales (see Hale (1972), Silverstein (1976)): For instance, the internal argument is 1. person, and the external argument is 3. person inanimate. Alternatively, an external argument DP is differentially case-marked if both arguments qualify as atypical in this sense. Especially the first case poses an interesting problem for derivational theories without look-ahead or backtracking: At the point where the question of differential case-assignment to an internal argument DP must be decided, the external argument DP is not yet part of the structure, assuming an Earliness principle (see Pesetsky (1989) and the references given on page 194 above) and/or the Strict Cycle Condition (e.g., in the formulation given in (23) in chapter 4). In line with this, all previous accounts of global case splits permit themselves to simultaneously scan properties of both the external and the internal argument DP in order to decide the question of whether differential case marking does nor does not apply; see Aissen (1999), de Hoop & Malchukov (2008), Béjar & Řezáč (2009), and Keine (2010). Consider, in contrast, the maraudage approach developed for global case splits in Yurok in Georgi (2010), Georgi et al. (2009): In Yurok, an internal argument DP_{int} is accusative-marked only if it is higher on the person scale than an external argument DP_{ext} – i.e., if DP_{int} is 1. or 2. person, and DP_{ext} is 3. person. On Georgi's analysis, this is so because v expects DP_{int} to be third person (the unmarked case for internal arguments); if DP_{int} is 1./2. person, it needs additional features (given an approach to person features as in Harley & Ritter (2002)), and consequently marauds the ϕ-features reserved for DP_{ext}. This maraudage operation is ultimately held responsible both the the occurrence of differential case marking and

bedded C domain first and may discharge more features in this position than just the edge feature that was inserted for it; in particular, it may maraud the embedded C's original feature set and thereby make subsequent Merge of the regular specifier of C impossible. However, before delving further into the maraudage approach to operator islands, let me first introduce some data from German that shed some light on the status of operator islands as variable or strict (i.e., as *weak* or *strong*).

4. Wh-Islands and Topic Islands

As noted in section 3.6 of chapter 1, extraction from *wh*-islands may sometimes lead to full ungrammaticality, and may sometimes yield more variable results, depending on a variety of factors (argument vs. adjunct, finite vs. non-finite *wh*-clause, and the type of landing site that the extracted item ends up in). In contrast, extraction from a topic island is much less subject to variation in the degree of ungrammaticality. For German, the most basic generalization would seem to be this (see Fanselow (1987) and Müller & Sternefeld (1993)): First, *wh-islands* block *wh*-movement but not (argument) topicalization in German. And second, *topic islands* block both *wh*-movement and topicalization in German.[8]

The variable nature of *wh*-islands for *wh*-movement vs. topicalization in German is shown in (12) (see Fanselow (1987, 57), Müller & Sternefeld (1993, 485), and Grewendorf (2002, 88)).

(12) a. *Welches Radio$_1$ weißt du nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_1$ t$_2$ repariert]] ?
        which radio know you not   how   one   fixes

    b. ?Radios$_1$ weiß ich nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_1$ t$_2$ repariert]]
        radios know I   not   how   one   fixes

The strict nature of topic islands for both *wh*-movement and topicalization in German is illustrated by the examples in (13).

(13) a. *Welches Radio$_1$ glaubst du [$_{CP}$ der Maria$_2$ [$_C$ hat] [$_{TP}$ er t$_2$ t$_1$
        which book think you   the Mary   has   he

---

the restriction of such case marking to contexts where DP$_{ext}$ is 3. person.

[8] This is reminiscent of the difference between relativization and *wh*-movement from *wh*-islands in Italian (see Rizzi (1982)), where only the former operation yields acceptable results; recall subsection 2.2 of chapter 1. A relevant pair of examples is repeated here in (i).

(i) a. Tuo fratello [$_{CP_3}$ [$_{PP_1}$ a cui] mi   domando [$_{\mathbf{CP_4}}$ [$_{DP_2}$ che storie] abbiano
       your brother   to whom myself I ask   which stories they have
       raccontato t$_2$ t$_1$]] era molto preoccupato
       told   was very   worried

    b. *[$_{DP_1}$ Chi] ti   domandi [$_{CP_3}$ [$_{DP_2}$ chi] t$_2$ ha incontrato t$_1$] ?
        who yourself you ask   who   has met

    gegeben]] ?
    given

    b. *Radios$_1$ glaube ich [$_{CP}$ der Maria$_2$ [$_C$ hat] [$_{TP}$ er t$_2$ t$_1$ gegeben]]
        radios   think I   the Mary   has   he   given

Before moving on, it may be noted that the fact that there is an asymmetry between movement types in (12), and no asymmetry between movement types in (13), poses severe problems for a Relativized Minimality type approach. A standard Relativized Minimality approach (as in Rizzi (1990; 2001)) distinguishes three kinds of interveners: Head, A, A-bar. This would uniformly rule out all sentences in (12) and (13) (given that the moved items are subject to Relativized Minimality in the first place). This seems hardly adequate in view of the (relative) wellformedness of (12-b). In contrast, a more fine-grained Relativized Minimality approach that distinguishes between different kinds of A-bar interveners (topic vs. *wh* in the case at hand) will invariably be too liberal. Such an approach is developed in Rizzi (2004), where a distinction is made between *argumental* (person, number, gender, case), *quantificational* (*wh*, neg, measure, focus), *modifier* and *topic* features. On this basis, we would wrongly predict both (12-b) and (13-a) to be well formed because the two A-bar-moved items bear different features (quantificational/*wh* vs. topic) (also see Unger (2010, 22)). Thus, if Relativized Minimality is to account for the three-out-of-four pattern in (12) and (13), further assumptions will be required.[9] This can be seen as additional evidence that a new approach to the phenomenon is called for, quite independently of the general conceptual considerations concerning properties of a good syntactic constraint (given the general assumptions about the derivational organization of grammar adopted in this monograph) that are summarized in (11) in chapter 2.
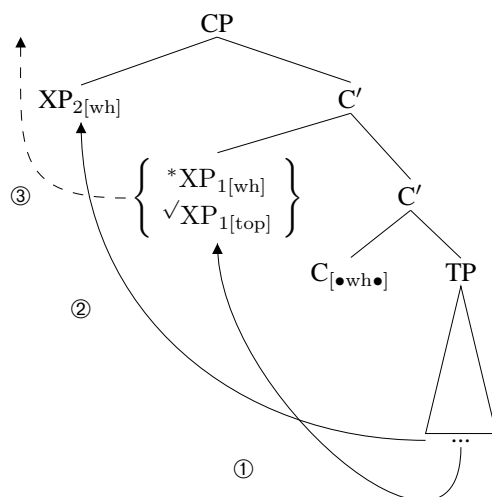
5.   Analysis: Operator Islands as Maraudage

Given the *Intermediate Step Corollary* in (9), the order of rule application with extractions from a *wh*-island and from a topic island must look as in (14-ab), respectively, with the intermediate movement step of the item that is to be long distance-moved (marked by ①) taking place prior to the one that would create the operator island
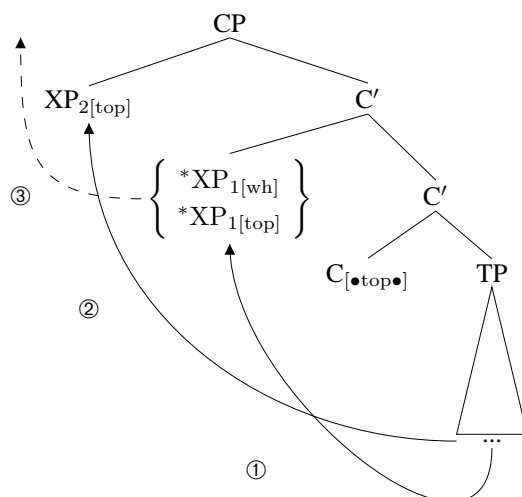
---

[9] See, for instance, Starke (2001), who builds an inherent asymmetry into his version of the Relativized Minimality condition, such that α may cross β if α has more (relevant) features than β, whereas α may not cross β when α has fewer (relevant) features than β. It is not inconceivable that the pattern emerging from (12) and (13) could be accounted for along these lines (given appropriate assumptions about the feature specifications of *wh*-phrases and topics); but the fact remains that a special version of Relativized Minimality would be needed for this, whereas the present approach dispenses in toto with a separate locality constraint blocking intervention. Furthermore, as will become clear momentarily, the present proposal tends to predict the opposite of what Starke's proposal predcits: Intervention effects can be voided if the moved item realizes a subset of the feature set associated with the (putative) intervener.

(marked by ②).

(14)   a. *Wh-island*

$$
\begin{array}{l}
\text{CP} \\
\quad \text{XP}_{2[\text{wh}]} \quad\quad \text{C}' \\
\quad\quad \left\{ \begin{array}{l} {}^*\text{XP}_{1[\text{wh}]} \\ \checkmark\text{XP}_{1[\text{top}]} \end{array} \right\} \quad \text{C}' \\
\quad\quad\quad\quad \text{C}_{[\bullet\text{wh}\bullet]} \quad \text{TP} \\
\end{array}
$$

① ② ③

b. *Topic island*

$$
\begin{array}{l}
\text{CP} \\
\quad \text{XP}_{2[\text{top}]} \quad\quad \text{C}' \\
\quad\quad \left\{ \begin{array}{l} {}^*\text{XP}_{1[\text{wh}]} \\ {}^*\text{XP}_{1[\text{top}]} \end{array} \right\} \quad \text{C}' \\
\quad\quad\quad\quad \text{C}_{[\bullet\text{top}\bullet]} \quad \text{TP} \\
\end{array}
$$

① ② ③

The hypothesis that I would now like to pursue is this: Given that the item that undergoes the intermediate movement step reaches the domain of a C head before the item that is supposed to ultimately check the [●F●] of C in this position, it may maraud C's stack of structure-building features, making regular specifier placement impossible. Thus, *wh*-island and topic island effects are due not to *wh*-islands or topic islands but, somewhat paradoxically, to the fact that the *wh*-island/topic island cannot be generated in the first place.

To flesh out this hypothesis and make it possible to account for variation in maraudage, a more fine-grained system of A-bar-related features is needed (see Rizzi (1997; 2004) for the basic idea; and also Lahne (2007) for a version that does not correlate such A-bar-related features with separate functional heads). For concreteness, suppose that topicalization (in German) is bare operator movement to SpecC, and *wh*-movement is movement of a certain kind of operator (viz., a *wh*-operator) to SpecC.[10] The (relevant) feature specifications of C heads that trigger *wh*-movement, of C heads that trigger topicalization, and of *wh*-phrases and topics are given in (15).[11]

(15)   a. $C_{[wh]}$ = $C_{[\bullet op, wh \bullet]}$
        b. $C_{[top]}$ = $C_{[\bullet op \bullet]}$
        c.   *wh*-phrase = $XP_{[op, wh]}$
        d.   topic   = $XP_{[op]}$

Finally, let me make explicit again the assumption that certain kinds of goal features may trigger maraudage as soon as they enter a syntactic context in which they can legitimately be checked. The principle of feature maraudage in Georgi et al. (2009) is repeated here in a modified form (see (11)): The maraudage-inducing features that are relevant in the present context are [op] and [wh] rather than [person] and [animate].

(16)   *Maraudage:*
        Certain goal features of Agree or Merge operations (among them [op] and [wh]) are checked if the structural conditions for checking are met.

Since [op] and [wh] are features that are matched by structure-building features (rather than by probe features) in German (assuming that there is no specific Agree requirement on *wh*-in situ or topic-in situ in this language), this means that maraudage by a *wh*-phrase or a topic can only take place after the item has undergone movement to SpecC for an independent reason. However, when an item bearing [op], or [op] and [wh], undergoes intermediate movement to a SpecC position, driven by edge feature

---

[10] This classification of topicalization as an underspecified movement operation chimes in well with its often-noted multi-functionality in German (see, e.g., Grewendorf (1989); also Culicover (1991) on English): An item that is subject to the syntactic operation of topicalization may variously be interpreted as a (semantic) topic, as a (regular) focus, as a contrastive focus, or it may remain without any information-structural effect (i.e., it may be subject to semantic reconstruction). (Also cf. Fanselow (2011) for an even stronger position which would nevertheless support the system in (15) in essentially the same way: Fanselow argues that the formal, syntactic operation of topicalization in German actually *never* has an information-structural effect (or source).

[11] All these specifications are particular to German. It is very well conceivable (in fact, almost certainly the case) that other languages have subtly different specifications for topicalization and *wh*-movement, which will then have interesting consequences for the strictness of the operator islands created by (and for) topics and *wh*-phrases.

insertion, it tries to get as many of its operator features ([op], [wh], ...) checked as possible. Structure-building operator features thus attach to the newly inserted [•X•], are checked with the moved item, and undergo deletion. Therefore, no feature may be left to trigger subsequent, regular operator movement. This accounts for operator island effects. To see how this approach works in detail, let me go through the four relevant derivations one by one, beginning with *wh*-movement from a *wh*-island.

5.1 Wh-Movement from a Wh-Island

Recall that *wh*-movement (of arguments as well as adjuncts) from a *wh*-island is ungrammatical in German; (12-a) is repeated here as (17).

(17) *Welches Radio$_1$ weißt du nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_1$ t$_2$ repariert]] ?
     which radio know you not   how   one   fixes

Wh-phrases in German have two A-bar-related features: [op] and [wh]. Therefore, they accomplish full maraudage in a SpecC$_{[wh]}$ position that they use as an escape hatch in a CP phase, thereby blocking subsequent regular *wh*-movement. (18) is a step-by-step illustration of what happens to the interrogative C head in (17) in the course of the derivation once it has been merged with TP (driven by a top-most subcategorization feature [•T•] on C).

(18)   *Wh-movement from a wh-island: complete maraudage*
     (i)   C: [•op, wh•]   ($\rightarrow$ edge feature insertion)
     (ii)   C: [•X•] $\succ$ [•op, wh•] ($\rightarrow$ movement of *wh*-phrase)
     (iii) C: [•X•] $\succ$ [•op, wh•] ($\rightarrow$ edge feature discharge)
     (iv) C: [•op, wh•]   ($\rightarrow$ further discharge: maraudage)
     (v)   C: $\emptyset$ (no structure-building features left)

First (cf. the step from (i) to (ii)), the C head that bears the complex *wh*-feature (consisting of the primitive features [op] and [wh]) is affected by edge feature insertion for XP$_{1[op,wh]}$. This is permitted by the Edge Feature Condition, and then required at exactly this step of the derivation because it is the only step where the item that needs to move is already part of the structure (TP has been merged), and C is still active because it has an undischarged structure-building feature left (viz., the complex *wh*-feature consisting of two primitive features [op] and [wh]). Next, movement of XP$_{1[op,wh]}$ to SpecC takes place (step (iii)). This movement is legitimate (vis-à-vis Last Resort) because the edge feature on C can now be discharged and deleted (step (iv)).[12] However, since the syntactic context for maraudage is met (the *wh*-phrase is in

---

[12] This presupposes that the formal operation of movement should be kept apart from the feature discharge operation that licenses it. Alternatively, the two steps could be seen as a single composite operation (Merge – internal or external – plus discharge). So far, I have been uncommited as concerns this issue, but the

the edge domain of a C head that is equipped with exactly the A-bar-related features [●op, wh●] that can be used by the features [op, wh] on the *wh*-phrase which trigger checking according to (16)), the *wh*-phrase in SpecC (i.e., $XP_{1[op,wh]}$) will now also carry out complete maraudage (step (v)), effecting discharge (thus, deletion) of all structure-building features on C. This means that the A-bar related features that C had reserved for the 'proper' *wh*-phrase to move to its specifier (whichever of the two that is) are now gone; there is no feature left for attracting $XP_{2[op,wh]}$. What is more, the C head is now not active anymore: The last structure-building feature has gone, and the phase head also cannot have a probe feature left at this point that would render it active (because this would imply a violation of either the c-command condition on Agree, or the Strict Cycle Condition; recall subsection 3.2.3 of chapter 4). Therefore, an edge feature cannot be inserted for the remaining *wh*-phrase which might move it to SpecC.[13]

As a consequence, there is no way to get the remaining $XP_{2[op,wh]}$ (the item we would expect to create the *wh*-island) to the edge domain of the embedded CP (from the edge domain of the TP phase to which it has undergone an intermediate movement step earlier in the derivation, in accordance with the Edge Feature Condition). Depending on assumptions about criterial freezing (see Rizzi (2006; 2007), and the Criterial Freezing constraint in (48) in chapter 2), it may or may not be possible now for the *wh*-phrase in the embedded SpecC position to move on, into the matrix clause (to satisfy the demands of another $C_{[wh]}$). However, such movement would still lead to a violation of a general visibility requirement for the embedded *wh*-clause; cf. (19).[14] In any case, the prediction is that, somewhat paradoxically, *wh*-island effects are characterized by the property that the *wh*-island cannot be erected.

---

wording in the preceding chapters would arguably have been closer to the second possibility. Nothing depends on this issue for the purposes of the present monograph; but there might eventually be some non-trivial differences. See Heck & Müller (2007).

[13] As observed by Stefan Keine (p.c.), this reasoning is compatible with the hypothesis that [wh]-features are accompanied by scope indices (see section 5.1 of chapter 3) if it is assumed that such scope indices play a role for potential availability but not for feature checking. Thus, maraudage of [●op, wh●]$_i$ by [wh]$_j$ is possible.

[14] As noted, $XP_{2[op,wh]}$ is really not in situ but in SpecT at this point. This does not affect the general reasoning, though, and I stick to genuine in-situ variants here and in what follows. There are basically two reasons for this. First, the crucial parts of the derivation are arguably more perspicuous if irrelevant movements are omitted. And second, while the approach to operator islands developed in the present chapter relies on the Intermediate Step Corollary and feature maraudage, it is not *not* inherently tied to the assumption that all phrases are phases – it works just as well if only vP and CP are phases, or even CP alone.

(19) *Welches Radio$_1$ fragst du   dich   [$_{CP}$ t$'_1$ (dass) [$_{TP}$ man t$_1$ wie$_2$ repariert]] ?
       which   radio   ask   you yourself   that   one   how fixes

The maraudage operation accounting for the strict barrier status of *wh*-clauses for *wh*-movement under the Intermediate Step Corollary is illustrated in (20) (where ① and ② signal movement and massive feature discharge, respectively, and movement step ③ is then blocked).

(20) *The decisive stage of the derivation – wh-extraction from wh-island*:
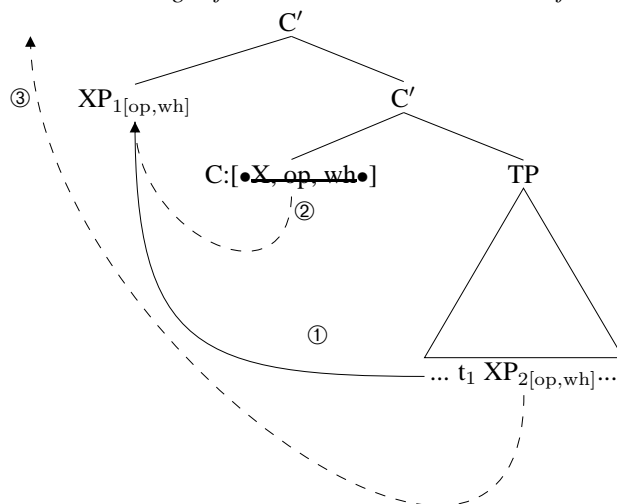
5.2   Topicalization from a Wh-Island

As noted, topicalization from a wh-island yields much better results; examples of the type in (12-b) occur frequently in unmarked discourse, and can be considered as minimally deviant at worst in various varieties of German; (12-b) is repeated here as (21).

(21) ?Radios$_1$ weiß ich nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_1$ t$_2$ repariert]]
       radios   know I   not   how   one   fixes

Given (15), topics have fewer A-bar-related features than *wh*-phrases. Therefore, they do not accomplish full maraudage in a SpecC$_{[wh]}$ position that they use as an escape hatch in a CP phase. Consequently, subsequent regular *wh*-movement is not blocked. (22) shows how the phase head C acquires and discharges its A-bar-related features in the course of the derivation generating (21).

(22)   *Topicalization from a wh-island: minor maraudage*

    (i)   C: [●op, wh●]   (→ edge feature insertion)

    (ii)   C: [●X●]   ≻ [●op, wh●] (→ movement of topic)

    (iii) C: [●X●]   ≻ [●op, wh●] (→ edge feature discharge)

    (iv) C: [●op, wh●]   (→ minimal maraudage)

    (v)   C: [●wh●]   (→ movement of *wh*-phrase)

    (vi) C: [●wh●]   (→ discharge)

    (vii) C: Ø   (two items in edge domain)

The important difference to the unsuccessful derivation in (18) is that the topic that needs to move to the edge domain first (given the Intermediate Step Corollary) discharges only the [●op●] feature of the interrogative C head, and leaves the [●wh●] feature available for further structure-building (step v). This [●wh●] feature then suffices to attract the *wh*-phrase to its target position, thereby erecting a *wh*-'island' that comes too late to block extraction.[15]

    The relevant stage of the derivation is shown in (23); operations ① and ② leave [●wh●] accessible, and therefore do not preclude operation ③.

(23)   *The decisive stage of the derivation – topicalization from wh-island*:

---
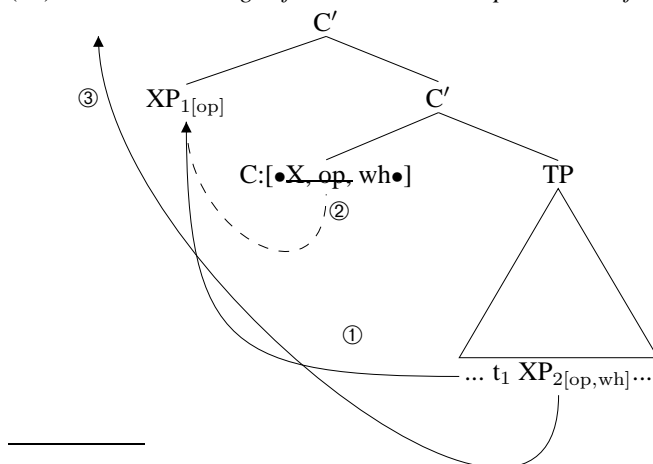
[15] One might speculate that a mild deviance of examples like (21) might be correlated with the fact that not all of the *wh*-phrase's goal features are eventually checked in this context, with [op] remaining unchecked. Note, however, that an unchecked goal feature does not pose a problem as such under present assumptions: In-situ *wh*-phrases have all their A-bar-related goal features left unchecked; and the condition in (16) only requires checking of [op] if the structural conditions are met (which they are not if there is no matching structure-building feature left on the phase head). Still, one might claim that whereas complete checking of A-bar-related features on operator XPs is unproblematic, and complete absence of checking is unproblematic, too (as long as (16) and the Feature Condition are respected), *incomplete* checking may give rise to minimal deviance. Still, the issue might emerge as more complex once other kinds of constructions are taken into consideration; see Georgi (2010) on A-related features.

5.3   Wh-Movement from a Topic Island

Consider next the case of *wh*-movement from a topic island, which always leads to ungrammaticality. (13-a) is repeated here as (24).

(24) *Welches Radio$_1$ glaubst du [$_{CP}$ der Maria$_2$ [$_C$ hat] [$_{TP}$ er t$_2$ t$_1$ gegeben]] ?
      which   book   think   you   the Mary   has   he   given

As we have seen above, *wh*-phrases are perfect marauders in a SpecC$_{[wh]}$ posi-
tions; consequently, it does not come as a surprise that they are just as successful in
SpecC$_{[top]}$ positions, where C is characterized by a proper subset of structure-building
features. (25) illustrates how C is affected by complete maraudage by the *wh*-phrase
in the course of the derivation, which leaves no features for attracting a topic.

(25)   *Wh-movement from a topic island: complete maraudage*
       (i)   C: [•op•]   ($\rightarrow$ edge feature insertion)
       (ii)   C: [•X•] $\succ$ [•op•] ($\rightarrow$ movement of *wh*-phrase)
       (iii) C: [•X•] $\succ$ [•op•] ($\rightarrow$ edge feature discharge)
       (iv)   C: [•op•]   ($\rightarrow$ further discharge: maraudage)
       (v)   C: $\emptyset$ (no structure-building features left)

The derivation proceeds in essentially the same way as in the case of *wh*-movement
from a *wh*-island. First, an edge feature is inserted on C for the *wh*-phrase that needs
to undergo long-distance movement (the *wh*-phrase is in SpecT at this point); see (i).
Next, the *wh*-phrase undergoes movement to SpecC and discharges the edge feature
(see (ii), (iii)). Given (16), maraudage of C's structure-building feature [op] by the
wh-phrase applies next (see (iv)). As a consequence, C does not have any structure-
building feature left for carrying out topicalization (see (v)). Therefore, it is impossible
to move the remaining XP$_{2[op]}$ (the item that supposedly creates the topic island) to
the edge domain of the embedded CP (C is now inert, which precludes further edge
feature insertion). Thus, it follows that a structure like (24) can never be generated;
the topic island effect is derived (for *wh*-phrases). Still, the question arises of what the
fate of a derivation is that incorporates the operations in (25).

As with the derivation based on (18), there are in principle two possibilities: The
wh-phrase may not be able to undergo further movement now, or it may be able to
do this. According to the first possibility, it could be assumed that the topicalized
wh-phrase is precluded from moving on, as an instance of criterial freezing (see Rizzi
(2006; 2007)). The other possibility is to either dispense with the notion of criterial
freezing altogether, or weaken it in such a way that a topicalized item may move on to
another position in which it discharges some inherent structure-building feature. On
this view, the continuation of a derivation that involves (25) may give rise to sentences

like (26).

(26)   Welches Radio$_1$ glaubst du   [$_{CP}$ t$'_1$ [$_C$ hat] [$_{TP}$ er der Maria t$_1$ gegeben]] ?
       which   book   think you   has   he the Mary   given

Closer inspection reveals that the standard analysis of extraction from verb-second clauses in German requires a weakening (or abandonment) of criterial freezing anyway: A declarative verb-second C head triggers topicalization, and the topicalized item may then move on into the matrix clause, exactly as in (26).[16] Thus, if we just implement standard assumptions about extraction from verb-second clauses in the present approach to movement, extraction from verb-second clauses looks almost exactly as sketched in (25). The only differences would be (i) that in the standard approach to verb-second clauses, an edge feature does not have to be inserted for the *wh*-phrase because the *wh*-phrase, by assumption, *is* also a topic; and (ii) that there does not have to be some other item around that would act as a topic, so that discharge of the [op] feature by the *wh*-phrase would not technically count as an instance of maraudage. On this view, we end up with the following situation: Structures like the one in (26) may alternatively be derived via maraudage (as in (25)), or without maraudage (if the *wh*-phrase counts as the only topic).[17] At present, I take it to be an open question whether this kind of potential derivational ambiguity is harmless, or should be filtered out in some way.[18] To sum up, given the Intermediate Step Corollary, maraudage of C's [op] feature accounts for the impossibility of structures like the one in (24) that involve topicalization in the embedded clause (a topic island cannot be generated in the first place); whether the derivation that can be generated then necessarily crashes, or may be continued to a legitimate case of *wh*-movement from a verb-second clause is an orthogonal issue that can be resolved in either way.

---

[16] The only alternative would be to assume that extraction from verb-second clauses is not actually an option in German (possibly just because of a strict notion of criterial freezing), and that strings like the one in (26) involve a parenthetical construction; see Reis (1996).

[17] At least, this is so if we abstract away from Edge Feature Condition-driven (ultimately: Phase Balance-driven) displacement of *der Maria* to a SpecT phase edge in (26). Note that this particular example would, as a matter of fact, lead to ungrammaticality (since a non-pronominal object DP in SpecT would precede an unstressed subject pronoun at the left edge of vP); but such EFC-driven fronting to SpecT would be entirely unproblematic, and often indistinguishable from free scrambling, in many other cases. See footnote 14 above.

[18] Note that the same situation arises systematically in optimality theory. Prince & Smolensky (2004) suggest that such an ambiguity should be systematically resolved by an additional operation known under the label of *input optimization* (or *lexicon optimization*, if the focus is just on phonology): Basically, the derivation that employs fewer vacuous features is to be preferred. A similar strategy would be possible in the present context.

As before, it may be helpful to illustrate the crucial part of the derivation by an abstract tree that exhibits the relevant operations; see (27): The intermediate step of *wh*-movement (see ①) triggers full maraudage (see ②), which blocks subsequent regular topicalization (see ③).

(27) *The decisive stage of the derivation – wh-extraction from topic island*:
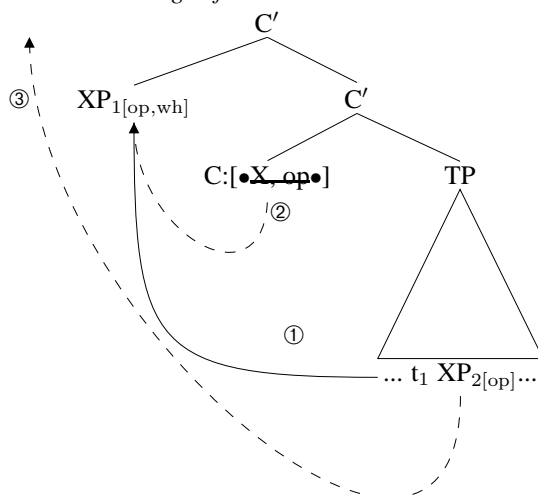
5.4 Topicalization from a Topic Island

Unlike *wh*-islands, topic islands are also strict for topicalization in German; recall (13-b), which is repeated here as (28).

(28) *Radios$_1$ glaube ich [$_{CP}$ der Maria$_2$ [$_C$ hat] [$_{TP}$ er t$_2$ t$_1$ gegeben]]
    radios   think I   the Mary   has   he   given

The analysis is straightforward. Topics have only one operator feature: [op]. However, this feature suffices to block subsequent topicalization via maraudage if a topic undergoes movement to an intermediate SpecC position on its way to the left periphery of the matrix clause. (29) shows what happens to the embedded C head that would normally trigger (embedded) topicalization.

(29) *Topicalization from a topic island: complete maraudage*
    (i)   C: [●op●]   (→ edge feature insertion)
    (ii) C: [●X●] ≻ [●op●] (→ movement of topic)
    (iii) C: [●X●] ≻ [●op●] (→ edge feature discharge)
    (iv) C: [●op●]   (→ further discharge: maraudage)
    (v)   C: ∅   (no structure-building features left)

(29) looks exactly like (25); for maraudage of C's [●op●] feature, it does not matter whether the marauding item is equipped with the features [op, wh], or just with the

feature [op].

As in the case of *wh*-movement from a clause with a C head that triggers topical-
ization, the consequence is that a topic island cannot be created in the first place. The
same issues arise with respect to criterial freezing and input optimization as before,
i.e., concerning the question of whether the derivation can then legitimately be contin-
ued to produce (30), or whether (30) can only be generated without maraudage (based
on a slightly different numeration).

(30)   Radios$_1$ glaube ich [$_{CP}$ t$'_1$ [$_C$ hat] [$_{TP}$ er der Maria t$_1$ gegeben]]
        radios   think I   has   he the Mary   given

The relevant part of the derivation looks identical to the one illustrated for *wh*-
extraction from a topic island in (27); see (31).

(31)   *The decisive stage of the derivation – topicalization from topic island*:
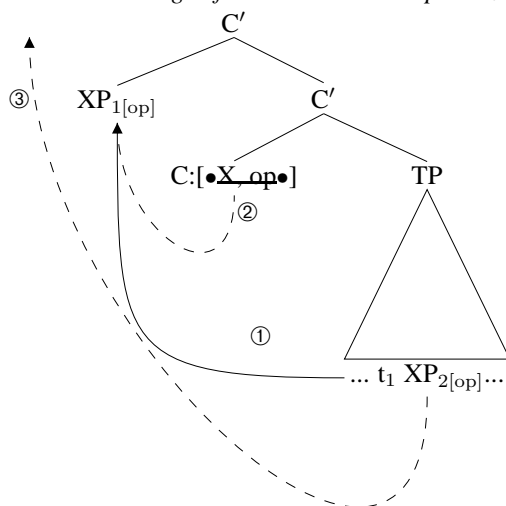
6.   Conclusion, Consequences, and Extensions

6.1   *Conclusion*

To sum up, the present approach to *wh*-islands and topic islands works without in-
voking the idea that *wh*-elements or topics create islands. More generally, there is
no minimality/intervention constraint on movement (like the (G)MLC or Relativized
Minimality). Rather, an item undergoing long-distance movement targets the same
domain and marauds the inventory of movement-inducing features of C before the
items that are supposed to show up in a specifier position of this C node permanently
have had a chance to get there. Thus, as with the analysis in chapters 3 and 4, the
crucial factor is *timing*: Which item arrives first in the C domain? I take it to be an
interesting result that the answer needed to derive *wh*-island and topic island effects is

one that follows automatically, given the approach to CED effects in chapter 4. The reason is that this approach has as an immediate consequence what I have called the *Intermediate Step Corollary*: Intermediate movement steps to specifiers of X (as required by the PIC) must take place before a final specifier is merged in XP. Needless to say, the analysis of operator islands just sketched has a number of further non-trivial consequences; it also suggests extensions in various domains. In the remaining two subsections, I will briefly address these two issues.

6.2   Consequences

6.2.1   Multiple Movement

First, multiple movement of the same type to a single domain has been proposed for various kinds of phenomena, e.g., multiple scrambling to the vP domain in German (and many other languages), and multiple *wh*-movement to the CP domain in (some) Slavic languages (cf. Rudin (1988)); see (32-ab), from German and Bulgarian, respectively.

(32)   a.   dass [$_{vP}$ dem Mann$_1$ [$_{v'}$ das Buch$_2$ [$_{v'}$ keiner   [$_{VP}$ t$_1$ t$_2$ gab] v]]]
              that   the man$_{dat}$ the book$_{acc}$ no-one$_{nom}$ gave

       b.   [$_{CP}$ Koj$_1$ [$_{C'}$ kogo$_2$ [$_{C'}$ kakvo$_3$ [$_{C'}$ C [$_{TP}$ t$_1$ e pital t$_2$ t$_3$]]]]] ?
              who   whom   what   asked

Clearly, irrespective of the question whether, e.g., [Σ] features are subject to maraudage in the same way that [wh] features are, such structures should be permitted throughout in approaches based on the option of multiple specifiers; thus, (32-ab) must not be ruled out by repeated maraudage of the phase head's (i.e., v's or C's) movement-inducing features by the first-moved item. This follows without further ado if it is assumed that checked features (like [Σ] and [wh] in the case at hand) do not have to be checked again with genuine movement-inducing features.[19]

6.2.2   'Whether'-Clauses

A second consequence concerns extraction from *whether* clauses. Chomsky (1986a, 50) claims that "*whether* yields a much weaker *wh*-island effect than moved *wh*-phrases." As a matter of fact, the prediction of the present analysis is that there will be no effect at all if *whether* is the lexical realization of an interrogative C. However, this prediction is presumably not correct. I will assume that at least for adjuncts, and arguably also for arguments, a *wh*-island effect can be detected with extraction from *whether* clauses, as in (33).

---

[19] If one were to assume that such features in fact cannot be checked again once they have been checked by a structure-building feature (except for pure edge features), this would amount to postulating criterial freezing; see above.

(33)   a.   *How$_1$ do you wonder [$_{CP}$ whether Mary fixed the car t$_1$] ?

       b.   ?*What$_1$ do you wonder [$_{CP}$ whether John likes t$_1$] ?

The conclusion I would like to draw from this is that *whether* is not the lexical realization of an interrogative C; more specifically, suppose that *whether* is never base-generated in the position in which it shows up, but undergoes movement to a C$_{[\bullet wh \bullet]}$ head just like other *wh*-items. Principles of semantic interpretation will then ensure that there can be no multiple *wh*-questions involving both *whether* and and a regular *wh*-phrase.[20] There is potential additional evidence for a movement approach to *whether* clauses. In a language like Dutch, *of* ('whether') can co-occur with an uncontroversial complementizer *dat* ('that') following it (see, e.g., Zwart (1993, 265)). Under present assumptions, this might suggest an approach in terms of *whether*-movement from within TP to SpecC. Of course, this analysis raises further questions. For reasons of space and coherence, I will confine myself to pointing out two of them: First, where exactly is *whether* externally merged? And second, why can *whether* not undergo long-distance movement? However, none of these questions strikes me as particularly troublesome, and I take it that answers can be given that are relatively straightforward.

6.2.3   *Underspecification and Overspecification*

A third issue concerns underspecification and overspecification of *wh*-phrases and topics with respect to operator features, as it is presupposed by the approach developed here. Two clear predictions can be derived from this. First, given that topics are specified as [op], and *wh*-phrases as [op, wh], it follows that a topic can never discharge all the A-bar-related features of an interrogative C$_{[\bullet op,wh \bullet]}$. Therefore, we expect that a moved topic can never eventually satisfy the demands of an interrogative C. In other words: topics should not be able to act as *wh*-phrases. This prediction appears to be borne out (in German, but also cross-linguistically).

   Second, the case might be different with *wh*-phrases. A *wh*-phrase bearing the features [op, wh] in the specifier of a topic C$_{[\bullet op \bullet]}$ can discharge all of C's features. This means that it might be possible for a *wh*-phrase to ultimately show up in the specifier domain of a non-interrogative C. If there is no other interrogative C head around in the sentence, such a configuration will most likely lead to semantic uninterpretability. However, suppose that the *wh*-phrase has undergone regular, feature-driven *wh*-movement to a specifier of an interrogative C head first. Then, depending (again) on assumptions about criterial freezing, it might or might not be an option for the *wh*-phrase to move on, and discharge a non-interrogative C head's [$\bullet$op$\bullet$] feature in

---

[20] Note in passing that movement of *whether* is in fact proposed in Chomsky (1986a), if only for the covert component of LF.

a higher clause. Marked (but still halfway acceptable) examples of *wh*-topicalization in imperative clauses in German (see Reis & Rosengren (1992)) might suggest that this option does exist; compare (34-a) (with regular *wh*-movement to a SpecC$_{wh}$ position in an indirect question clause) with (34-b) (with further fronting to the specifier position of the matrix – imperative – C head).

(34) a. $[_{CP}$ – Sag mal $[_{CP}$ wen$_1$ (dass) du t$_1$ getroffen hast $]]$!
                tell me   whom that you   met   have

b. ??$[_{CP}$ Wen$_1$ $[_{C_{imp}}$ sag $]$ mal $[_{CP}$ t$'_1$ dass du t$_1$ getroffen hast $]]$!
            whom   tell me   that you   met   have
         (both: 'Tell me who you met!')

Again, further questions will eventually have to be answered if the analysis is to prove viable. For instance, something needs to be said about the illformedness of examples like (35-b) (based on (35-a)), which differs from (34-b) only in that topicalization of the *wh*-phrase ends up in a declarative (rather than imperative) root clause. ((35-b) does not seem to become acceptable even if the topicalized is interpreted as heavily focussed, even though some improvement can arguably be detected in this case.)

(35) a. $[_{CP}$ – Maria fragt $[_{CP}$ wen$_1$ (dass) Fritz t$_1$ getroffen hat $]]$
                Maria asks   whom that Fritz   met   has

b. *$[_{CP}$ Wen$_1$ $[_{C_{decl}}$ fragt $]$ Maria $[_{CP}$ t$'_1$ dass Fritz t$_1$ getroffen hat $]]$
            whom   asks Maria   that Fritz   met   has
         (both: 'Maria asks whom Fritz met.')

6.3   Extensions

There are two general correlations that can be derived from the analysis of operator islands via maraudage and the Intermediate Step Corollary. They are given in (36).

(36) a.   The more A-bar-related (goal) features a moved item is equipped with, the more likely it is that it cannot cross an operator island (i.e., the "operator island" cannot be created in the first place).

b.   The fewer A-bar-related (structure-building) features a C head is equipped with, the more likely it is that it creates an operator island that cannot be crossed.

Combining (36-a) and (36-b), the prediction is that an operator island effect can best be circumvented if the item that is to undergo long-distance movement has few operator features, and the C head that triggers embedded operator movement has many operator features – this is why movement of a topic (with few operator features) across a C$_{wh}$ head (with many operator features) is the only combination leading to wellformedness in German. Both (36-a) and (36-b) suggest interesting extensions.

Turning to (36-a) first, it offers a new outlook on argument/adjunct (-type) asymmetries with extraction from islands (*weak islands*; see, e.g., Koopman & Sportiche (1986) on Vata). The general logic underlying an account of an argument/adjunct asymmetry with extraction from *wh*-islands might look as follows. Suppose that moved adjuncts are inherently characterized by more A-bar-related features than moved arguments. This implies that, as operators, adjuncts maraud the feature structure of a C head whose domain they enter in the course of successive-cyclic (edge feature-driven) movement more drastically than arguments. Consequently, an intermediately-moved argument may leave further structure-building features on a C head, whereas an intermediately-moved adjunct "eats up" all structure-building features, and thus makes the regular movement operation to be carried out for this head impossible. I take this general logic to be by and large on the right track. However, as is well known, drawing the line between arguments and adjuncts with respect to extractability from weak islands (like *wh*-islands) is an oversimplification. For many languages, there is a split between referential and non-referential adjuncts (*when, where* vs. *how, why*; see Aoun (1986, 125); also see Harbert & Pet (1988) on *alon* 'where' vs. *halika* 'how' in Arawak), such that referential adjuncts can leave weak islands whereas non-referential adjuncts cannot. Similarly, in some languages, there is a split between referential (D-linked, etc.) arguments and non-referential arguments, such that only the former can escape weak islands.

A (still somewhat simplified) description of the situation in German in the case of *wh*-islands may look as follows. First, topic islands are strict throughout. Second, *wh*-islands are transparent if the moved item is a topic which is not a non-referential adjunct or a non-referential argument (see Cinque (1990), Rizzi (1990)). Thus, (37-a) (with topicalization of a non-referential adjunct from a *wh*-clause) is ungrammatical, and (37-b) (with topicalization of a non-referential argument) also seems to be fully ungrammatical. In contrast, (37-c) (with topicalization of a referential adjunct from a *wh*-clause) is halfway acceptable (see footnote 15 on the possible source of degradedness in (37-c)).

(37) a. *Deshalb$_1$ weiß ich nicht mehr   [$_{CP}$ wer$_2$ C t$_2$ t$_1$ gekommen ist]]
           therefore know I   not   anymore   who   came

      b. *90   Kilo$_1$ weiß ich nicht [$_{CP}$ wer$_2$ C [$_{TP}$ t$_2$ t$_1$ wiegt]]
           ninety kilos know I   not   who   weighs

      c. ??In Hannover$_1$ weiß ich nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_2$ t$_1$ das sagt]]
           in Hannover know I   not   how   one   that says

This state of affairs can be accounted for if it is assumed that all *wh*-phrases and non-referential topics bear the feature [wh] in addition to [op] in German, so there is full maraudage of the embedded C's structure-building features in (37-ab) (but not in (37-c)). To this one might object that non-referential topics do not morphologically

look like *wh*-phrases, and behave differently syntactically. However, such an objection would be based on a misunderstanding of the nature of feature decomposition. The primitive feature [wh] is not the same as the standard *wh*-feature; rather, it defines a part of the standard *wh*-feature; *wh*, on this view, is composed of primitive features like [op] and [wh], and it is just the combination of primitive features that makes a *wh*-phrase a *wh*-phrase, not any single part of it. (Accordingly, one might just as well replace the feature [wh] with a feature like [nonref], which would capture exactly the same natural class – *wh*-phrases are inherently non-referential.) Given that, say, topics can behave "*wh*-like" in at least certain respects (see Chomsky (1977)), this is captured by postulating features that are common to both but not other categories – i.e., by postulating a natural class via shared features (Bierwisch (1967)). Clearly, though, there are differences between non-referential topics and *wh*-phrases, and by assuming that they are both characterized as [op, wh], these differences are not yet captured. This can be done by assuming one or more additional features that distinguish between the two types of categories; again, this is a standard procedure in feature-based approaches that rely on decomposition and underspecification.[21]

Consider next the situation in English. A simple characterization of the state of affairs here might maintain that topic islands are strict throughout (but cf. Culicover (1991) for a systematic class of exceptions), whereas *wh*-islands are transparent if the moved item is an argument. (This is a simplification: Notions like D-linking, referentiality, and case may also enter the picture.) This can be analyzed by assuming that interrogative C has two A-bar-related structure-building features (e.g., [•op, wh•]), whereas topic C has only one A-bar-related structure-building feature (e.g., [•op•]). All adjuncts (topics and *wh*-phrases) have the complete set of A-bar-related features (e.g., [op, wh]); all arguments (but see the qualification just made) have only one (e.g., either [op] or [wh]).

More generally, the present approach may shed new light on the classical question of why certain classes of items may escape certain types of weak islands in some language whereas other classes of items cannot do so, and on how the variation in this domain can be accounted for. The next step in this research programme would be to look at other instances of weak islands, and investigate whether a maraudage analysis would also be tenable, or, indeed, directly supported by the empirical evidence; but

---

[21] Note that, on this view, issues of over- and underspecification become relevant again: For instance, one has to ensure that non-referential topics, even though they can carry out complete maraudage of an interrogative C head's set of structure-building features, cannot act as *wh*-phrases themselves, saturating an interrogative C head's demands.

this is beyond the scope of the present chapter.[22]

Turning next to (36-b), I would like to contend that this correlation offers a new outlook on the variable effects incurred by different types of intervening C heads. Heads with few movement-inducing features generate stricter islands (because they are easier to maraud fully). I have argued that this is the case with topic vs. *wh*-islands in German: An interrogative C is equipped with more features than a declarative C that triggers topicalization. From this perspective, one may profitably look at other asymmetries between movement types; see, for instance, Bayer & Salzmann (2009) on the different behaviour of relativization in comparison to other movement types, and Müller & Sternefeld (1993) on asymmetries between topicalization, *wh*-movement, and scrambling. A further extension that suggests itself concerns finite/non-finite asymmetries in extraction (see Frampton (1990)).

To end this chapter, let me point out that the approach to operator islands advanced here leaves sufficient room for parametrization, such that variation between languages (even micro-variation, possibly within a single language) can be captured. On the one hand, particular types of moved items (arguments, adjuncts, referential items, *wh*-phrases, topics, etc.) may not be characterized by exactly the same kinds of A-bar-related features in different languages. And on the other hand, particular movement operations (topicalization, *wh*-movement, relativization, scrambling, etc.) may also not be cross-linguistically homogeneous as far as the A-bar-related features that are involved in them are concerned.

The final two chapters of this monograph address two cases where a CED effect might be expected but fails to occur, viz., certain kinds of extractions from verb-second clauses in German, and island repair by ellipsis. In both cases, I argue that there is reason to assume that a phase head that might initially be thought to be fatally inactive at the relevant stage of the derivation is actually active because it triggers subsequent operations (verb-second and deletion, respectively).

---

[22] In this context, it is worth pointing out that Manzini (2009) develops an approach to the blocking of mesoclisis and enclisis by negation in varieties of Italian that does not employ the Head Movement Constraint or another intervention-based constraint, and that is similar to the one developed here. In her analysis, negation blocks verb movement not because it intervenes between the base position and the landing site, but because it consumes features that would be needed by the verb in the fronted position.

Chapter 6

Movement from Verb-Second Clauses

1. Introduction

Extraction from declarative complement clauses in German reveals a curious pattern: Whereas verb-final complement clauses headed by *dass* ('that') permit movement both into a higher verb-second and a higher verb-final clause, complement clauses in which verb-second has applied only permit movement into a higher verb-second clause again, but not movement into a higher verb-final clause.[1] This systematic pattern has been addressed in a number of approaches of various kinds in the last decades, with varying degrees of success, but it seems fair to conclude that a conceptually simple analysis on the basis of the minimalist program (see Chomsky (2001b; 2008)) is still outstanding. The goal of the present chapter is to show that such an analysis is readily available if certain insights of Staudacher's (1990) approach in terms of barriers are incorporated into the phase-based approach to Condition on Extraction Domain (CED) phenomena developed in chapter 4 above, and if verb-second clauses are derived by reprojection movement. In a nutshell, the restriction on movement from verb-second clauses into verb-final clauses will be shown to follow from the PIC because an edge feature that is needed to trigger movement from the verb-second complement to the next phase edge cannot be inserted on the matrix verb; in contrast, movement from verb-second clauses into verb-second clauses again can circumvent a PIC-based CED effect because here, the matrix verb independently has an operation-inducing feature left (that will subsequently trigger verb-second) which keeps it active at the relevant stage of the derivation.

I will proceed as follows. Secion 2 introduces the relevant data and gives a very concise, and certainly incomplete, overview over the existing literature on the phenomenon. Section 3 sketches Staudacher's (1990) approach. Section 4 summarizes

---

[1] A version of this chapter has appeared as Müller (2010a). The analysis owes an obvious debt to Peter Staudacher's work.

the main consequences of the phase-based analysis of CED effects developed in chapter 4 that will play a role in the analysis. Section 5 introduces an approach to verb-second movement that is based on the concept of reprojection; the approach combines features of the analyses in Fanselow (2009) and Georgi & Müller (2010). Section 6 then shows that the ban on movement from verb-second clauses into verb-final clauses can be derived given the assumptions laid out in sections 4 and 5, in a way that integrates some of Staudacher's original insights. Finally, section 7 addresses some further issues.

2. The Problem

2.1 Data

Two types of finite declarative clauses can be embedded under certain kinds of verbs (often bridge verbs, though the correlation is far from perfect) in German: (i) verb-final clauses headed by a complementizer *dass* ('that'); (ii) verb-second clauses with finite V in the C position and some XP in SpecC. Both types of complements as such appear to be transparent for *wh*-movement to SpecC. *Wh*-movement from a *dass* clause may go to a *dass* clause or to a verb-second clause; see (1-ab).[2] In contrast, as shown in (1-cd), *wh*-movement from a verb-second clause may only end up in a verb-second clause again (see Tappe (1981), Haider (1984), Reis (1985)).[3]

(1) a. (Ich weiß nicht) [$_{CP_1}$ wen$_4$ (dass) du meinst [$_{CP_2}$ t$'_4$ dass sie t$_4$ getroffen
   I know not   whom that you think   that she met
   hat]]
   has

   b. [$_{CP_1}$ Wen$_4$ meinst du [$_{CP_2}$ t$'_4$ dass sie t$_4$ getroffen hat]] ?
      whom think you   that she met has

   c. [$_{CP_1}$ Wen$_4$ meinst du [$_{CP_2}$ t$'_4$ hat sie t$_4$ getroffen]] ?
      whom think you   has she met

   d. *(Ich weiß nicht) [$_{CP_1}$ wen$_4$ (dass) du meinst [$_{CP_2}$ t$'_4$ hat sie t$_4$ getroffen]]
      I know not   whom that you think   has she met

---

[2] A complementizer of CP$_1$ in sentences like (1-a) must then be deleted in Standard German, but not in dialects and colloquial varieties. Following Pesetsky (1998), I assume that complementizer deletion is a PF phenomenon in languages like German and English, with a *that/dass* complementizer present in syntax proper.

[3] The representations in (1) presuppose that long-distance movement proceeds successive-cyclically, via intermediate SpecC positions, and leaves traces (*t*) in the positions it targets on its way to the ultimate landing site. At this point, these assumptions are mainly for convenience; they are shared by Staudacher's approach sketched below, though.

The same restriction as in (1-d) holds when movement from SpecV2 to Spec*dass* is followed by further *wh*-movement, as in (2); so the ultimate target position is not relevant.

(2) *$[_{CP_0}$ Wen$_4$ glaubt er $[_{CP_1}$ t$''_4$ dass du meinst $[_{CP_2}$ t$'_4$ hat sie t$_4$ getroffen]]] ?
   whom believes he that you think has she met

Furthermore, the restriction is also active when the moved item is a topic or relative pronoun, as in (3-ab) and (4-ab), respectively.[4]

(3) a. $[_{CP_0}$ Den Fritz$_4$ glaubt er $[_{CP_1}$ t$''_4$ dass du meinst $[_{CP_2}$ t$'_4$ dass sie t$_4$
      the Fritz believes he that you think that she
      getroffen hat]]]
      met has

   b. *$[_{CP_0}$ Den Fritz$_4$ glaubt er $[_{CP_1}$ t$''_4$ dass du meinst $[_{CP_2}$ t$'_4$ hat sie t$_4$
      the Fritz believes he that you think has she
      getroffen]]]
      met

(4) a. die Frau $[_{CP_1}$ die$_4$ (wo) du meinst $[_{CP_2}$ t$'_4$ dass sie t$_4$ getroffen hat]]
      the woman whom REL you think that she met has

   b. *die Frau $[_{CP_1}$ die$_4$ (wo) du meinst $[_{CP_2}$ t$'_4$ hat sie t$_4$ getroffen]]
      the woman whom REL you think has she met

2.2 Analyses

The data have proven remarkably robust over the years, and many attempts have been made to account for the asymmetry involved. First, it has been suggested that illicit instances of movement from a verb-second clause into a verb-final *dass* clause reveal a *locality* effect. On this view, the verb-second clause acts as a barrier in (1-d) (as well as in (3-b) and (4-b)). This then may or may not require some extra assumption about (1-c), where barrier status of the verb-second clause seems to be voided. Proposals of this type include Staudacher (1990), Sternefeld (1989) on the one hand, and Reis (1996), Müller (2004b), and Heck (2010) on the other. The former type of locality approach assigns verb-second clauses barrier status if they are embedded by a verb-final clause (as in (1-d)) but not if verb-second movement takes place in the higher clause (as in (1-c)). The latter type of locality approach treats verb-second clauses as barriers throughout (i.e, in (1-d) *and* (1-c)), and attributes the well-formedness of the string in (1-c) to an alternative option for analysis: (1-c) is assumed to involve a special kind of ("integrated") parenthesis (*meinst du* in (1-c); also cf. Kiziak (2007)).

---

[4] (3-a), (4-a) and, perhaps to a lesser extent, (1-b) are marked, or even not possible at all, in some varieties of German. In what follows, I will only be concerned with varieties that permit extraction from *dass* clauses across the board.

This option is assumed to be non-existent in (1-d).

Second, it has been proposed that the asymmetry in (4) follows from *directionality* constraints on movement of the kind postulated in Kayne (1984) (via 'g-projections') and Koster (1987) (via 'Global Harmony'); see Müller (1989, ch. 6) and Haider (1993b)). The basic idea here is that the apparent clash in directionality of government (or selection) by an embedded V in a verb-second position and a matrix V in a verb-final position blocks movement.

Third, the data have been approached in terms of constraints against *improper movement*; see Haider (1984), Stechow & Sternefeld (1988, ch. 11.7), Sternefeld (1992), Müller & Sternefeld (1993), Williams (2003). These approaches all differ substantially from one another (except perhaps for Sternefeld (1992) and Williams (2003), which both rely on the version of the cycle proposed in Williams (1974); cf. footnote 41 of chapter 4), but they share a common core: SpecV2 and SpecC positions are viewed as sufficiently different to be able to block movement from one to the other as improper. Furthermore, an asymmetry must be built into the theory of improper movement so as to make a mixing of landing sites possible in (1-b) (movement may take place from SpecC to SpecV2) but not in (1-d) (movement may not proceed from SpecV2 to SpecC).

Fourth, I have tentatively pursued an approach that relies on *shape conservation* (in the sense of Williams (2003)) in Müller (2003). The idea here is that there is a general but violable constraint that requires left edges of CP to have an identical shape. This constraint is satisfied in (1-a) and (1-c) but not in (1-b) and (1-d). Again, an asymmetry needs to be imposed, and this is achieved by local, cyclic, bottom-up optimization of CPs. In this approach, it turns out that the shape conservation constraint can be violated with verb-second movement in CP_1 in (1-b) because, given that CP_2 has already been optimized, other options that would be shape-preserving (specifically, failing to carry out verb-second) violate higher-ranked constraints. In contrast, shape conservation cannot be violated in (1-d) because, given that CP_2 (with verb-second movment) is already in place, verb-second movement can and must apply in CP_1 because this will give rise only to a violation of a lower-ranked constraint (the one that precludes movement which is not feature-driven).

Finally, there is a strategy that solves the problem in a very simple way, and that could be referred to as *data denial*: Ćavar (1996) and Fanselow & Mahajan (1996) claim that extraction from verb-second clauses into *dass* clauses as in (1-d) is possible after all.[5] This last view is at variance with the clear robustness that the effect has exhibited over the last thirty years or so. For this reason I will disregard this option in

---

[5] In a similar vein, Shin (1988, 253) claims that potential problems with the construction in (1-d) are not syntactic but pragmatic in nature.

what follows.

I cannot possibly make an attempt here to discuss all of these approaches in detail and highlight their merits as well as their shortcomings. I do believe, however, that virtually all of them rely on assumptions that are not really independently motivated, and that very often turn out to be construction-specific upon closer inspection. Furthermore, it can be noted that none of them is compatible with basic tenets of the minimalist program, where, e.g., there can be no designated locality constraints employing notions like barrier, no constraints that mention directionality, and no special constraints blocking improper movement; and where shape conservation effects must be treated as an epiphenomenon (recall the requirements imposed on 'good' constraints in (11) of chapter 2). Thus, if one adopts a minimalist perspective, as I will do here, it seems fair to conclude that the problem in (1) has not yet received a satisfying solution.

That said, I think that many of the basic insights of Staudacher (1990) (and, to some extent, also of Sternefeld (1989)) can be integrated into a new phase-based approach to the ban on extraction from verb-second clauses into verb-final clauses in German that does without construction-specific assumptions and meets basic minimalist demands; and I will try to show this below. To this end, I first summarize the proposal in Staudacher (1990) in the next section.

3.   Staudacher's (1990) Approach

Staudacher (1990) presents two analyses, which share a number of properties (in particular, they are both based on the hypothesis that there is a barrier present in (1-d) that blocks movement, and that verb-second in the matrix clause as in (1-c) opens up this barrier via transitivity of indexing) but differ in others (the first analysis, pp. 330-334, assumes the matrix VP to be a minimality barrier, whereas the second analysis, pp. 334-338, assumes the embedded verb-second clause to be a barrier because of non-selection). In what follows, I will focus on the second version of the approach since it is this version that the analysis to be developed below will resemble most.

The first thing to note is that embedded verb-second clauses that are embedded by a verb-final *dass* clause (as in (1-d)) are *strong* islands, in the sense that they uniformly block all kinds of extraction and do not distinguish between arguments and adjuncts; compare the cases of licit and illicit adjunct movement in (5-ab), which are fully parallel to the argument movement cases in (1-c), (1-d).

(5) a.   Ich weiß nicht $[_{CP_1}$ wie$_4$ du   meinst $[_{CP_2}$ t$'_4$ dass das t$_4$ gehen soll $]]$
            I   know not   how you think   that that   work should
     b. *Ich weiß nicht $[_{CP_1}$ wie$_4$ du   meinst $[_{CP_2}$ t$'_4$ soll   das t$_4$ gehen $]]$
            I   know not   how you think   should that   work

A standard version of the Condition on Extraction Domain (CED; Huang (1982),

Chomsky (1986a)) blocks both argument and adjunct movement from non-complements (where complements are sisters of lexical items); see (6).[6]

(6) *Condition on Extraction Domain* (CED):

    a.   Movement must not cross a barrier.

    b.   An XP is a barrier iff it is not a complement.

Assuming the CED to be responsible for the illformedness of (1-d), (5-b) and similar examples, we are led to look for a barrier in these contexts. A straightforward application of the CED in (6) to the data in (1) and (6) faces two basic problems. First, embedded verb-second clauses as in (1-c), (6-b) look like genuine complements of V – a verb like *meinen* ('think') subcategorizes for a complement and assigns a θ-role to it, and in the absence of an obvious alternative candidate, there is no good reason to assume that CP_2 is not the complement that the matrix verb is looking for, irrespective of whether verb-second movement has or has not applied in CP_2. And second, given the notion of barrier in (6), it is completely unclear how verb-second movement in the matrix clause could remove the barrier status of the embedded verb-second CP.

To solve the first problem, Staudacher proposes that the concept of barrier is to be refined in such a way that complement status alone does not suffice for an XP to avoid barrier status: Being a complement is viewed as a necessary but not yet sufficient condition for non-barriers. His revised notion of barrier is given in (7).

(7) *Condition on Extraction Domain* (CED, based on Staudacher (1990)):

    a.   Movement must not cross a barrier.

    b.   An XP is a barrier iff it is not *head-marked*.

Head-marking is a somewhat stricter notion than being a complement because it requires that, in addition to being a complement of a lexical item, *selection* of (or *co-indexing* with) the head of XP must take place:[7]

(8) *Head-marking* (Staudacher (1990, 336)):

    α head-marks β iff

---

[6] For the purpose of presenting and discussing Staudacher's (1990) analysis, I will to disregard the new version of the CED that was shown to follow as a theorem in chapter 4; but see below. – That said, Staudacher (1990) does not actually invoke the CED as such; rather he embraces the more complex system incorporating the CED that Chomsky (1986a) develops on the basis of Lasnik & Saito (1984). Where it does not affect the main points to be made, I tacitly simplify Staudacher's analysis in various respects, and also adjust it to more current terminology.

[7] (8) is a simplified version of Staudacher's definition of head-marking, but it suffices for present purposes – the complications Staudacher introduces in the "proper" definition in Staudacher (1990, 336) (and also already in Staudacher (1987, 13)) do not bear on the issue of extraction from V/2 clauses.

    a. α is a lexical item, and

    b. α is a sister of β that selects or is co-indexed with the head of β.

The basic idea is that CPs with a complementizer *dass* are always head-marked by the matrix verb (because the verb selects the head of the *dass* clause), whereas embedded verb-second clauses are not. Although embedded verb-second clauses are selected complements, their verb-second head, by assumption, is not selected by the matrix verb, and thus turns CP into a barrier. The assumption that heads of verb-second CPs are not selected receives independent justification – so Staudacher argues – from the observation that these heads are not filled by lexical material before verb-second movement applies. More generally, if *dass* complements are head-marked by V (via head-selection) and verb-second complements are not head-marked by V (via head-selection), this captures the insight that *dass* complements have a closer relationship with the matrix verb than verb-second complements, and it also captures the observation that embedded verb-second clauses, while having complement status as such, also exhibit several root properties. This also partially accounts for the conditions under which they are licensed in the first place.[8]

    Thus, a CP with a verb-second head is a barrier due to a lack of head-marking, in contrast to a CP with a lexical complementizer. This accounts for the contrast between examples like (1-a), (1-b) on the one hand (i.e., movement from *dass* clauses into *dass* clauses or into verb-second clauess), and examples like (1-d) on the other (i.e., movement from verb-second clauses into *dass* clauses).

    What remains to be explained is why verb-second movement in the matrix clause removes the barrierhood of an embedded CP with a verb-second head, as in (1-c). At this point, the option of bringing about head-marking by co-indexing (as an alternative to head selection) envisaged by (8) becomes relevant. After verb-second movement in the matrix clause, the following co-indexing relations obtain. First, since movement, by assumption, implies chain formation, and members of the same chain share an index, the verb-second head *meinst* and its trace in the V position are co-indexed. Second, for the same reason, the *wh*-phrase *wen* is co-indexed both with its original trace in the object position of the embedded clause and with the intermediate trace in the embedded SpecC position. Third, and this is the crucial step in the argument, Staudacher assumes that there is a process of abstract agreement between specifier and head, and that this agreement relation is also indicated by co-indexing. Given that verb-second movement is movement to C, from this it follows that the indices of the moved *wh*-phrase in the matrix SpecC position and the moved verb in the verb-second

---

[8] On root properties of embedded verb-second clauses and similar constructions, see Hooper & Thompson (1973), Meinunger (2004), and Bentzen (2009), among others.

position in the matrix clause are also identical. Fourth, by the same reasoning, the indices of the intermediate *wh*-trace in the embedded SpecC position and the embedded verb-second head are also identical as a consequence of abstract specifier/head agreement. And finally, by transitivity, this implies that the indices of the V trace of the matrix verb and the verb-second C head of the embedded clause are also identical in cases like (1-c). Consequently, joint verb-second movement and *wh*-movement in the matrix clause leads to head-marking of the embedded verb-second clause, and hence to a removal of barrier status there. This explains why (1-c) can respect the CED. Note that this way out is not available in (1-d) as long as it is assumed that co-indexing cannot arise accidentally: In (1-d), verb-second does not take place in the matrix clause, so the *wh*-item in the matrix SpecC position and the matrix verb are not co-indexed, and the transitivity chain is broken; as a result, matrix V cannot head-mark the embedded verb-second clause.

Summarizing, the gist of Staudacher's proposal is that embedded verb-second clauses are barriers due to a lack of head-marking by head selection, but the barrier status can be removed by combined verb-second and *wh*-movement in the matrix clause because this results in head-marking by head co-indexing.

I take this approach to be elegant and intuitively appealing, but it seems clear that it cannot be maintained as such under more recent minimalist assumptions as they are laid out in Chomsky (2001b; 2008) and much related work. For one thing, the analysis is based on a version of the CED (more precisely, on a system of constraints incorporating a version of the CED; see footnote 6) where the requirements imposed on good constraints in (11) of chapter 2 are not met. To the extent that it makes the right predictions, the CED should be derived as a theorem from more basic assumptions. This means that a different, arguably less stipulative reason must be sought that predicts barrier status of verb-second complement clauses in (1-d). For another thing, Staudacher's account of the removal of barrier status of verb-second complement clauses by verb-second movement in the matrix clause is not available under minimalist assumptions for a number of reasons. First, this account relies on indices and transitivity of co-indexing, but, following Chomsky's arguments, co-indexing is not a technical device that is legitimate under minimalist assumptions (let alone inhomogeneous co-indexing of the type that is needed in Staudacher's approach). Second, there is no room for a concept like specifer/head agreement if all Agree relations have their source on a lexical item (more precisely, a probe feature) and imply c-command by this item. And third, as it stands, Staudacher's account presupposes either a representational syntax or a derivational syntax that has look-ahead capacity (because the question of whether extraction from the lower clause is possible can only be answered once the edge domain of the matrix clause is reached); but phase-based minimalist analyses are characterized by the property of being (a) strictly derivational and (b) without look-ahead capacity. It is not clear whether the approach could be transferred

into a derivational system without look-ahead (or back-tracking).

In view of this state of affairs, Staudacher's approach must either be abandoned or modified appropriately. In what follows, I will show that a new analysis that maintains main insights of Staudacher's approach suggests itself on the basis of the approach to CED phenomena in chapter 4 (section 4) and a version of the approaches to repro-jection movement in Fanselow (2009) and Georgi & Müller (2010) (section 5). More specifically, following Staudacher (1990), I will develop an analysis in section 6 that rests on the idea that embedded verb-second clauses are barriers because they are not as strictly selected as *dass* clauses, and that the barrier status can be removed by verb-second movement in the matrix clause because such movement in effect produces (more or less, as in the head selection/co-indexing disjunction in (8-b)) the same con-figuration as it is present with *dass* clauses from the start.

4. CED Effects: A Phase-Based Approach

Recall from chapter 4 that CED effects can be derived from the PIC in (9); the version of the CED that thus emerges as a theorem is (10).

(9) *Phase Impenetrability Condition* (PIC):
 The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations.

(10) *Condition on Extraction Domain* (derived from the PIC):
 a. Movement must not cross a barrier.
 b. α is a barrier if the operation that has merged α in a phase Γ is the final operation in Γ.

Also recall that the analysis rests on four assumptions: First, all syntactic operations are driven by *structure-building* features ([•F•]) or by *probe* features that trigger Agree ([∗F∗]). Second, features on lexical items are ordered. Third, all phrases are phases. And finally, the crucial assumption is that edge features that are required to trigger intermediate movement steps to phase edges (which in turn are required by the PIC) can only be inserted on a phase head as long as the phase head is still active, i.e., as long as it has some structure-building or probe feature left. This is captured by the Edge Feature Condition in (11) (as before, to simplify matters I leave out the requirement that edge feature insertion is furthermore subject to the requirement that such an insertion is the only way to produce a balanced phase).

(11) *Edge Feature Condition* (EFC):
 An edge feature [•X•] can be assigned to the head γ of a phase only if (a) and (b) hold:
 a. γ has not yet discharged all its structure-building or probe features.

b. $[\bullet X\bullet]$ ends up on top of γ's list of structure-building features.

The following four consequences follow from these assumptions:

(12) a. A last-merged specifier is always a barrier (because the phase head is necessarily inert at the stage of the derivation where the last-merged specifier has come into being).

b. A non-last merged specifier is transparent (because, by definition, the phase head still has a structure-building feature left at the point where such a specifier has just entered the structure; this includes the melting phenomenon).

c. A non-last merged complement is transparent (for exactly the same reason that non-last merged specifiers are transparent).

d. A last-merged complement is transparent if the phase head still has a probe feature left after they have been merged, and opaque for extraction otherwise.

The last consquence is the one that will prove particularly important for the analysis of the pattern in (1). (13) illustrates again why last-merged complements do not have to be barriers. They are barriers if there is an empty stack of probe features on γ (see (13-a)), but they are not barriers if some probe feature is present (see (13-b)).

(13) *Last-merged complements as non-barriers*:

a. Edge feature insertion follows complement feature discharge, no probe feature:

$$\begin{array}{l}\gamma\colon\ [\bullet\beta\bullet]\\ \rightarrow\ \gamma\colon\ \varnothing\\ \rightarrow\ \gamma\colon\ [\bullet X\bullet]\end{array}\qquad\rightsquigarrow\ \text{violates (11-a)}$$

b. Edge feature insertion follows complement feature discharge, with probe feature:

$$\begin{array}{l}\gamma\colon\ [\bullet\beta\bullet]\\ [\ast F\ast]\\ \rightarrow\ \gamma\colon\ [\ast F\ast]\\ \rightarrow\ \gamma\colon\ [\bullet X\bullet]\\ [\ast F\ast]\end{array}\qquad\rightsquigarrow\ \text{violates nothing}$$

Recall finally that the question arises of which operation is carried out next with the phase head γ in (13-b): Is it discharge of the probe feature or is it discharge of the structure-building feature? The Strict Cycle Condition (SCC) provides the answer. A simple version of the SCC is given in (14) (= (23) in chapter 4, with Greek letters changed to avoid confusion in the present context).

(14) *Strict Cycle Condition* (SCC):

Within the current domain ζ, a binary syntactic operation may not exclusively apply to positions ϵ, τ if ϵ and τ are both included within another domain π that is dominated by ζ.

If the phase head γ is merged with α that has undergone movement out of β, with α therefore becoming γ's specifier, subsequent discharge of the probe feature [∗F∗] on γ with some item in β will have to violate the SCC because this latter operation exclusively applies to positions that are included in γ' (= π in (14)), which is dominated by the current domain γP including α (= ζ in (14)). An a priori possible way out might be to carry out Agree between [∗F∗] on γ and α in the specifier position (or with some category included in α); however, this would violate the c-command requirement on Agree operations. Thus, the next operation carried out with the phase head in (13-b) is discharge of [∗F∗] by Agree with some item in β, and this is then followed by movement of α from the edge of β to the edge of γ, which discharges [●X●]. Finally, the interaction of the SCC and the c-command requirement on Agree operations ensures that this way out is available only for complements, not for specifiers.

As noted, for an account of the pattern in (1), the barrier status of last-merged complements is highly relevant: Only when (a) a complement is last-merged and (b) there is no Agree operation with the head that it is a complement of can a complement become a barrier. In section 6, I will argue that such an exceptional situation arises in the case of verb-second complements embedded by a verb-final clause. However, before that, something needs to be said about verb-second movement.

5. Verb-Second by Reprojection

Verb-second movement in German is often conceived of as adjunction to C that proceeds by intermediate adjunction to T and v, as in (15).

(15) *Verb-Second, a standard view*:
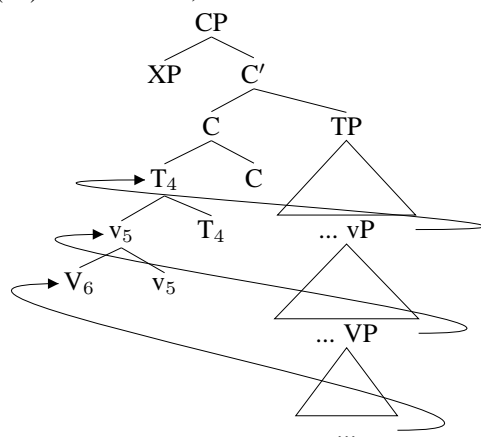
There are three potential problems with this view. The first problem is that there do not seem to strong empirical arguments for assuming verb-second movement to proceed from V to v, then from v to T, and finally from T to C (see Haider (1993a) and Roberts (2009; 2010), *pace* Sabel (1996)). And assuming that it does so nevertheless gives rise to various technical problems: Strictly speaking, C attracts T (not V), T attracts v (not V), and only v attracts V, so the existence of full verb-second movement of V to C emerges as a fortunate coincidence going back to a conspiracy of three separate movement rules. If the view embodied in (15) is correct, we might expect phenomena (in minimally differing varieties of German, e.g., earlier stages of the language, or regional variants) like bare T-to-C movement or V-to-v-to-T movement stopping there, for which there is no convincing evidence (see the above references for discussion). These considerations lead me to conclude that verb-second movement does not proceed via successive adjunction to higher heads.

The second problem is that one never finds cases in German where there is a realization of C as *dass* together with an adjacent verb-second head, as in (16-a). Given that there are varieties of German that do not respect the Doubly-Filled Comp Filter (i.e., that allow a simultaneous realization of *dass* and some other item in the CP edge domain), this is unexpected. Furthermore, it seems that there *are* substandard instances of a simultaneous realization of *dass* and verb-second in Modern German after all (and, depending on the analysis, perhaps also in Old High German), but these look exactly like their Scandinavian counterparts, with a sequence C+topic+V2, as in (16-b) (see Lenerz (1984), Müller & Sternefeld (1993); and Axel (2007) and Freywald (2009) for recent discussion concerning Old High German and Modern German substandard varieties, respectively).

(16) a.   *(Ich glaube,) [$_{CP}$ sie hat$_4$-dass [$_{TP}$ ihn getroffen t$_4$]]
            I   believe   she has-that   him met

     b.   */$^{#}$(Ich glaube,) [$_{CP}$ dass [den Fritz hat [$_{TP}$ sie getoffen t$_4$]]
            I   believe   that   the Fritz has   she met

The third problem concerns the nature of head movement as adjunction to another head in general. As is well known, this view of head movement creates several problems with respect to highly general (and independently motivated) constraints on movement, e.g., the c-command requirement for traces (which is a subcase of the SCC; cf. (23) of chapter 4); see Brody (2001), Mahajan (2001), Abels (2003), Müller (2004b), and Matushansky (2006) (among others) for discussion. This problem can be solved by assuming that head movement is not adjunction but reprojection, in the sense that a head moves out of a phrase and remerges with it, projecting its label in the derived position. Reprojection movement of a finite verb is explicitly assumed for German verb-second clauses in Stechow & Sternefeld (1988), based to some extent on Reis (1985) who had argued that *dass* clauses and verb-second clauses should be

treated as categorially different (essentially CP vs. VP). Both Stechow & Sternefeld (1988) and Sternefeld (1989) assume that treating verb-second as reprojection movement is a precondition for deriving the asymmetry between *dass* clauses (which are assumed to be CPs) and verb-second clauses (VPs) with respect to extraction into *dass* clauses. Subsequent approaches that analyze instances of verb movement (including, in some cases, verb-second movement) by reprojection include Holmberg (1991), Ackema, Neeleman & Weerman (1993), Koeneman (2000), Haider (2000a), Hornstein & Uriagereka (2002), and Fanselow (2003; 2009).[9]

In what follows, I adopt a reprojection analysis of verb-second clauses in German that combines aspects of the approaches to reprojection in Fanselow (2009), Georgi & Müller (2010) and other work just mentioned (but that nevertheless differs from all these approaches in certain minor respects).[10]

The basic idea is that a head γ may be equipped with a certain probe feature [∗F∗] that it cannot possibly check in situ, for the simple reason that there is no goal feature around that it might check it with. There are two possibilities instantiating this general state of affairs: Either there is no matching goal feature in the c-command domain of γ at all (recall that Agree requires c-command), or there is a matching goal feature that has already been checked with some other probe earlier in the derivation, so it is not available anymore for γ. In this case, γ may, as a last resort operation (a concept which is not to be confused with the Last Resort condition introduced in (6) of chapter 3 and modified in (10) of chapter 4 above), undergo reprojection movement in order to find a matching goal and discharge its probe feature [∗F∗] under c-command. [∗F∗] thus in effect acts as a *Münchhausen* feature.[11]

---

[9] As noted by Fanselow (2009), reprojection analyses of verb-second in German to some degree resemble the classic approaches developed in Bierwisch (1963) and Thiersch (1978), which had been superseded by the standard analysis in (15) going back to den Besten (1977).

[10] As a side remark, and anticipating the analysis of the data in (1) given in the next section, let me point out that, unlike Stechow & Sternefeld (1988) and Sternefeld (1989), I will not so much exploit reprojection movement and the categorial difference (CP vs. VP) between *dass* clauses and verb-second clauses that goes along with it to account for the difference in grammaticality between (1-a) and (1-d); i.e., the basic status of verb-second clauses as barriers will not necessarily be tied to their being VPs rather than CPs (although this will turn out to be one of several possibilities). Rather, reprojection movement will be required to explain the difference between (1-c) and (1-d); i.e., it will offer a simple account of the fact that verb-second movement in the matrix clause destroys barrierhood and turns an embedded verb-second clause into a transparent domain.

[11] The name is self-explanatory: Baron Münchhausen escapes from a swamp (where he is trapped on the back of his horse) by pulling himself up by his hair. The use of the name 'Münchhausen' in syntactic theory arguably goes back to Sternefeld's (1991b) characterization of an operation employed in Chomsky's (1986a) theory of barriers: Here, VP is a barrier, but a V moved to I can belatedly justify its own (originally impossible) movement across the VP barrier by L-marking VP and removing barrierhood – clearly a case

The cases of reprojection movement investigated in Georgi & Müller (2010) are typically highly local; in particular, in that paper we are concerned with word order variation in nominal projections that is derived by extremely local reprojection movement of N.[12] In contrast, verb-second movement by reprojection cannot be such an extremely local movement operation if verb-second clauses in German have a full TP-vP-VP structure, and if verb-second movement targets a position outside of TP. Indeed, there seems to be evidence in abundance for assuming that verb-second clauses in German have a full TP-vP-VP structure. For instance, in *dass* clauses as well as in non-subject initial verb-second clauses, only subjects can precede unstressed pronouns (except for items in the *Vorfeld*), which follows if there is a designated TP category whose specifier acts as a target for optional subject raising (as argued in Grewendorf (1989)), and scrambling cannot go beyond the vP/VP domain (see chapters 3 and 4 above). Moreover, asymmetries between (in-situ) subjects and other arguments with respect to subcategorization by V, verb phrase topicalization, extraction, binding, and so on, are the same in *dass* clauses and verb-second clauses, and to the extent that these differences motivate a vP/VP distinction, they do so in both contexts. Similarly, it is clear that, since verb-second movement may end up in a position preceding a subject that in turn precedes an unstressed object pronoun (which in turn may precede a non-pronominal object), it must target a position outside of TP. The question then is: How does V get there?

Here is a suggestion. Suppose that the Münchhausen probe feature on a V that is to undergo verb-second movement by reprojection is [∗T∗] (see Fanselow (2009, sect. 3.3)); such a feature is optionally instantiated on V in the numeration, and it indicates a special relation that certain kinds of V (viz., verb-second Vs) have to enter with a TP projection. Suppose further that, if V takes an object with clausal structure inside, [∗T∗] can never be discharged with a lower TP that it c-commands. This is obvious if the embedded clause is a verb-second clause itself (because in this case, the embedded verb-second head will have undergone Agree with TP already, and TP is not accessible for Agree anymore since it does not have an active goal). We may assume that it also holds in cases where a proper C (like *dass*) subcategorizes TP, perhaps because C also has a probe feature [∗T∗] in addition to the subcategorization feature [•T•] that triggers merge of C and TP; note that since a last-merged comple-

---

of pulling oneself up by one's own hair. (Incidentally, Staudacher's (1990) analysis in terms of transitivity of co-indexing is not too dissimilar in this respect.) Fanselow (2003) applies the concept to reprojection movement ('Münchhausen-style head movement').

[12] This in turn solves a number of problems that arise under the classic view that nominal projections are NPs, with DP as a specifier of N, rather than DPs, with NP as a complement of D. However, this issue is orthogonal to my present concerns, so I will continue to assume that nominal projections are DPs.

ment TP of C is transparent for extraction, the approach to CED effects developed in chapter 4 presupposes that there is *some* probe feature on C that keeps C active for edge feature insertion in order to move something out of TP's edge domain.[13] Thus, irrespective of whether there is an embedded T in the c-command domain or not, the only way to discharge a [∗T∗] feature of V is to carry out reprojection movement. Consequently, $V_{[∗T∗]}$ undergoes movement to the next phase edge, and from there to the next phase edge, until it finally reaches a position from which it can reproject a VP and discharge its Münchhausen feature [∗T∗] under c-command of T – i.e., the movement operation as a whole is non-local, but like all other movement operations, long-distance reprojection movement is composed of a series of smaller steps, as required by the PIC (see chapter 3). This derivation is shown in (17) (with ○ marking the intermediate positions successively occupied by V on its way to its reprojection position, and the dashed arrow from V to T indicating the Agree relation that provides the trigger for verb-second movement).

(17) *Verb-Second by reprojection*:

---
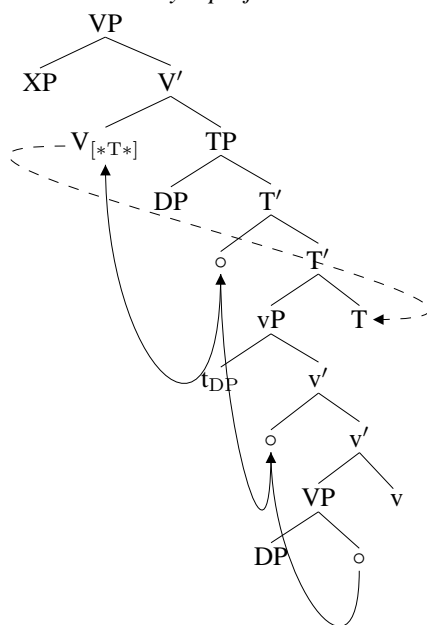
[13] If it is assumed that some other feature than [∗T∗] is involved in C-T Agree, this would make the same prediction as long as it is ensured that this Agree operation makes T an inactive goal for Agree with a higher V.

The assumption that verb-second in German looks as in (17) gives rise to various questions.[14] An obvious one concerns the Head Movement Constraint (HMC): This analysis of verb-second in terms of reprojection movement of V to the TP domain takes place via local steps (viz., via all intermediate edge domains); however, it clearly violates the HMC because the intervening heads v and T are skipped by the operation. Arguably, though, this is not a problem since the HMC can be shown to be too strict anyway; see in particular Roberts (2009; 2010), who argues that various kinds of head movement operations must be able to freely cross intervening head positions (among them long verb movement in Breton, V movement in predicate cleft constructions in Spanish and Hebrew, clitic-climbing in Italian and, incidentally, also verb-second movement in German).

Second, we may ask whether the intermediate movement steps of $V_{[*T*]}$ are triggered by edge features, like all other cases of successive-cyclic movement. Indeed, there is every reason to assume that they are (note particularly that all extraction takes place from complements, so edge features can be inserted in line with the Edge Feature Condition in (11)).

Third, why does V reproject, rather than just ending up as a further specifier of TP? A simple (and standard) answer here could be that if two items are combined, only one of them can have the feature that induces this operation, and the item that does will always be the one that projects.[15]

---

[14] It also makes some further interesting predictions. Consider, for instance, the well-known fact that some morphologically complex verbs can occur in finite form in a verb-final position, but fail to participate in V/2 fronting (see Haider (1992) and Koopman (1995), among many others):

(i) a.   dass sie die Oper hier ur-auf-führen
           that they the opera here perform first
     b.   *Sie führen die Oper hier ur-auf
     c.   *Sie auf-führen die Oper hier ur
     d.   *Sie ur-auf-führen die Oper hier

This is a priori difficult to account for if verb-second movement can uniformly be traced back to some attracting feature on an invariant C; in the present approach, one can simply assume a redundancy rule for certain complex verbs that systematically blocks the presence of [*T*] on such a verb.

[15] Upon closer scrutiny, a bit more must be said, though. Why can't V stay in the specifier position of T, discharging its probe feature [*T*] and projecting there? If it does, the structure-building feature for T's DP specifier (the EPP property) can never be satisfied. So T must be able to remain in a position for a while from which checking would in principle be possible, and then move again to a further position outside of TP to actually carry out the checking (otherwise the SCC would be violated). Note that since this last operation technically goes to a position *beyond* the TP phase, the earlier intermediate step to a SpecT position is required. – That said, nothing in what follows would inherently be incompatible with an analysis of verb-second as movement to a specifier of TP, with V' in (17) reinterpreted as T' or TP; to reach this result, one would have to give up the assumption that the label of a complex category is determined by the daughter that has contributed the operation-inducing feature that is responsible for creating that complex

Fourth, it is so far not clear how the specifier XP of the verb-second head V comes into existence that closes off the VP in (17). In the general type of approach adopted here, there must be a structure-building feature that triggers placement of some XP in the first position. There are various possibilities as to how this can be implemented in the present system. For concreteness, suppose that a V category that is equipped with the feature [∗T∗] inducing verb-second must (or may, depending on the analysis of apparent verb-first structures) also bear a structure-building feature creating a specifier in the reprojection position. Such a feature may trigger *wh*-movement ([●wh●]), as in the cases in (1), or topicalization ([●top●]), as in an example like (3), but typically not relativization ([●rel●]), since relative clauses in German standardly do not involve verb-second – cf. (4), but also Gärtner (2000) for some exceptions. In order to avoid a discharge of this structure-building feature while V is still in situ, it can be assumed that a feature like [●wh●] or [●top●] on $V_{[∗T∗]}$ comes with a diacritic that makes its discharge dependent on a prior discharge of the probe feature [∗T∗] driving verb-second. This is shown in (18) for *wh*-movement and topicalization respectively, on the basis of the entry for V in (17).[16]

(18) a.   V: [●DP●] ≻ [●wh●]$^{([∗T∗])}$, [∗T∗]
      b.   V: [●DP●] ≻ [●top●]$^{([∗T∗])}$, [∗T∗]

In the course of the derivation, a V as in (18-ab) first discharges [●DP●] by merging with a DP that becomes its complement; then it undergoes successive-cyclic movement steps via intermediate edge domains (forced by the PIC, and permitted by the EFC, given an obvious extension of the effect on outcome/balanced phase requirement) until it can discharge its Münchausen feature (the categorial probe feature [∗T∗]) by reprojection, taking TP as its complement; and finally, this opens up the possibility of discharging the movement-inducing feature ([●wh●] or [●top●]) by creating a specifier; in a sense, [∗T∗] on V locks an operator movement feature, and the discharge of [∗T∗] unlocks it.[17]

---

category.

[16] I abstract away here from other possible probe features on V that it may check with its argument DP or some item included in it. Also, to ensure compatibility with the analysis of operator island effects in chapter 5, [●top●] should really be viewed as [●op●], and [●wh●] as a shorthand for [●op,wh●]; see (15) in the previous chapter.

[17] As noted, alternative approaches are possible. One could, for instance, assume that structure-building features and probe features do not show up on two separate stacks, but are actually ordered with respect to one another on a single stack. In the case at hand, an initial order [●DP●] ≻ [∗T∗] ≻ [●wh●] would make it possible to do without the diacritic on the movement-inducing feature, but such a theory is arguably more stipulative than the approach envisaged in the main text because an order like [●DP●] ≻ [∗T∗] (or a generalized variant thereof) follows from the general make-up of the theory (a probe feature can only be

As a side remark, note that it might be that the present approach to filling the first position in verb-second clauses in German is not yet fine-grained enough. Frey (2004b) and Fanselow (2003; 2009) argue that one should formally distinguish between unmarked realizations and marked realizations of the first position in verb-second clauses, with the former case involving, apparently, just items at the left periphery of the verb-second clause that would normally also show up at the edge of the TP or vP domain, and the latter case involving information-structurally marked items, as, e.g., in *wh*-clauses and long-distance topicalizations. If this is right, one might want to address it by permitting an underspecified edge feature [•X•] as a third option for an intrinsic feature on a verb-second V: [•X•] discharge would then imply moving whatever happens to be the closest TP- or vP-internal item to the specifier of the reprojecting verb (given a constraint like the (generalized) Minimal Link Condition; see Fanselow (1991), Chomsky (2001b), and the PIC-based reconstruction in chapter 3), whereas [•wh•] (or [•top•]) discharge would imply moving the closest *wh*-item (or [top]-marked item). – Note incidentally that topic islands are strict independently of the issue of marked vs. unmarked realization of the pre-verb-second position. In order to fully maintain the account of operator islands given in the previous chapter (especially that of topic islands) under the assumption that [•X•] may be involved in the formal operation of topicalization, one would have to assume that [•X•] can be marauded by [•op,wh•] and by [•op•]. This assumption strikes me as plausible, given that X basically functions as a variable ranging over all kinds of A-bar related features.

Fifth, the question arises why C cannot embed a verb-second VP (abstracting away from marked configurations like the one in (16-b)). Standardly, the complementary distribution is derived by assuming categorial identity, which the present approach does not.[18] In the present approach, a simple analysis suggests itself: C is equipped with a structure-building feature [•T•], but reprojection movement of V in verb-second clauses has created a VP, which C cannot subcategorize. Consequently, we are led to the conclusion that C only shows up optionally in a numeration (or subarray) – if it does in the presence of verb-second movement, the derivation will

---

discharged if there is some category present that provides a goal) and thus does not have to be stated as such.

[18] However, note that categorial identity in the strict sense is not available under the standard approach either, given structures like (16-a); and also note the problem raised by (16-a) in this context. Furthermore, Stechow & Sternefeld (1988, 402-405) show that the idea of a complementary distribution of verb-second and *dass* is problematic to begin with.

crash.[19]

Much more would ultimately have to be said about a reprojection approach to German verb-second clauses. However, I will refrain from doing so since the only thing that is of fundamental importance for the PIC-based approach to the data in (1) to be developed in the next section is really just the assumption (adopted from Fanselow (2009)) that a V that undergoes verb-second movement in the derivation has an additional operation-inducing feature that is absent when V stays in situ, in a *dass* clause.

6. A PIC-Based Approach

6.1 *Extraction from Last-Merged Clausal Complements*

Now the system is basically in place that makes it possible to derive the the pattern in (1) from the PIC, in a way that incorporates Staudacher's (1990) hypothesis that verb-second clauses are barriers because they are not as strictly selected as *dass* clauses, and that the barrierhood is lifted by verb-second movement because this gives rise to a configuration that resembles the one found with *dass* clauses. There is just one more assumption that is needed: In the examples in (1), the most deeply embedded clause is both first-merged (i.e., a complement) and last-merged. This implies that the embedded clause can only be transparent for extraction if V is equipped with another operation-inducing feature that keeps it active (for edge feature insertion to be possible) after the clause is merged. As I will show momentarily, this state of affairs lies at the core of the 3-out-of-4 pattern in (1). However, the grammaticality status does not vary noticeably when an additional argument is added, as in (19-ab), which can be compared with (1-a) and (1-d), respectively (the added argument is in italics).[20]

---

[19] A further derivation that must be ruled out has C merging with TP before reprojection movement of V. The illformedness of such a derivation follows if last resort operations like reprojection movement obey an Earliness requirement, as assumed in Georgi & Müller (2010): The Münchhausen feature is discharged as soon as possible.

[20] At least, there is no systematic improvement of the verb-final/verb-second combination; quite on the contrary, as a tendency the results get somewhat worse throughout. Incidentally, (19-b) is not quite fully well formed even if there is no movement from the embedded verb-second clause; cf. (i-a) vs. (i-b).

(i)   a.   (Ich weiß) [dass du   gesagt hast [sie hat ihn getroffen]]
              I   know   that you said   have   she has him met

      b.   ??(Ich weiß) [dass du *ihm* gesagt hast [sie hat ihn getroffen]]
              I   know   that you him said   have   she has him met

If this effect could be shown to be systematic, no additional assumption would be required for the cases in (19). However, there still seems to be a contrast between (i-b) and (19-b), and this implies that the illformedness of (19-b) cannot solely be due to a general non-licensing of embedded verb-second in the presence of a second argument in VP.

(19) a. ?(Ich weiß nicht) [wen$_4$ (dass) du *ihm* gesagt hast [t$'_4$ dass sie t$_4$
            I   know not   whom that you him told   have   that she
          getroffen hat]]
          met   has

      b. *(Ich weiß nicht) [wen$_4$ (dass) du *ihm* gesagt hast [t$'_4$ hat sie t$_4$
            I   know not   whom that you him told   have   has she
          getroffen]]
          met

Against the background of the theory developed so far, the conclusion to be drawn
from this is that clausal complements are always the only items merged in a VP; a fur-
ther non-subject argument will then be merged in some separate projection between
vP and VP. Therefore, a structure-building feature for an indirect object can never
make extraction from a CP possible by keeping the V head active. Possibly, this as-
sumption could be generalized, such that indirect objects and direct objects are always
merged in different projections, even when the direct object is non-clausal. On this
view, extraction from a complement of V always depends on the availability of some
additional probe feature on V (also recall the pertinent discussion in chapter 4).

    Here, then, is the analyis of the pattern in (1): Both *dass* complements and verb-
second complements enter the structure via a subcategorization feature on V ([●C●]
and [●V●], respectively). However, the former clause type receives a special iden-
tification in the form of a probe feature on V that agrees with it; such a feature is
absent from V if it embeds a verb-second clause. For present purposes, it does not re-
ally matter what exactly this additional probe feature on V that discriminates between
dass clauses and verb-second clauses looks like. One might speculate that it is a case
feature (assuming C to be nominal, CPs might need case, which verbal categories do
not); but this question is discussed controversially in the literature. An alternative
would be to postulate abstract ϕ features on a *dass* clause (but not on a verb-second
clause). For now, I will simply assume that the probe feature in question is a *categorial*
probe feature of just the type encountered in the previous section, the only difference
being that it does not trigger reprojection movement because it can be checked by V
under c-command. Thus, in the case of a *dass* clause, matrix V may bear a probe
feature [∗C∗]; but there is no comparable categorial probe feature on matrix V for a
verb-second clause.[21] I take this to be a relatively straightforward and faithful transfer
of Staudacher's (1990) distinction in terms of head-marking via head-selection (*dass*

---

[21] At least at first sight, one may think that presence vs. absence of [∗C∗] (in addition to [●C●]) on V
captures the bridge/non-bridge distinction in much the same way that the the approach sketched in section
3.2.4 of chapter 4 does (also cf. the pseudo-melting approach tentatively envisaged in section 7.2 of the
appendix to chapter 4). However, we will later see that this is not the case. For the moment, the bridge/non-
bridge distinction with extraction from CP will be ignored.

clauses are head-selected, verb-second clauses are not).[22]

6.2 *Extraction from 'dass' Clauses*

Given the approach to CED effects in chapter 4, this accounts for the transparency of *dass* clauses: At the point of the derivation where an edge feature needs to be inserted on the matrix V to make movement from the edge of CP to the edge of VP possible, V is active because it still bears [∗C∗]. The reason why *dass* complements of V are transparent for extraction (see (1-a), (1-b)) thus follows in exactly the way outlined in (13-b) above for last-merged complements in general. Consider first movement from a *dass* clause into a *dass* clause, as in (1-a). (1-a) is repeated here as (20).[23]

(20) (Ich weiß nicht) [$_{CP_1}$ wen$_4$ (dass) du   [$_{VP}$ t$''_4$ meinst [$_{CP_2}$ t$'_4$ dass sie t$_4$
       I   know not   whom that you   think   that she
      getroffen hat]]]
      met   has

The changes in feature-composition on the matrix V *meinst* in (20) are illustrated in (21).

(21) *CP complements as non-barriers, matrix V in situ*:

	V:	[•C•]
		[∗C∗]
→	V:	[∗C∗]
→	V:	[•X•]
		[∗C∗]

⤳ violates nothing
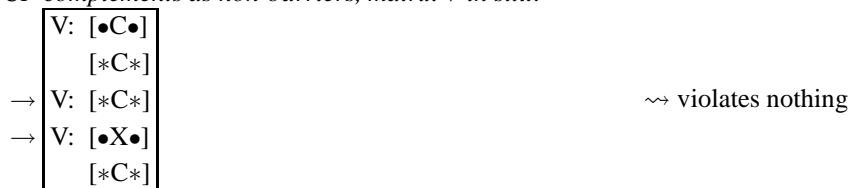
After the edge feature [•X•] is inserted, [∗C∗] is discharged, triggering Agree, and finally, [•X•] is discharged, triggering extraction from the CP complement. The relevant part of the derivation of (20) is shown in (22).[24]

---

[22] Why can a V with a feature [•V•] not merge with a (finite or non-finite) bare VP with an in situ V head that is not equipped with [∗T∗]? Perhaps the most straightforward answer might be that it can, but that the resulting structure is uninterpretable since the lower "clause" does not have a subject (in addition, an object in the lower VP could not have its case feature checked).

[23] Although there are many more intermediate movement steps, given that all phrases are phases, I add here only the trace in the matrix VP position that signals that the movement step from the CP$_2$ edge to the matrix VP edge is legitimate. (As before, I insert traces here only for convenience; see chapter 3.)

[24] Dashed lines indicate that the operation has not yet been carried out at this stage of the derivation, and that the structure has not yet been created; the arrow ╱ signals edge feature insertion.

(22) *CP complements as non-barriers, first case*:
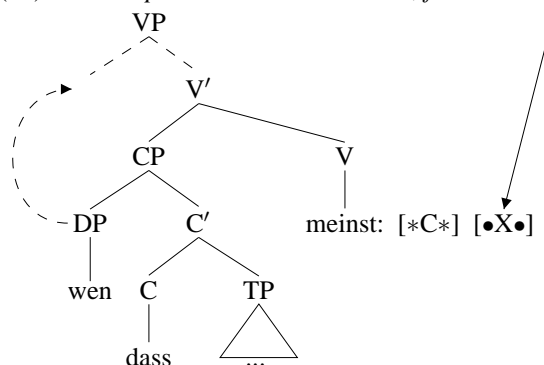
Note that this account of *dass* clause transparency is completely independent of the issue of verb movement in the higher clause. Therefore, sentences like (1-b), with movement from a *dass* clause into a verb-second clause, receive essentially the same treatment; (1-b) is repeated here as (23) (with CP_1 replaced with VP_1, in line with the new analysis of verb-second developed in section 5).

(23) $[_{VP_1}$ Wen$_4$ meinst du $[_{VP}$ t$_4''$ $[_{CP_2}$ t$_4'$ dass sie t$_4$ getroffen hat $]]]$?
           whom think   you   that she   met   has

The only relevant difference to (20) is that matrix V has an additional probe feature, viz., the [∗T∗] feature that triggers reprojection movement. Given that probe features are ordered on their stack in the same way that structure-building features are on theirs, it must be assumed that the order is [∗C∗] (to be discharged with the CP complement by V in situ) ≻ [∗T∗] (to be discharged with the matrix TP after reprojection raising of V).[25] Thus, edge feature insertion can respect the EFC on matrix V, as indicated in (24).

(24) *CP complements as non-barriers, matrix V undergoes V2*:

	V: [•C•]	
	[∗C∗] ≻ [∗T∗]	
→	V: [∗C∗] ≻ [∗T∗]	⤳ violates nothing
→	V: [•X•]	
	[∗C∗] ≻ [∗T∗]	

---
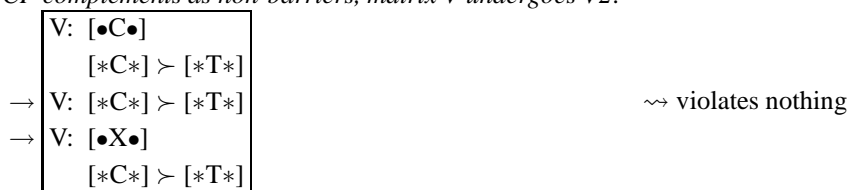
[25] Then again, no aspect of the present analysis really forces the view that probe features are ordered, too. In the present context, this assumption is mainly (and weakly) justified by considerations pertaining to homogeneity and uniformity of feature stacks.

Except for the additional [∗T∗] feature, the relevant part of the derivation of (23) is identical to what is shown in (22); see (25).

(25) *CP complements as non-barriers, second case*:

meinst: [∗C∗] ≻ [∗T∗] [•X•]

```
                  VP
               ⌒
              V′
          CP        V
       DP    C′      │
      wen  C   TP
           │
         dass   ...
```
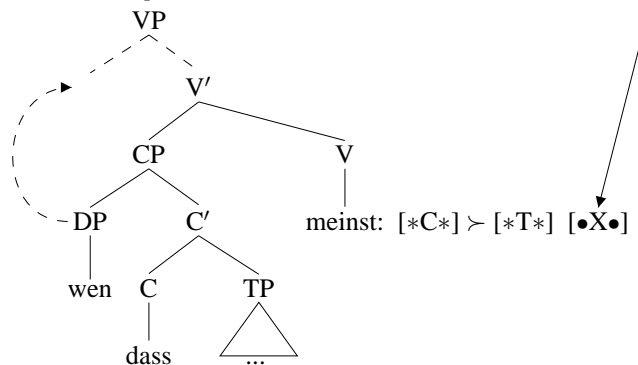
6.3   *Extraction from Verb-Second Clauses*

Things are different with embedded verb-second clauses. Let us turn to movement from a verb-second clause into a verb-second clause first; the example in (1-c) is given again in (26) (with CP replaced by VP, and an intermediate trace in the edge domain of the matrix VP added).

(26) [$_{VP_1}$ Wen$_4$ meinst du [$_{VP}$ t$_4''$ [$_{VP_2}$ t$_4'$ hat sie t$_4$ getroffen]]] ?
        whom think   you   has she   met

(26) is well formed, which implies that an edge feature can be inserted on matrix V for an item in the edge domain of its sole verb-second complement even though verb-second complements, by assumption, are not identified by a categorial probe feature (that would keep the V head active), like *dass* clauses are. It is at this point that the assumption made in section 5 becomes relevant according to which verb-second movement is brought about by a Münchhausen feature on V that triggers reprojection movement (viz., [∗T∗]). Crucially, this feature is present on matrix V in its in situ position in (26), and this is the reason why V is still active (after it has merged with the verb-second clause), and therefore accessible for edge feature insertion. The changing feature structure on matrix V in the relevant part of the derivation is shown in (27).

(27) *VP complements as non-barriers, matrix V undergoes V2*:

```
      ┌─────────┐
      │ V: [•V•] │
      │   [∗T∗]  │
  →   │ V: [∗T∗] │              ⤳ violates nothing
  →   │ V: [•X•] │
      │   [∗T∗]  │
      └─────────┘
```

The decisive stage of the derivation itself is illustrated in (28).

(28) *VP complements as non-barriers*:

```
                    VP
                    ╱ ╲
              ┌─ ▸     V′
              ┆      ╱    ╲
              ┆    VP       V
              ╲   ╱  ╲      │
               ─ DP   V′   meinst: [∗T∗] [•X•]
                  │   ╱  ╲
                 wen V    TP
                     │    ╱╲
                    hat  /__\
                         ...
```
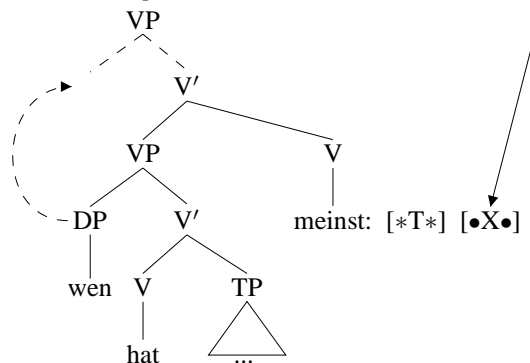
In the three cases discussed so far, there is at least one probe feature on matrix V that keeps the head active and thereby permits edge feature insertion for movement out of the clausal complement: In (20), there is [∗C∗]; in (26), there is [∗T∗], and in (23), there is both [∗C∗] and [∗T∗]. However, in the case of movement from a verb-second clause into a *dass* clause as in (1-d), repeated here as (29) (with slightly modified labelling), there is no [∗C∗] on the matrix V (because the verb-second complement is not strictly selected, or "head-selected", in Staudacher's (1990) terminology), and there is no [∗T∗] on the matrix V either (because the matrix clause is not a verb-second clause).

(29) *(Ich weiß nicht) [$_{CP_1}$ wen$_4$ (dass) du meinst [$_{VP_2}$ t′$_4$ hat sie t$_4$ getroffen]]
      I   know not   whom that you think   has she   met

Consequently, the matrix V head becomes inert after it has discharged its subcategorization feature and merged with the verb-second complement: Edge feature insertion cannot apply, and a PIC violation therefore cannot be avoided once the derivation has proceeded beyond the matrix VP phase. This accounts for the illformedness of (29). The impoverished feature structure of matrix V in this derivation is shown in (30).

(30) *VP complements as barriers, matrix V stays in situ*:

$$\rightarrow \boxed{\begin{array}{l} \text{V: } [\bullet V \bullet] \\ \text{V: } \text{---} \end{array}} \qquad \rightsquigarrow [\bullet X \bullet] \text{ insertion impossible}$$

The problematic step of the derivation itself is illustrated in (31).

(31) *VP complements as barriers*:
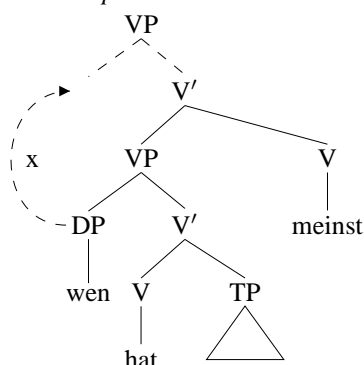
This concludes the account of the pattern in (1) that is the subject of the present chapter: The restriction on movement from verb-second clauses to verb-final clauses can be derived if Staudacher's (1990) main assumptions are transferred into the PIC-based account of CED effects in Müller (2010b) and enriched by a reprojection approach to verb-second movement that locates the operation-inducing feature on V, as in Fanselow (2009). The one thing that has been lost from Staudacher's original analysis in the course of doing so is transitivity of indexing as a means to remove barrier status from an embedded verb-second clause; since this device is either representational or demands look-ahead capacity, and since its work can straightforwardly be done by the reprojection-inducing probe feature on matrix V against the background of the theory developed in chapter 4, I take this loss to be unproblematic.

7.   Some Further Issues

Of course, the approach just sketched raises a number of further issues. I will confine myself to addressing three of them here.

7.1   Periphrastic Verbs Forms

First, the question arises of what happens with periphrastic verb forms where, say, an auxiliary undergoes verb-second movement in the matrix clause and the main verb stays in situ. As shown in (32-ab), the pattern remains the same as in (1-c), (1-d) (= (26), (29)):

(32) a. $[_{VP_1}$ Wen$_4$ hast du $[_{VP}$ t$''_4$ gemeint $[_{VP_2}$ t$'_4$ hat sie t$_4$ getroffen]]] ?
           whom have you   thought   has she   met

      b.   *(Ich weiß nicht) $[_{CP_1}$ wen$_4$ (dass) du gemeint hast $[_{VP_2}$ t$'_4$ hat sie t$_4$
           I   know not   whom that you thought have   has she
           getroffen]]
           met

The challenge here is to account for the fact that raising of the higher auxiliary can lift barrierhood of the verb-second clause even though the main verb stays in situ in the matrix clause. There are several ways to solve this problem.[26] A first strategy might be to assume that in fact not all phrases are phases after all: In extended (verbal) projections, only the highest projection counts as a phase. This would derive the pattern in (32), but it would require further assumptions for extraction from *dass* clauses into *dass* clauses in contexts where there is an auxiliary in situ in the matrix clause, as in (33) (compare (1-a) = (20)). The reason is that it would be unclear why a probe feature on the lower V head (i.e., the main verb) could render an otherwise inactive higher V head (i.e., the auxiliary) active.

(33) (Ich weiß nicht) $[_{CP_1}$ wen$_4$ (dass) du $[_{VP}$ t$_4''$ gemeint hast $[_{CP_2}$ t$_4'$ dass sie t$_4$
       I   know not   whom that you   thought have   that she
       getroffen hat]]]
       met   has

I will therefore pursue a slightly different approach here and postulate that an auxiliary (or modal) verb and its associate main verb are base-generated as a single complex head, and reprojection movement of the finite part of the complex head proceeds via excorporation, much as with particle stranding under verb-second.

7.2   Extraposition

Second, throughout this chapter I have presupposed that complement clauses are merged in the canonical (left-peripheral) object position in German. However, finite complement clauses typically undergo extraposition; in fact, extraposition is obligatory in the case of verb-second complements (see Stechow & Sternefeld (1988), Heck (2010)), and also if extraction from the clause has applied (arguably a syntactic reflex of successive cyclicity in German; see Müller (1999), Lahne (2009)). On the view adopted here, obligatory extraposition is unrelated to the status of a clausal complement as transparent or opaque for extraction; and one might argue that the transparency of *dass* clauses as in (1-a), (1-b) supports this view.[27]

---

[26] Closer inspection reveals that this problem shows up in a number of analyses of the pattern in (1), among them the one pursued by Staudacher (1990). In his system, one would have to assume that the auxiliary and the main verb are always co-indexed.

[27] Heck (2010) assumes that verb-second clauses are barriers because they are obligatorily extraposed; but this then requires a heterogeneous approach to verb-second clause and *dass* clause extraposition (i.e., two extraposition rules with different properties) so as to ensure that extraposed *dass* clauses can still be transparent.

7.3   Bridge Verbs

Third, it can be observed that the present proposal is incompatible with one small piece of the analysis of CED effects I give in chapter 4: Recall that I suggested there that the bridge verb/non-bridge verb distinction, as well as lexical variation with extraction from DP, can be tied to the presence or absence of an additional probe feature that may keep V active and accessible for edge feature insertion; cf. (34) (lexically determined extraction from CP) and (35) (lexically determined extraction from DP).

(34) a. (Ich weiß nicht) [$_{CP_1}$ wen$_4$ (dass) du [$_{VP}$ t_4'' meinst/glaubst [$_{CP_2}$ t_4' dass
          I   know not   whom that you   think   that
          du   t$_4$ getroffen hast]]]
          you met   have

    b. ?*(Ich weiß nicht) [$_{CP_1}$ wen$_4$ (dass) du [$_{VP}$ t_4'' bereust/weißt [$_{CP_2}$ t_4' dass
          I   know not   whom that you   regret/know   that
          du   t$_4$ getroffen hast]]]
          you met   have

(35) a. (Ich weiß nicht) [$_{PP_4}$ worüber] der Fritz   [$_{VP}$ t_4' [$_{DP}$ ein Buch   t$_4$]
          I   know not   about what the Fritz$_{nom}$ a   book$_{acc}$
          liest]
          reads

    b. ?*(Ich weiß nicht) [$_{PP_4}$ worüber] der Fritz   [$_{VP}$ t_4' [$_{DP}$ ein Buch   t$_4$]
          I   know not   about what the Fritz$_{nom}$ a   book$_{acc}$
          klaut]
          steals

According to the approach in chapter 4, V may not have the special probe feature that is needed to keep the head active and permit edge feature insertion for extraction from the (CP or DP) complement in (34-b) and (35-b); i.e., on this view, t_4'' in (34-b) and t_4' in (35-b) cannot be present, which gives rise to a PIC vioalation once the derivation moves on. As alluded to in footnote 21, this analysis cannot be maintained anymore under present assumptions. The reason is that we would then expect verb-second movement to improve the examples in the same way that verb-second movement in the matrix makes extraction from an embedded verb-second clause possible. This would clearly not be a correct prediction; see (36-ab).

(36) a. ?*[$_{CP_1}$ Wen$_4$ bereust/weißt du [$_{VP}$ t_4'' [$_{CP_2}$ t_4' dass du t$_4$ getroffen hast]]] ?
          whom regret/know   you   that you   met   have

    b. ?*[$_{PP_4}$ Worüber] klaut der Fritz   [$_{VP}$ t_4' [$_{DP}$ ein Buch   t$_4$]] ?
          about what steals the Fritz$_{nom}$ a   book$_{acc}$

In order to exclude sentences like those in (34-b), (35-b) and (36) as a CED effect reducible to the PIC (via the impossibility of edge feature insertion), an additional

assumption seems thus necessary. Irrespective of whether or not a categorial probe feature identifying the complement can be inserted on V in all these cases, there must be an independent difference (and this might then support the hypothesis that a categorial probe feature can always show up with complements, for reasons of uniformity alone). I cannot possibly develop a full-fledged account at this point. However, it seems that a simple hypothesis would do the job: Suppose that in the contexts where extraction from a last-merged complement is impossible even in the presence of verb-second, there are additional *shells* on top of the respective (CP or DP) categories; and suppose further that the heads of these shells (a) do *not* take specifiers and (b) do *not* carry any probe features (categorial or other); see (37) (where ZP is the shell intervening between V and CP/DP).

(37) [$_{VP}$ [$_{ZP}$ [$_{Z'}$ Z $\{C/D\}$P]] V]

This will lead to a failure of edge feature insertion for an item in the edge domain of CP or DP, and subsequent extraction from CP or DP to the edge domain of V will create a PIC violation. Incidentally, such a shell-based approach has been developed for the bridge/non-bridge distinction in Müller (1995) (based on Kiparsky & Kiparsky (1970) and others); whether it might be tenable for variation in the domain of extraction from DP remains to be seen.[28]

---

[28] Thus, adding more structure may produce CED effects that are reducible to the PIC which would otherwise not be present. Does this imply an incompatibility with the seemingly contradictory statement in the appendix of chapter 4, to the effect that more structure may help to circumvent CED effects that are reducible to the PIC? The answer is no: Adding structure *within* a phrase – as in the approach to the bridge phenomenon and variation with extraction from DP envisaged in the appendix of chapter 4 – may undo a CED effect that would otherwise occur (melting, pseudo-melting). In contrast, adding structure *outside* of a phrase – as in the approach to the bridge phenomenon and, possibly, variation with extraction from DP just sketched – may create a CED effect.

Chapter 7

Island Repair by Ellipsis

1. Introduction

Like the previous chapter, the final chapter of this monograph is concerned with the initially surprising absence of CED-type effects in a certain domain. More specifically, I want to show that the approach to CED effects in chapter 4 offers a new perspective on an old problem: the lack of CED effects under ellipsis in sluicing constructions, where deletion takes place. The line of argumentation will be similar to the one given in chapter 6: Recall that according to the approach developed in chapter 4, a CED effect arises with movement from XP if the next higher phase head Y is not active anymore, i.e., if it does not bear an operation-inducing feature at the relevant stage of the derivation. In sluicing constructions, a *wh*-phrase undergoes movement to a SpecC$_{[wh]}$ position, which is accompanied by deletion of all material to the right of the *wh*-phrase; interestingly, island effects that normally show up with movement can be avoided in this construction type. This phenomenon ultimately remains a mystery under many existing accounts, and where it does not, it is usually taken to require specific assumptions, and a representational approach. In contrast, I will argue that the effect follows without further ado from an independently motivated approach to deletion, given the appraoch to CED effects in chapter 4: There are features on phase heads that trigger deletion in ellipsis contexts, and these features keep the head active at the relevant stage of the derivation, thereby making a circumvention of island effects possible. Thus, non-occurrence of CED effects with extraction from a verb-second complement in the presence of verb-second movement in the matrix clause (as it was discussed in chapter 6) and non-occurrence of CED effects in sluicing contexts (as it will be discussed below) exhibit a similar pattern: What unites the two contexts is the presence of an operation-inducing feature on a phase head *after* the phase head has discharged its final structure-building feature (which would normally be responsible for the CED effect): a Münchhausen probe feature that triggers verb-second by reprojection movement in the first case; and a feature that triggers deletion of complements of phase heads in ellipsis contexts. These features keep the head active at the

relevant stage of the derivation.[1]

2. The Problem

2.1 Data

Run-of-the-mill examples involving sluicing in English and German are given in (1-ab), respectively (see Ross (1969), Merchant (2001)).[2]

(1) a. John bought something, but I don't know what$_1$ ~~John bought t$_1$~~.
    b. Fritz hat irgendwas falsch gemacht, aber mir   ist nicht klar, was$_1$ ~~Fritz t$_1$~~
        Fritz has something wrong done   but me$_{dat}$ is not   clear what Fritz
        ~~falsch gemacht hat~~
        wrong done   has

A standard analysis of the phenomenon that I will presuppose in what follows is that sluicing is to be analyzed as TP deletion triggered by a (certain kind of) interrogative C, and subject to recoverability and a parallelism constraint (see, e.g., Lasnik (1999), Merchant (2001), and Heck & Müller (2003)).

There is one generalization that is important in the present context, and that is that slucing may repair (or, more neutrally, circumvent) island effects (see Ross (1969), Chomsky (1972), Chung et al. (1995), Merchant (2001), Fox & Lasnik (2003), Bošković (2011), among many others).[3] In what follows, this is shown for a number of island contexts, often with examples that go back to Ross (1969) and that are by now legendary. Thus, the example pair in (2) illustrates that sluicing in English can circumvent Sentential Subject Constraint effects (cf. (69) from chapter 1).

(2) a. That he'll hire someone is possible, but I won't divulge who$_1$ [$_{TP}$ [$_{CP}$ ~~that~~
        ~~he'll hire t$_1$~~] ~~is possible~~]
    b. *That he'll hire someone is possible, but I won't divulge who$_1$ [$_{TP}$ [$_{CP}$ that
        he'll hire t$_1$] is possible]

Next, Subject Condition effects with DPs (see (72) from chapter 1) are also absent in sluicing constructions; compare the German examples with and without sluicing in (3).[4]

---

[1] Note that the argument is completely parallel to the one given in Assmann (2010) for the absence of certain island effects in parasitic gap constructions, which instantiates a third kind of a priori unexpected non-occurrence of CED effects.

[2] Here and in what follows, deletion is indicated by crossing out the relevant material.

[3] In this respect, sluicing differs from another syntactic deletion operation, viz., VP ellipsis, which cannot repair island effects.

[4] English examples showing the same pattern are given in (i) (from Merchant (2001).

(3) a. Es hat wohl ein Buch   über irgendwen den Karl   beeindruckt, aber
      it has PRT a   book$_{nom}$ about someone   the Karl$_{acc}$ impressed   but
      ich weiß nicht, über wen$_1$ ~~wohl~~ [$_{DP}$ ~~ein Buch~~   ~~t$_1$~~] ~~den Karl~~
      I   know not   about whom PRT   a   book$_{nom}$ the Karl$_{acc}$
      ~~beeindruckt~~ ~~hat~~
      impressed   has

   b. *Es hat wohl ein Buch   über irgendwen den Karl   beeindruckt, aber
       it has PRT a   book$_{nom}$ about someone   the Karl$_{acc}$ impressed   but
       ich weiß nicht, über wen$_1$ wohl [$_{DP}$ ein Buch t$_1$] den Karl
       I   know not   about whom PRT   a   book$_{nom}$ the Karl$_{acc}$
       beeindruckt hat
       impressed   has

The same pattern arises with the Adjunct Condition (see (94) from chapter 1): Regular movement as in (4-b) in English yields strong illformedness; but the example becomes acceptable if sluicing takes place, as in (4-a).

(4) a. Ben left the party because one of the guests insulted him, but he wouldn't
      tell me [$_{DP_1}$ which guest] ~~he left the party because t$_1$ insulted him~~

   b. *Ben left the party because one of the guests insulted him, but he wouldn't
       tell me [$_{DP_1}$ which guest] he left the party because t$_1$ insulted him

Island repair by deletion in sluicing constructions is shown for the relative clause part of the Complex NP Constraint (see (65) from chapter 1) on the basis of English data such as those in (5-a) and (5-b). The example pair in (5-b) is particularly instructive because it shows that a violation of the Left Branch Condition (according to which the leftmost item of a DP cannot be moved out of that DP; see Ross (1967)) is also possible with sluicing.

(5) a. (i) She kissed a man who bit one of my friends, but Tom doesn't realize
            [$_{DP_1}$ which one of my friends] ~~she kissed a man who bit t$_1$~~

      (ii) *She kissed a man who bit one of my friends, but Tom doesn't realize
            [$_{DP_1}$ which one of my friends] she kissed a man who bit t$_1$

   b. (i) They want to hire someone who speaks a Balkan language, but I don't
            remember which$_1$ ~~they want to hire someone who speaks~~ [$_{DP}$ ~~t$_1$ Balkan~~
            ~~language~~]

      (ii) *They want to hire someone who speaks a Balkan language, but I don't
            remember which$_1$ they want to hire someone who speaks [$_{DP}$ t$_1$ Balkan

---

(i) a. *Which Marx brother$_1$ did she say that [$_{DP_2}$ a biographer of t$_1$] interviewed her ?

    b. A biographer of one of the Marx brothers interviewed her, but I don't remember which ~~Marx~~
        ~~brother$_1$ a biographer of t$_1$ interviewed her~~

language]

The same asymmetry arises with the noun complement part of the Complex NP Constraint; see (6-a) vs. (6-b).

(6) a. I believe the claim that he bit someone, but they don't know who$_1$ ~~I believe the claim that he bit t$_1$~~

     b. *I believe the claim that he bit someone, but they don't know who$_1$ I believe the claim that he bit t$_1$

Freezing effects (cf. (36) in section 4 of chapter 4) also do not show up under sluicing; see the following pair of examples taken from Merchant (2001).

(7) a. *Which Marx brother$_1$ did she say that [$_{DP_2}$ a biography of t$_1$] she refused to read t$_2$?

     b. A: [$_{DP_2}$ A biography of one of the Marx brothers] she refused to read t$_2$.
         B: Which one$_1$ ~~(did)~~ [$_{DP_2}$ ~~a biography of t$_1$~~] ~~she refuse to read t$_2$~~ ?

Finally, consider the example pair in (8). (8-a) shows that *wh*-movement in English obeys the Coordinate Structure Constraint (according to which no conjunct may be moved in a coordinate structure, nor may any element contained in a conjunct be moved out of that conjunct; see Ross (1967)); and (8-b) shows that sluicing lifts the ban on moving an item out of a coordinate structure.

(8) a. Irv and someone were dancing together, but I don't know who$_1$ ~~Irv and t$_1$ were dancing together~~

     b. *Irv and someone were dancing together, but I don't know who$_1$ Irv and t$_1$ were dancing together

As mentioned in footnote 3, VP ellipsis behaves differently from sluicing (i.e., TP ellipsis) in that it does not give rise to island repair of any kind. This is shown for Adjunct Condition effects by the lack of contrast in (9) (see Chung et al. (1995)).

(9) a. *What$_1$ did you leave [$_{CP}$ before they started playing t$_1$] ?
          (We left before they started playing party games.)

     b. *What$_1$ did you leave before they did [$_{VP}$ ~~start playing t$_1$~~] ?

2.2 A Standard Analysis

The classical approach to island repair by ellipsis in sluicing constructions is due to Chomsky (1972). In brief, the analysis works as follows. First, when movement crosses an island, a diacritic signalling ungrammaticality (like #) is assigned to the island. Second, an output filter blocks structures in which # shows up. And third, if deletion applies, # is removed together with the lexical material of the clause. This is why sluicing can circumvent island effects. Could such an approach be compati-

ble with the approach to islands and movement developed in this monograph? The answer would seem to be yes if one assumes that violations of the PIC do not necessarily lead to a breakdown of the derivation but may only result in the assignment of # to the phrase structure, which may then eventually be deleted in sluicing constructions.[5] Then again, such an approach would be inherently representational in that it would depend on inspecting full phrase structures at the end of the derivation in order to detect diacritics which are possibly very deeply embedded. Moreover, it would be stipulative because of the very presence of the diacritic # that is assigned after an island violation.[6] These are good reasons to investigate whether some other, derivational approach to island repair by ellipsis is available, one that fits into the overall system of assumptions developed in the preceding chapters of this monograph. The key to such an approach is the concept of cyclic, phase-based deletion, to which I will now turn.[7]

3.   Deletion by Phase

As a starting point, I would like to point out that there are two widely held assumptions about spell-out and deletion that may easily be construed as mutually incompatible. These assumptions are given in (10).

(10) a.   Sluicing is TP deletion.
       b.   Spell-out is cyclic and proceeds phase-by-phase.

The first assumption has already been introduced above (cf. Lasnik (1999), Merchant (2001), Heck & Müller (2003)). On this view, there is only one deletion operation in sluicing contexts: A TP that is the complement of an interrogative C is deleted, and with it all the material that it contains. Note that this deletion operation must be syntactically visible because it interacts with syntactic operations like multiple *wh*-movement in multiple sluicing constructions in German (see Heck & Müller (2003)

---

[5] This option is of course not available if the PIC is associated with (ultimately, reduced to) a non-metaphorical concept of cyclic spell-out of complements of phase heads; recall footnote 5 in chapter 2, and see below.

[6] Also compare the use of the feature $[\pm\gamma]$ for the purposes of the ECP; see the discussion of (5) in chapter 1; and see Merchant (2004) on the similarity of the two concepts. In addition, see Kitahara (1999) for an attempt to get rid of these kinds of features that only serve the purpose of flagging earlier events for later reference (led by considerations related to the Inclusiveness Condition).

[7] Needless to say, many other approaches to island repair by ellipsis have been developed since Chomsky's (1972) original analysis. I cannot possibly list all proposals here, let alone go through them one by one. Let me just mention Merchant's (2001) influential idea that for at least some types of island effects, the non-occurrence of the island effect in sluicing (i.e., PF deletion) constructions may be explained by the hypothesis that the island effect in question is *itself* a PF phenomenon.

and references cited there). Thus, consider the two examples in (11), which instantiate a legitimate and an illegitimate case of multiple wh-movement in German sluicing constructions, respectively.

(11) a. Irgend jemand hat irgend etwas geerbt   aber der Fritz weiß   nicht
          someone   has something   inherited but ART Fritz knows not
          mehr [$_{CP}$ wer$_1$ was$_2$ [$_{TP}$ t̶₁̶ t̶₂̶ g̶e̶e̶r̶b̶t̶   h̶a̶t̶]]
          more   who what   inherited has

      b. *Irgend jemand hat behauptet dass die   Maria irgend etwas geerbt   hat
          someone   has claimed   that ART Maria something   inherited has
          aber der   Fritz weiß   nicht mehr [$_{CP}$ wer$_1$ was$_2$ [$_{TP}$ t̶₁̶ b̶e̶h̶a̶u̶p̶t̶e̶t̶ h̶a̶t̶ [$_{CP}$
          but ART Fritz knows not   more   who what   claimed has
          d̶a̶s̶s̶ [$_{TP}$ d̶i̶e̶   M̶a̶r̶i̶a̶ t̶₂̶ g̶e̶e̶r̶b̶t̶   h̶a̶t̶]]]]
          that   ART Maria   inherited has

As argued in Heck & Müller (2003), both *wh*-phrases in (11-a) have indeed undergone overt *wh*-movement to an embedded SpecC position. Such multiple *wh*-movement is exceptionally possible in German (which otherwise only tolerates one *wh*-phrase per SpecC position) because it is the only way to fulfill the general recoverability requirement in (11-a) – if the second *wh*-phrase (whichever one that is; cf. chapter 3) stays in situ, important information will fatally be lost, leading to ungrammaticality (or uninterpretability). However, as shown by the illformedness of (11-b), multiple *wh*-movement in sluicing constructions in German is not always possible; if one of the *wh*-phrases has to undergo long-distance movement from a CP that is further embedded, ungrammaticality arises. Independently of which analysis of this phenomenon is ultimately adopted, the contrast in (11) shows that TP deletion in sluicing constructions must be syntactically visible – it clearly interacts with syntactic movement.[8]

Now let us turn to the second assumption in (10), according to which spell-out applies cyclically; see Uriagereka (1999) for the original idea, and Chomsky (2001b; 2008) and Epstein & Seely (2002) for two further developments of this assumption.[9] Chomsky's phase-based approach to cyclic spell-out is arguably the most widely adopted approach among those that incorporate the general concept. Here it is as-

---

[8] On the analysis in Heck & Müller (2003), (11-b) qualifies as an instance of *counter-feeding* (see Chomsky (1975), Kiparsky (1973)): The ultimate representation is opaque; viewed representationally, it looks as though multiple *wh*-movement should be able to apply to make the multiple question reading possible under sluicing (in analogy to (11-a), which shows that recoverability can in principle force multiple *wh*-movement in German). However, the trigger for multiple *wh*-movement (viz., sluicing, i.e., deletion of TP, and the recoverability problem it creates) comes too late in the derivation; when it is there, the chance for multiple sluicing to apply has passed (because the in-situ *wh*-phrase is now too deeply embedded).

[9] Also recall from chapter 2 that Uriagereka's analysis is radically incompatible with the phase-based approach proposed by Chomsky.

sumed that when a phase head has carried out all the operations that it can trigger and the phase is completed, its complement is sent to PF and thereby rendered inaccessible for any further operations in the derivation.

The dilemma that arises if both these assumptions are taken seriously is the following: If cyclic spell-out proceeds phase-by-phase and sluicing is a deletion operation applying to syntactic structures, the latter operation will only be able to affect that part of the structure created by the derivation so far which has not yet undergone spell-out. In other words: If some part of a structure has already been affected by regular (cyclic) spell-out, and thus been sent off to PF, subsequent deletion of the whole structure will only be able to affect what is left. More specifically, if a subpart of a TP that is to be deleted in a sluicing construction has already been sent off to PF via cyclic spell-out of phase head complements, subsequent TP deletion cannot affect all material that was originally present in TP. This will give rise to selective (and sometimes non-constituent) deletion effects which do, however, not seem to be attested in sluicing constructions. Based on the standard assumption that vP and CP are phases (which I have abandoned in the present work), combining cyclic spell-out and sluicing as TP deletion would give rise to sentences like those in (12). In (12-a), we should expect only the subject pronoun to be successfully deleted by sluicing since *ein Buch bringt* is the complement of the next-lower phase head (v), and already PF-realized when TP deletion applies. In (12-b), the non-constituent sequence *Karls*-D would undergo deletion, but the rest of the the embedded clause (i.e., *Katze, von uns mag*) would be immune to deletion because it has already been spelled out as the complement of a phase head (v again), and is thus PF-realized when TP deletion applies.[10]

(12) a. *Sie bringt irgendwem ein Buch, aber ich weiß nicht, wem$_1$ [$_{TP}$ [$_{vP}$
     she brings someone   a   book but   I   know not   whom$_{acc}$
     ~~sie~~   [$_{v'}$ [$_{VP}$ t$_1$ ein Buch   [$_{V}$ bringt]] v]]]
     she$_{nom}$ a   book$_{acc}$ brings

    b. *Karls Katze mag einen von uns, aber ich sage nicht, wen$_1$ [$_{TP}$ [$_{vP}$ [$_{DP}$
     Karl's cat   likes one   of   us   but I   say   not   whom
     ~~Karls~~ ~~D~~ [$_{NP}$ Katze]] [$_{v'}$ [$_{VP}$ [$_{DP}$ t$_1$ von uns] mag] v]]]]
     Karl's   cat   of   us   likes

It is clear that these problems can only get worse (especially in other deletion constructions) if every phrase is a phase, as assumed throughout this monograph. The solution that I would like to suggest in view of this state of affairs is the following:

---

[10] I assume here that the main verb V stays in situ. If it undergoes movement to v (notwithstanding what has been said in chapter 6), it will also be affected by phase-based deletion, producing non-constituent deletion throughout.

Deletion must be successive-cyclic, proceeding phase by phase.[11] As a matter of fact, this looks like null hypothesis from a conceptual perspective: On this view, deletion is just a special form of PF realization, and there is no reason to assume that it proceeds differently from other ('regular') instances of PF realization. So, if regular PF realization applies phase-by-phase, deletion should do so, too.[12]

Here is a proposal that implements this assumption. Suppose that deletion, as a special kind of cyclic spell-out, is brought about by a designated operation-inducing feature [†]. A feature [†] on a phase head triggers a certain kind of spell-out of the complement, viz., zero realization. In sluicing (and other deletion) contexts, there must then be an uninterrupted sequence of [†] features on all heads dominated by the ultimate deletion site TP (except for those heads that do not take complements, i.e., that qualify as phrases), ending with the interrogative C that takes TP as a complement. (The sequence of [†]s must not be interrupted because every phrase is a phase.) Of course, such a spreading of [†] features must be effected in some way. For the moment, I will abstract away from this issue, but I will come back to it at the end of the chapter. Let us for now just take the presence of [†] features on all phase heads within a TP deletion site for granted, and see what this predicts for the occurrence of CED effects.

4.   A PIC-Based Approach

4.1   Analysis

Spell-out of a complement is the very last operation that takes place in a phase; after this operation, the derivation moves into the next phase cycle, merging a further phase head with the current phase. Given the proposal made in the previous section, there are two kinds of complement spell-out. On the one hand, there is regular spell-out that leads to a phonological realization of the lexical material contained in the structure that is sent off to PF; on the other hand, there is deletion. The latter operation is triggered by a feature on a phase head: [†]. Crucially, as long as a phase head is equipped with an operation-inducing feature [†] for complement deletion, it is active. Technically, this implies that the Edge Feature Condition (see (11) in chapter 6) must be enriched as in (13) (as before, to simplify matters I leave out the additional requirement that the edge feature can only be inserted if this is the only way to produce a balanced phase; see chapter 3); it now recognizes three kinds of operation-inducing features which are

---

[11] A qualification is in order here: I have nothing to say about other deletion operations like gapping and pseudo-gapping, which may require a very different analysis.

[12] There may be alternatives to this conclusion. For instance, one might assume that TP deletion may affect all pieces of structure that have so far undergone regular cyclic spell-out after all, by *reversing*, in the phonological component, earlier decisions concerning the type of spell-out as they result from regular cyclic spell-out. However, this looks like a clear deviation from the null hypothesis.

located on three separate feature stacks.

(13) *Edge Feature Condition* (EFC):
      An edge feature [•X•] can be assigned to the head γ of a phase only if (a) and (b) hold:

      a.   γ has not yet discharged all its structure-building (•), probe (∗) or deletion (†) features.

      b.   [•X•] ends up on top of γ's list of structure-building features.

Since a [†] feature on a phase head is necessarily discharged as the final operation in the phase, it keeps the phase head active *after* a last-merged specifier (or complement) has been introduced into the structure. Therefore, a phase head that is part of a deletion structure permits edge feature insertion for some item in the edge domain of a last-merged phrase, and this item is thus accessible on the next cycle without violating the PIC. This, in a nutshell, accounts for the lack of CED effects in sluicing constructions. Thus, [†] plays a role that is completely analogous to the roles played by expletives in pseudo-melting (chapter 4), and by Münchhausen probe features on matrix verbs in extraction from verb-second clauses (chapter 6).

   As a case study of how CED effects can be circumvented with the help of [†] features in sluicing constructions, consider Subject Condition effects and their a priori unexpected absence in sluicing constructions (see (2), (3); and (i) of footnote 4). Focussing on the German examples in (3), two cases can be distinguished: In the first case, there is movement from a DP that is last-merged in vP, and there is no deletion of v's complement VP. In the second case, there is movement from a DP that is last-merged in vP, but this time deletion of v's complement VP must take place (as part of a sluicing construction). I address these two cases in turn.

4.2   Extraction from Specv without VP Deletion

The ill-formed example in (3-b) is repeated here as (14).

(14) *Es hat wohl ein Buch   über irgendwen den Karl   beeindruckt, aber ich
      it has PRT a   book$_{nom}$ about someone the Karl$_{acc}$ impressed   but I
      weiß nicht, über wen$_1$ wohl [$_{DP}$ ein Buch t$_1$] den Karl   beeindruckt hat
      know not   about whom PRT   a   book$_{nom}$ the Karl$_{acc}$ impressed has

This is just a regular CED effect as it has been extensively discussed in chapter 4. The subject DP that is last-merged in vP is a barrier because, at the point where DP has been merged in vP, no operation-inducing feature can be left on v. This is so because (i) v has discharged its last structure-building feature by definition; (ii) v cannot retain any probe features (e.g., case and φ-features, but the same conclusion can be drawn for any other kind of potential probe feature on v, including perhaps features encoding status-government, as envisaged on page 182 in chapter 4) after the subject DP has

been merged because these features could then not be discharged anymore, due to a conspiracy of the Strict Cycle Condition (which precludes Agree with an item in the complement VP) and the c-command requirement on Agree (which blocks Agree with the subject DP itself); and (iii) there is no feature triggering deletion present on v that might keep v active. The changes in feature composition on the phase head v at the crucial stages of the derivation are shown in (15).

(15) *Specifiers of v as barriers, no* [†] *on v*:

$$
\begin{array}{ll}
& \text{v: } [\bullet V\bullet] \succ [\bullet D\bullet] \\
& \quad [*\text{acc}*], [*\phi*] \\
\rightarrow & \text{v: } [\bullet D\bullet] \\
& \quad [*\text{acc}*], [*\phi*] \\
\rightarrow & \text{v: } [\bullet D\bullet] \\
& \quad [*\phi*] \\
\rightarrow & \text{v: } [\bullet D\bullet] \\
& \text{v: } \text{---}
\end{array}
$$

$\rightsquigarrow [\bullet X\bullet]$ insertion impossible

Since v is therefore inactive at the stage of the derivation when its DP specifier has come into being, no edge feature can be inserted on v that would permit extraction of some item from the subject DP; consequently, the PIC will have to be violated once the derivation moves beyond the vP domain and tries to attract the *wh*-PP$_1$ to a higher position (i.e., SpecT).

4.3 Extraction from Specv in Sluicing Constructions

Now consider the well-formed example involving extraction from a subject DP in a sluicing construction given in (3-a); the example is repeated here as (16).

(16) Es hat wohl ein Buch   über irgendwen den Karl   beeindruckt, aber ich
     it has PRT a   book$_{nom}$ about someone   the Karl$_{acc}$ impressed   but I
     weiß nicht, über wen$_1$ ~~wohl~~ [$_{DP}$ ~~ein Buch~~   t$_1$] ~~den Karl~~   ~~beeindruckt~~
     know not   about whom PRT   a book$_{nom}$ the Karl$_{acc}$ impressed
     ~~hat~~
     has

There is one minimal difference to (14); this time, v bears a [†] feature because it will eventually be part of a bigger TP deletion structure. This feature keeps the phase head active after the subject DP has been merged. Consequently, an edge feature can be instantiated on v at this point, and movement can take place out of the subject to a further (higher) Specv position. This way, a PIC violation can be avoided when the derivation moves beyond the vP domain. (17) shows how the stacks of operation-inducing features on v change in the relevant part of the derivation; in particular, it is shown here that an edge feature can be inserted even though both structure-building

and probe features are absent when this needs to be done.

(17) *Specifiers of v as non-barriers,* [†] *on v*:

$$
\begin{array}{l}
\boxed{
\begin{array}{ll}
\text{v:} & [\bullet V\bullet] \succ [\bullet D\bullet] \\
& [\ast \text{acc}\ast], [\ast \phi \ast] \\
& [\dagger] \\
\rightarrow \text{v:} & [\bullet D\bullet] \\
& [\ast \text{acc}\ast], [\ast \phi \ast] \\
& [\dagger] \\
\rightarrow \text{v:} & [\bullet D\bullet] \\
& [\ast \phi \ast] \\
& [\dagger] \\
\rightarrow \text{v:} & [\bullet D\bullet] \\
& [\dagger] \\
\rightarrow \text{v:} & [\dagger] \\
\rightarrow \text{v:} & [\bullet X\bullet] \\
& [\dagger] \\
\rightarrow \text{v:} & [\dagger] \\
\rightarrow \text{v:} & \text{---}
\end{array}
}
\end{array}
\qquad \rightsquigarrow \text{violates nothing}
$$

4.4 [†] Discharge and the Strict Cycle Condition

At this point, a question needs to be addressed that arises for derivations incorporating a sequence of operations like (17) with respect to the Strict Cycle Condition; this constraint is repeated once more in (18) (= (14) in chapter 6).

(18) *Strict Cycle Condition* (SCC):
Within the current domain ζ, a binary syntactic operation may not exclusively apply to positions ϵ, τ if ϵ and τ are both included within another domain π that is dominated by ζ.

Discharge of [†] seems to violate the SCC: The derivation has reached the vP level, but [†] discharge applies only to v (which bears the feature) and VP (which is the target of the operation, thus clearly making [†] discharge a *binary* operation in the sense of (18)). However, v and VP are included in v′, which in turn is dominated by vP; so the operation would seem to be counter-cyclic. Thus, if this option is available with [†] features, there is a fundamental difference to probe features: Recall that the option of counter-cyclic feature-discharge must not be available with probe features; otherwise the barrier status of (last-merged) specifiers could not be derived.

Still, there is good evidence that such an asymmetry between probe features and [†] features triggering deletion is not an artefact of the theory. Complement spell-out violates strict cyclicity virtually by definition: The complement of a phase head X

is spelled out within the local domain X′ *after* the more inclusive XP has been completed; and [†]-driven complement deletion is just complement spell-out of a certain type. Thus, we may conclude that SCC violations are unavoidable under a cyclic spell-out approach, and therefore unproblematic for [†]-feature discharge.[13]

5.   Some Further Issues

The account of the absence of island effects just given raises a number of further questions. I will confine myself to briefly discussing three of them, concerning (i) the generality of the phenomenon, (ii) the consequences for VP ellipsis, and (iii) the spreading of [†] features through syntactic structures.

5.1   Island Types

First, notwithstanding Merchant's (2001) arguments that absence of islands effects with sluicing should not be treated as a homogeneous phenomenon (with clausal islands differing from non-clausal ones, and only the latter motivating the hypothesis that islands can be a PF phenonomenon; also see Abels (2011) for an even more radical view), I assume that a homogeneous approach to the non-occurrence of island effects with sluicing should ideally be given. The approach developed here works for all legitimate island violations with sluicing where the island can be traced back to a CED (or barrier) effect derivable from the PIC, like the Sentential Subject Constraint, the Subject Condition, the Adjunct Condition, the Complex Noun Phrase Constraint (with both the relative clause and complement clause subcases), and the Freezing Generalization. At present, I take it to be an open question whether the approach can be generalized to other legitimate island violations, like the lack of Coordinate Structure Constraint and Left Branch Condition effects with sluicing. The simplest approach would seem to suggest PIC-based approaches to these phenomena, too; but I will not make any attempt at deriving Coordinate Structure Constraint and Left Branch Condition effects from the PIC (or at least from the inactivity of a phase head at the relevant stage of the derivation) in this monograph.[14]

5.2   VP Ellipsis

Second, the question arises of why VP ellipsis, unlike TP ellipsis, cannot circumvent islands; see (9), repeated here in (19).

---

[13] Strictly speaking, this means that either the Strict Cycle Condition in (18) would have to be altered so as to reflect this, or that cyclic spell-out (including [†]-feature discharge) would have to be classified as not purely a "binary syntactic operation". But since both options are straightworward, I refrain from doing so here.

[14] As far as I can see, such an attempt would not have to be classified as futile from the start, but it would need to involve highly complex reasoning, and abstract structures.

(19) a. *What₁ did you leave [_CP before they started playing t₁] ?
          (We left before they started playing party games.)
      b. *What₁ did you leave before they did [_VP ~~start playing t₁~~] ?

A simple analysis that covers the examples of sluicing and VP ellipsis mentioned so
far on the basis of the standard approach is discussed in Merchant (2001) and Fox &
Lasnik (2003): In (19-b), the island that is crossed by movement, and consequently
assigned the diacritic #, shows up outside of the deletion site; therefore, repair is im-
possible. In contrast, in all the sluicing examples involving island repair discussed so
far, the island is within the deletion site; therefore, a # that shows up here disappears
once deletion takes place. The very same consequence would in fact also directly
follow under present assumptions: The head *before* of the adjunct CP in (19-b) does
not bear a [†] feature. Therefore, a CED effect that is reducible to the PIC cannot
be avoided: Deletion in lower parts of the structure is irrelevant. However, Merchant
(2001) and Fox & Lasnik (2003) show that this kind of analysis can ultimately not be
the whole story because it turns out that there are cases where VP ellipsis fails to cir-
cumvent an island even though the island is within the deletion site. The two accounts
given by Merchant (2001) and Fox & Lasnik (2003) in view of this are not compatible
with the present approach. (For instance, Fox and Lasnik's analysis necessitates the
assumption that one-step long-distance movement is in principle an option.) At the
moment, I cannot offer a full-fledged solution to this problem. An assumption that
might work from a purely technical point of view (even though it is not substantiated
by independent syntactic evidence) might be that different kinds of deletion-inducing
features may be involved in the two kinds of constructions; this then may or may not
be related to the fact that different syntactic licensing requirements (concerning, e.g.,
lexicalization requirements for the head adjacent to the deletion site) are clearly im-
posed on the two deletion operations (in English at least). Assuming that there are
two types of deletion features, one might simply stipulate that one type is in effect too
weak to keep the head active at the decisive stage of the derivation, thereby making a
circumvention of island effects impossible in the case of VP ellipsis even if the island
itself also undergoes deletion.

5.3 *Spreading of [†]*

Third and finally, something must be said about the required spreading of [†] features
through all head positions (except for those that are non-complex phrases) up to the
ultimate TP deletion site in sluicing constructions. There are various standard tech-
niques of passing on non-local information locally in a bottom-up fashion that have
been suggested in strictly derivational approaches that strive to do without look-ahead
(involving, i.a., feature movement to head or specifier positions, or checking or selec-
tion mechanisms, as they have been proposed for binding (see Fischer (2006)), or for

agreement (as in the concept of cyclic Agree, see Legate (2005), Frank (2006), Lahne (2008)). However, these kinds of mechanisms do not seem applicable in the case at hand. They can propagate features along a projection line, thereby creating a path, but they cannot easily capture the intended top-down effect of spreading from the sister of the ultimate deletion source (i.e., the $C_{[wh]}$ head that is eventually the trigger of sluicing) to (nearly) all the embedded phase heads (again, except for heads that are trivial phrases) of the whole subtree under the deleted TP. There are two reasons for this failure: First, the intended relation is not one-to-one (as with standard Agree or movement operations) but one-to-many. And second, as noted, at least at first sight it looks as though the spreading could best be described by a top-down algorithm. To see this, consider the abstract representation in (20). In the cyclic, phase-by-phase approach to deletion via repeated complement spell-out as zero, [†] features must show up on the phase heads Z, X, and Y (but not on J, H, R, and U, assuming that these phase heads are trivial XPs – JP, HP, RP, and UP).

(20)
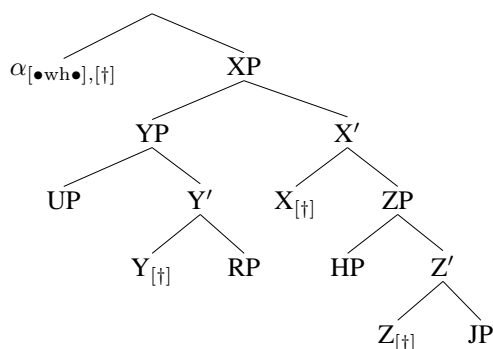
In view of this situation, it seems to me that the best way to proceed is not to try to assimilate [†]-induced operations to well-established syntactic operations (e.g., by adopting a less restrictive approach to Agree operations that might permit multiple Agree (see Hiraiwa (2005)) or feature sharing (see Pesetsky & Torrego (2007), Bhatt (2005), Heck & Cuartero (2011), among many others)); rather, I take it that one should dispense with the idea that cyclic deletion works exactly like other syntactic operations (driven by probe or structure-building features).[15] For concreteness, I would like to suggest the following approach to [†] feature spreading (but I hasten to add that alternative approaches are also conceivable, and readily available). Suppose that [†] can

---

[15] This reinforces the conclusions reached earlier, based on the inherently counter-cyclic nature of phase-based spell-out and the assignment of [†] features to a third stack.

optionally be inserted on any lexical item in the numeration.[16] Spreading of this feature in syntactic structures is then brought about by (21), which acts as a constraint on external Merge operations.[17]

(21) *Spreading of* [†]:
Suppose that α externally merges with β as a consequence of [$\bullet\beta\bullet$] on the head of α. Then, the head of α bears [†] iff the head of β bears [†], unless the head of β is $C_{[wh]}$.

The first thing to note here is that discharge of [†] by complement spell-out (as zero) must not imply that the feature is not available on the phase head aynmore that has triggered the operation. The feature persists on a phase head in some form, and is visible for further selection.[18] From (21), it then follows that a TP deletion operation in a sluicing construction can only be successfully carried out on the basis of a given numeration if (i) a [†] feature has initially been instantiated on all lexical items that will eventually end up in the c-command domain of the licensing $C_{[wh]}$ head;[19] and (ii) a [†] feature has not been instantiated on any lexical item that will end up outside of $C_{[wh]}$'s c-command domain. The only way to stop spreading of [†] at some point is to introduce $C_{[wh]}$ into the structure. $C_{[wh]}$ must then also bear [†], but when some other head (typically V) is subsequently merged with the CP headed by $C_{[wh]}$, this head does not have to bear [†], and [†] propagation (hence, deletion) is stopped. Nothing in (21) *precludes* the placement of [†] on, say, a matrix V selecting CP with a $C_{[wh],[†]}$ head; if [†] is inserted on such a V, spreading will continue, eventually yielding a syntactically legitimate sentence without PF-realization, which will (presumably) be uninterpretable because of recoverability.

---

[16] Thus, [†] can never be added in the course of the derivation, which avoids problems related to the Inclusiveness Condition and (more importantly, given the Inclusiveness Condition's minimal violability in favour of edge feature insertion) the Strict Cycle Condition.

[17] "Head of some category" is to be understood reflexively, such that the head of XP, X′, and X is X.

[18] Thus, the last two steps in (17) might actually look as in (i).

(i) *Specifiers of v as non-barriers,* [†] *on v, last two steps*
$\rightarrow$ | v: [†]° |
$\rightarrow$ | v: [†]— |

On this view, [†]° would be the basic, active feature, and [†] would be the form that the feature takes when it has been discharged and becomes inactive.

[19] Note that lexical items that are trivial phrases now also have [†] features even though there is no complement spell-out to trigger. However, this consequence is harmless: Either [†] can be discharged vacuously, or it need not be discharged at all if there is no spell-out operation. Alternatively, if this consequence were to be avoided, an additional clause that ensures this could be added in (21).

How can moved items (in particular, moved *wh*-phrases in sluicing constructions) avoid being assigned [†]? The answer to this is twofold. First, (21) is a constraint on external Merge; so internal Merge does not require the presence of [†] on the moved item even if the head that attracts it ($C_{[wh],[†]}$, in the case at hand) bears this feature. Second, how can the moved item avoid bearing [†] in situ where it has undergone *external* Merge with some other item? Here the answer may be based on an assumption that I have extensively motivated in the preceding chapters: Movement must be successive-cyclic (given the PIC), and assuming that all phrases qualify as phrases, it must take place via all phrase edges intervening between its base position and the ultimate target position. For a *wh*-phrase externally merged as a complement of, say, V, this means that is has to undergo extremely local movement to SpecV first before moving on to SpecC. Assuming that (21) is not a derivational constraint but a (mildly) *representational* constraint checked at the phase level (see the discussion in chapter 1, and note that (21) may be viewed as a requirement imposed by the PF interface), extremely local movement of β to a specifier of a phase head α can satisfy (21) even if α bears [†] and β does not. For a *wh*-phrase externally merged as a specifier of V or v, things are slightly more complicated. Here the most straightforward assumption might be that externally merged specifiers are in fact not part of the edge domain of a phase head; only internally merged specifiers are; see the pertinent discussion in section 5 of chapter 4 (page 210); thus, string-vacuous movement would be required here, and this would exempt a moved *wh*-phrase (or another item, for that matter), from obligatory deletion.

Such an approach to [†] spreading answers the basic questions that it was supposed to answer, but, needless to say, it also raises new ones. For instance, how can the analysis be conservatively extended to VP ellipsis? How can the apparently construction-specific assumption concerning $C_{[wh]}$ be avoided in (21)? And finally, is there a way to make numerations "smarter" (in the sense made precise in chapter 3) so as to avoid the consequence that given general optionality of [†] feature insertion, the vast majority of derivations will now be doomed to crash from the start, even though they may be filtered out fairly late in syntactic derivations? I firmly believe that the enterprise of providing reasonably principled answers to these (and other) questions is far from hopeless. However, since these questions do not affect the main point of the current chapter, I will, for now, leave it at that.

References

Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Connecticut, Storrs, Connecticut.

Abels, Klaus (2008): Towards a Restrictive Theory of (Remnant) Movement, *Linguistic Variation Yearbook* 7, 53–120.

Abels, Klaus (2011): Movement and Deletion: Locality and Sluicing. Ms., University College London.

Abels, Klaus & Kristine Bentzen (2011): Are Movement Paths Punctuated or Uniform?. In: A. Alexiadou, T. Kiss & G. Müller, eds., *Local Modelling of Non-Local Dependencies in Syntax*. Linguistische Arbeiten, De Gruyter, Berlin.

Abney, Steven (1987): The English Noun Phrase in Its Sentential Aspect. PhD thesis, MIT, Cambridge, Mass.

Ackema, Peter & Ad Neeleman (2003): Context-Sensitive Spell-Out, *Natural Language and Linguistic Theory* 21, 681–735.

Ackema, Peter & Ad Neeleman (2004): *Beyond Morphology*. Oxford University Press, Oxford.

Ackema, Peter, Ad Neeleman & Fred Weerman (1993): Deriving Functional Projections. In: *Proceedings of NELS*. Vol. 23, Amherst: GSLA, pp. 17–31.

Adger, David (2003): *Core Syntax*. Oxford University Press, Oxford, New York.

Adger, David & Daniel Harbour (2007): Syntax and Syncretisms of the Person Case Constraint, *Syntax* 10, 2–37.

Adger, David & Gillian Ramchand (2005): Merge and Move: Wh-Dependencies Revisited, *Linguistic Inquiry* 36, 161–193.

Adger, David & Peter Svenonius (2003): Beyond the Interface, *GLOW Newsletter* 50, Spring 2003.

Agbayani, Brian (1998): Feature Attraction and Category Movement. PhD thesis, UC Irvine.

Aissen, Judith (1999): Markedness and Subject Choice in Optimality Theory, *Natural Language and Linguistic Theory* 17, 673–711.

Alexiadou, Artemis (1997): *Adverb Placement*. Benjamins, Amsterdam.

Alexiadou, Artemis & Gereon Müller (2008): Class Features as Probes. In: A. Bachrach & A. Nevins, eds., *Inflectional Identity*. Oxford University Press, Oxford, pp. 101–155.

Alexiadou, Artemis & Gisbert Fanselow (2000): On the Correlation between Morphology and Syntax: The Case of V-to-I. In: W. Abraham & J.-W. Zwart, eds., *Proceedings fromthe 15th Comparative Germanic Syntax Workshop*. Benjamins, Amsterdam, pp. 219–242.

Alexiadou, Artemis, Liliane Haegeman & Melita Stavrou (2007): *Noun Phrase in the Generative Perspective*. De Gruyter, Berlin, New York.

Anagnostopoulou, Elena (2005): Strong and Weak Person Restrictions. A Feature Checking Analysis. In: L. Heggie & F. Ordoñez, eds., *Clitics and Affix Combinations*. Benjamins, Amsterdam, pp. 199–235.

Aoun, Joseph (1986): *Generalized Binding*. Foris, Dordrecht.

Aoun, Joseph, Norbert Hornstein & Dominique Sportiche (1981): Aspects of Wide Scope Interpretation, *Journal of Linguistic Research* 1, 69–95.

Assmann, Anke (2010): Parasitic Gaps in Derivational Grammar. Master's thesis, Universität Leipzig.

Assmann, Anke & Fabian Heck (2011): The Retreat of the Base Generators. Ms., Universität Leipzig.

Assmann, Anke, Fabian Heck, Johannes Hein, Stefan Keine & Gereon Müller (2010): Does Chain Hybridization in Irish Support Movement-Based Approaches to Long-Distance Dependencies?. In: S. Müller, ed., *The Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford, pp. 27–46.

Axel, Katrin (2007): *Studies on Old High German Syntax: Left Sentence Periphery, Verb Placement and Verb-Second*. Benjamins, Amsterdam.

Bach, Emmon & George Horn (1976): Remarks on 'Conditions on Transformations', *Linguistic Inquiry* pp. 265–299.

Baker, Carl L. (1970): Notes on the Description of English Questions: The Role of an Abstract Question Morpheme, *Foundations of Language* 6, 197–219.

Baker, Mark (1988): *Incorporation. A Theory of Grammatical Function Changing*. University of Chicago Press, Chicago.

Barbiers, Sjef (2002): Remnant Stranding and the Theory of Movement. In: *Dimensions of Movement*. Benjamins, Amsterdam, pp. 47–67.

Barss, Andrew (1986): Chains and Anaphoric Dependence. Ph.d. thesis, MIT, Cambridge, Mass.

Barss, Andrew & Howard Lasnik (1986): A Note on Anaphora and Double Objects, *Linguistic Inquiry* 17, 347–354.

Bayer, Josef (1990): Notes on the ECP in English and German, *Groninger Arbeiten zur Germanistischen Linguistik* 30, 1–51.

Bayer, Josef & Jaklin Kornfilt (1994): Against Scrambling as an Instance of Move Alpha. In: N. Corver & H. van Riemsdijk, eds., *Studies on Scrambling*. Mouton de Gruyter, Berlin, pp. 17–60.

Bayer, Josef & Martin Salzmann (2009): That-Trace Effects and Resumption as the Result of Improper Movement. Ms., Universität Konstanz.

Bech, Gunnar (1955/1957): *Studien über das deutsche Verbum Infinitum*. Niemeyer, Tübingen. Reprint 1983.

Beck, Sigrid (1997): Wh-Constructions and Transparent Logical Form. PhD thesis, Universität Tübingen.

Beckmann, Jill (1998): Positional Faithfulness. PhD thesis, University of Massachusetts, Amherst.

Béjar, Susana & Milan Řezáč (2009): Cyclic Agree, *Linguistic Inquiry* 40, 35–73.

Belletti, Adriana & Luigi Rizzi (1988): Psych-Verbs and Θ-Theory, *Natural Language and Linguistic Theory* 6, 291–352.

Bentzen, Kristine (2009): Embedded Root Phenomena, Assertion, Presupposition, and Main Point of Utterance. Ms., University of Tromsø.

Bhatt, Rajesh (2005): Long Distance Agreement in Hindi-Urdu, *Natural Language and Linguistic Theory* 23, 757–807.

Bickel, Balthasar (2004): Constraining Focus and Extraction Domains. Ms., Universität Leipzig. (Talk presented at *Syntax of the World's Languages 1*).

Bierwisch, Manfred (1963): *Grammatik des deutschen Verbs*. Studia Grammatica II, Akademie Verlag, Berlin.

Bierwisch, Manfred (1967): Syntactic Features in Morphology: General Problems of So-Called Pronominal Inflection in German. In: *To Honor Roman Jakobson*. Mouton, The Hague/Paris, pp. 239–270.

Bierwisch, Manfred (1988): On the Grammar of Local Prepositions. In: M. Bierwisch, W. Motsch & I. Zimmermann, eds., *Syntax, Semantik und Lexikon. Rudolf Růžička zum 65. Geburtstag*. Akademie-Verlag, Berlin, pp. 1–65.

Biskup, Petr (2009): The Phase Model and Adverbials. PhD thesis, Universität Leipzig.

Bittner, Maria & Ken Hale (1996): The Structural Determination of Case and Agreement, *Linguistic Inquiry* pp. 1–68.

Bobaljik, Jonathan (1995): Morphosyntax: The Syntax of Verbal Inflection. PhD thesis, MIT, Cambridge, Mass.

Bobaljik, Jonathan (2002): Realizing Germanic Inflection: Why Morphology Does Not Drive Syntax, *Journal of Comparative Germanic Linguistics* 6, 129–167.

Bobaljik, Jonathan & Susanne Wurmbrand (2003): Relativized Phases. Ms., University of Connecticut, Storrs.

Bobaljik, Jonathan & Susanne Wurmbrand (2008): Word Order and Scope. Transparent Interfaces and the 3/4 Signature. Ms., University of Connecticut.

Boeckx, Cedric (2003): *Islands and Chains. Resumption as Stranding*. Benjamins, Amsterdam.

Boeckx, Cedric (2004): Long-distance Agreement in Hindi: Some Theoretical Implications, *Studia Linguistica* 58, 23–36.

Boeckx, Cedric (2008): *Understanding Minimalist Syntax*. Blackwell, Oxford.

Boeckx, Cedric & Kleanthes K. Grohmann (2007): Putting Phases in Perspective, *Syntax* 10, 204–222.

Boeckx, Cedric, Norbert Hornstein & Jairo Nunes (2010): *Control as Movement*. Cambridge University Press, Cambridge.

Borer, Hagit (2004): *Structuring Sense. An Exo-Skeletal Trilogy. Vols 1 and 2*. Oxford University Press, Oxford.

Bouma, Gosse, Robert Malouf & Ivan Sag (2001): Satisfying Constraints on Extraction and Adjunction, *Natural Language and Linguistic Theory* pp. 1–65.

Bošković, Željko (1997): *The Syntax of Nonfinite Complementation. An Economy Approach*. MIT Press, Cambridge, Mass.

Bošković, Željko (2002): A-Movement and the EPP, *Syntax* 5, 167–218.

Bošković, Željko (2007): Agree, Phases, and Intervention Effects, *Linguistic Analysis* 33, 54–96.

Bošković, Željko (2011): Rescue by PF Deletion, Traces as (Non)interveners, and the That-Trace Effect, *Linguistic Inquiry* 42, 1–44.

Bošković, Željko & Howard Lasnik (1999): How Strict is the Cycle?, *Linguistic Inquiry* 20, 691–703.

Bresnan, Joan (1976): On the Form and Functioning of Transformations, *Linguistic Inquiry* 7, 3–40.

Brody, Michael (2001): Some Aspects of Elegant Syntax. Ms., University College London.

Brody, Michael (2002): On the Status of Representations and Derivations. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 19–41.

Broekhuis, Hans (2005): Extraction from Subjects: Some Remarks on Chomsky's 'On Phases'. In: H. Broekhuis, N. Corver, R. Huybregts, U. Kleinherz & J. Koster, eds., *Organizing Grammar. Studies in Honor of Henk van Riemsdijk*. Mouton de Gruyter.

Brosziewski, Ulf (2003): *Syntactic Derivations. A Nontransformational View*. Number 470 *in* 'Linguistische Arbeiten', Niemeyer Verlag, Tübingen.

Browning, Marguerite (1987): Null Operator Constructions. PhD thesis, MIT, Cambridge, Mass.

Browning, Marguerite (1991): Bounding Conditions on Representations, *Linguistic Inquiry* 22, 541–562.

Büring, Daniel (2005): *Binding Theory*. Cambridge University Press, Cambridge.

Büring, Daniel & Katharina Hartmann (1994): The Dark Side of Wh-Movement, *Linguistische Berichte* 149, 56–74.

Butt, Miriam (1995): *The Structure of Complex Predicates in Urdu*. CSLI Publications, Standford, California.

Cable, Seth (2007): The Grammar of Q. PhD thesis, MIT, Cambridge, Mass.

Caha, Pavel (2007): The Shape of Paradigms. Ms., University of Tromsø.

Caha, Pavel (2008): The Case Hierarchy as Functional Sequence. In: M. Richards & A. Malchukov, eds., *Scales*. Vol. 86 of *Linguistische Arbeitsberichte*, Universität Leipzig, pp. 247–276.

Cattell, Ray (1976): Constraints on Movement Rules, *Language* 52, 18–50.

Ćavar, Damir (1996): Untitled. Ms., Universität Potsdam.

Chandra, P. (2005): Hindi-Urdu Long Distance Agreement: Agree, AGREE or Spec-Head?. Ms, University of Maryland, College Park.

Chomsky, Noam (1964): *Current Issues in Linguistic Theory*. Mouton, The Hague.

Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, Noam (1972): Some Empirical Issues in the Theory of Transformational Grammar. In: S. Peters, ed., *Goals of Linguistic Theory*. Prentice-Hall, Englewood Cliffs.

Chomsky, Noam (1973): Conditions on Transformations. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle*. Academic Press, New York, pp. 232–286.

Chomsky, Noam (1975): *The Logical Structure of Linguistic Theory*. Plenum Press, New York.

Chomsky, Noam (1977): On Wh-Movement. In: P. Culicover, T. Wasow & A. Akmajian, eds., *Formal Syntax*. Academic Press, New York, pp. 71–132.

Chomsky, Noam (1981): *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam (1982a): *The Generative Enterprise. A Discussion with Riny Huybregts and Henk van Riemsdijk*. Foris Publications, Dordrecht.

Chomsky, Noam (1982b): *Some Concepts and Consequences of the Theory of Government and Binding*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1986a): *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1986b): *Knowledge of Language*. Praeger, New York.

Chomsky, Noam (1991): Some Notes on Economy of Derivation and Representation. In: R. Freidin, ed., *Principles and Parameters in Comparative Grammar*. MIT Press, Cambridge, Mass., pp. 417–454.

Chomsky, Noam (1993): A Minimalist Program for Syntactic Theory. In: K. Hale & S. J. Keyser, eds., *The View from Building 20*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. In: R. Martin, D. Michaels & J. Uriagereka, eds., *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001a): Beyond Explanatory Adequacy. Ms., MIT, Cambridge, Mass.

Chomsky, Noam (2001b): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (2002): *On Nature and Language*. Cambridge University Press, Cambridge.

Chomsky, Noam (2005): Three Factors in Language Design, *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam (2007): Approaching UG from Below. In: U. Sauerland & H.-M. Gärtner, eds., *Interfaces + Recursion = Language?*. Mouton de Gruyter, Berlin, pp. 1–31.

Chomsky, Noam (2008): On Phases. In: R. Freidin, C. Otero & M. L. Zubizarreta, eds., *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, Mass., pp. 133–166.

Chomsky, Noam & Howard Lasnik (1977): Filters and Control, *Linguistic Inquiry* 8, 425–504.

Chung, Sandra (1994): Wh-Agreement and 'Referentiality' in Chamorro, *Linguistic Inquiry* 25, 1–44.

Chung, Sandra (1998): *The Design of Agreement: Evidence from Chamorro*. Chicago University Press, Chicago.

Chung, Sandra, William Ladusaw & James McCloskey (1995): Sluicing and Logical Form, *Natural Language Semantics* 3, 1–44.

Cinque, Guglielmo (1990): *Types of A-bar Dependencies*. MIT Press, Cambridge, Mass.

Cinque, Guglielmo (1999): *Adverbs and Functional Heads*. Oxford University Press, Oxford.

Clements, George, James McCloskey, Joan Maling & Annie Zaenen (1983): String-Vacuous Rule Application, *Linguistic Inquiry* 14, 1–17.

Clifton, Charles, Gisbert Fanselow & Lyn Frazier (2006): Amnestying Superiority Violations: Processing Multiple Questions, *Linguistic Inquiry* 37, 51–68.

Cole, Peter & Gabriella Hermon (2000): Partial Wh-Movement: Evidence from Malay. In: U. Lutz, G. Müller & A. von Stechow, eds., *Wh-Scope Marking*. Benjamins, Amsterdam, pp. 101–130.

Collins, Chris (1993): Topics in Ewe Syntax. PhD thesis, MIT, Cambridge, Mass.

Collins, Chris (1994): Economy of Derivation and the Generalized Proper Binding Condition, *Linguistic Inquiry* 25, 45–61.

Collins, Chris (1997): *Local Economy*. MIT Press, Cambridge, Mass.

Collins, Chris (2003): Eliminating Labels. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 42–64.

Collins, Chris & Höskuldur Thráinsson (1996): VP-internal Structure and Object Shift in Icelandic, *Linguistic Inquiry* 27, 391–447.

Culicover, Peter (1991): Topicalization, Inversion, and Complementizers in English. In D. Delfitto (et al.), eds., *Going Romance, and Beyond*. OTS Working Papers, Utrecht, 85-118.

Culicover, Peter (2008): Beyond Simpler Syntax: Complexity and Explanation. Ms.,

Ohio State University. (Talk at Maryland Mayfest 2008 on *Island Perspectives*).

D'Alessandro, Roberta & Ian Roberts (2008): Movement and Agreement in Italian Past Participles and Defective Phases, *Linguistic Inquiry* 39, 477–491.

Dalrymple, Mary (2001): *Lexical Functional Grammar*. Vol. 34 of *Syntax and Semantics*, Academic Press, San Diego.

Davies, William & Stanley Dubinsky (2003): On Extraction from NPs, *Natural Language and Linguistic Theory* 21, 1–37.

Dayal, Veneeta (1994): Scope Marking as Indirect Wh-Dependency, *Natural Language Semantics* 2, 137–170.

de Cuba, Carlos (2007): On (Non)Factivity, Clausal Complementation and the CP-Field. PhD thesis, Stony Brook University.

de Hoop, Helen & Andrej Malchukov (2008): Case-Marking Strategies, *Linguistic Inquiry* 39, 565–587.

den Besten, Hans (1977): On the Interaction of Root Transformations and Lexical Deletive Rules. Ms., University of Amsterdam.

den Besten, Hans & Gert Webelhuth (1990): Stranding. In: G. Grewendorf & W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 77–92.

den Dikken, Marcel (2007): Phase Extension: Contours of a Theory of the Role of Head Movement in Phrasal Extraction, *Theoretical Linguistics* 33, 1–41.

den Dikken, Marcel (2008): Small Clauses, Phases, and Phase Extension – The Case of Scope and Object Shift. Ms.,CUNY Graduate Center. (Talk at Glow 2008, Newcastle).

Di Sciullo, Anna Maria & Daniela Isac (2008): The Asymmetry of Merge, *Biolinguistics* 2, 260–290.

Diesing, Molly (1992): *Indefinites*. MIT Press, Cambridge, Mass.

Embick, David & Rolf Noyer (2001): Movement Operations after Syntax, *Linguistic Inquiry* 32, 555–595.

Emonds, Joseph (1978): The Verbal Complex V′-V in French, *Linguistic Inquiry* 9, 151–175.

Epstein, Sam, Erich Groat, Ruriko Kawashima & Hisatsugu Kitahara (1998): *A Derivational Approach to Syntactic Relations*. Oxford University Press, Oxford and New York.

Epstein, Samuel David (1992): Derivational Constraints on $\overline{\text{A}}$-Chain Formation, *Linguistic Inquiry* 23, 235–259.

Epstein, Samuel David & T. Daniel Seely (2002): Rule Applications as Cycles in a Level-Free Syntax. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65–89.

Fanselow, Gisbert (1987): *Konfigurationalität*. Narr, Tübingen.

Fanselow, Gisbert (1988): Aufspaltung von NPn und das Problem der 'freien' Wortstellung, *Linguistische Berichte* 114, 91–113.

Fanselow, Gisbert (1990a): On the Nature of Proper Government and Syntactic Barriers. In: W. Abraham, W. Kosmeijer & E. Reuland, eds., *Issues in Germanic Syntax*. Mouton de Gruyter, Berlin, pp. 33–48.

Fanselow, Gisbert (1990b): Scrambling as NP-Movement. In: G. Grewendorf & W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 113–140.

Fanselow, Gisbert (1991): Minimale Syntax. Habilitation thesis, Universität Passau.

Fanselow, Gisbert (1992): The Return of the Base Generators. Ms., Universität Passau.

Fanselow, Gisbert (1996): The Proper Interpretation of the Minimal Link Condition. Ms. Universität Potsdam.

Fanselow, Gisbert (2001a): Features, Theta-Roles, and Free Constituent Order, *Linguistic Inquiry* 32, 405–436.

Fanselow, Gisbert (2001b): When Formal Features Need Company. In: C. Féry & W. Sternefeld, eds., *Audiatur Vox Sapientiae. A Festschrift for Arnim von Stechow*. Akademie-Verlag, Berlin, pp. 131–152.

Fanselow, Gisbert (2003): Münchhausen-Style Head Movement and the Analysis of Verb-Second. In: A. Mahajan, ed., *Head Movement and Syntactic Theory*. Vol. 3 of *Syntax at Sunset*, UCLA & Universität Potsdam Working Papers in Linguistics, Los Angeles & Potsdam, pp. 40–76.

Fanselow, Gisbert (2004): The MLC and Derivational Economy. In: A. Stepanov, G. Fanselow & R. Vogel, eds., *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 73–123.

Fanselow, Gisbert (2009): Bootstrapping Verb Movement and the Clausal Architecture of German (and Other Languages). In: A. Alexiadou, J. Hankamer, T. McFadden, J. Nuger & F. Schäfer, eds., *Advances in Comparative Germanic Syntax*. Benjamins, Amsterdam, pp. 85–118.

Fanselow, Gisbert (2011): Deutsche Wortstellung: Drei Gründe für Ihre Wahlentscheidung. Ms., Universität Potsdam.

Fanselow, Gisbert & Anoop Mahajan (1996): Partial Movement and Successive Cyclicity. In: U. Lutz & G. Müller, eds., *Papers on Wh-Scope Marking*. Number 76 *in* 'Arbeitspapiere', SFB 340, Universität Stuttgart and Universität Tübingen, pp. 131–177.

Fanselow, Gisbert & Anoop Mahajan (2000): Towards a Minimalist Theory of Wh-Expletives, Wh-Copying, and Successive Cyclicity. In: U. Lutz, G. Müller & A. von Stechow, eds., *Wh-Scope Marking*. Benjamins, Amsterdam, pp. 195–230.

Fanselow, Gisbert & Damir Ćavar (2001): Remarks on the Economy of Pronunciation. In: G. Müller & W. Sternefeld, eds., *Competition in Syntax*. Mouton de Gruyter, Berlin, pp. 107–150.

Fanselow, Gisbert & Damir Ćavar (2002): Distributed Deletion. In: A. Alexiadou, ed., *Theoretical Approaches to Universals*. Benjamins, Amsterdam, pp. 65–107.

Fanselow, Gisbert & Denisa Lenertová (2011): Left-Peripheral Focus. Mismatches between Syntax and Information Structure, *Natural Language and Linguistic Theory* 29, 169–209.

Featherston, Sam (2001): Universals and Grammaticality: Wh-Constraints in German and English. Ms., Universität Tübingen.

Ferguson, K. Scott (1993): Notes on the Shortest Move Metric and Object Checking, *Harvard Working Papers in Linguistics* 3, 65–80.

Ferguson, Scott & Erich Groat (1994): Defining 'Shortest Move'. Ms., Harvard University.

Fiengo, Robert (1980): *Surface Structure*. Harvard University Press, Cambridge, Mass.

Fiengo, Robert, Cheng-Teh James Huang, Howard Lasnik & Tanya Reinhart (1988): The Syntax of Wh-in-situ. In: H. Borer, ed., *Proceedings of WCCFL 7*. CSLI Publications, Stanford, pp. 81–98.

Fischer, Silke (2004): Towards an Optimal Theory of Reflexivization. PhD thesis, Universität Tübingen.

Fischer, Silke (2006): Matrix Unloaded: Binding in a Local Derivational Approach, *Linguistics* 44, 913–935.

Fitzpatrick, Justin (2002): On Minimalist Approaches to the Locality of Movement, *Linguistic Inquiry* 33, 443–463.

Fox, Danny (2000): *Economy and Semantic Interpretation*. MIT Press, Cambridge, Mass.

Fox, Danny & Howard Lasnik (2003): Successive-Cyclic Movement and Island Repair: The Difference between Sluicing and VP-Ellipsis, *Linguistic Inquiry* 34, 143–154.

Frampton, John (1990): Parasitic Gaps and the Theory of Wh-Chains, *Linguistic Inquiry* 21, 49–77.

Frampton, John & Sam Gutman (1999): Cyclic Computation, *Syntax* 2, 1–27.

Frank, Robert (2002): *Phrase Structure Composition and Syntactic Dependencies*. MIT Press, Cambridge, Mass.

Frank, Robert (2006): Phase Theory and Tree Adjoining Grammar, *Lingua* 116, 145–202.

Frantz, Donald (1980): Ascensions to Subject in Blackfoot. In: M. Niepokuj, M. V. Clay, V. Nikiforidou & D. Feder, eds., *Proceedings of the 6th Annual Meeting of the Berkeley Linguistics Society*. BLS, Berkeley, University of California, pp. 293–299.

Freidin, Robert (1992): *Foundations of Generative Syntax*. MIT Press, Cambridge, Mass.

Freidin, Robert (1994): Generative Grammar: Principles and Parameters Framework. In: R. Asher & J. Simpson, eds., *The Encyclopedia of Language and Linguistics, vol. III*. Pergamon Press, Oxford, pp. 1370–1385.

Freidin, Robert (1999): Cyclicity and Minimalism. In: S. D. Epstein & N. Hornstein, eds., *Working Minimalism*. MIT Press, Cambridge, Mass, pp. 95–126.

Frey, Werner (1993): *Syntaktische Bedingungen für die Interpretation*. Akademieverlag, Berlin.

Frey, Werner (2000): Über die syntaktische Position der Satztopiks im Deutschen. In: *ZAS Papers in Linguistics*. Vol. 20, pp. 137–172.

Frey, Werner (2004a): A Medial Topic Position for German, *Linguistische Berichte* 198, 153–190.

Frey, Werner (2004b): The Grammar-Pragmatics Interface and the German Prefield, *Sprache & Pragmatik* 52, 1–39.

Frey, Werner & Thilo Tappe (1991): Zur Interpretation der X-bar-Theorie und zur Syntax des Mittelfeldes. Grundlagen eines GB-Fragmentes. Ms., Universität Stuttgart.

Freywald, Ulrike (2009): V2 in German Complement Clauses Introduced by *dass* 'that'. Ms., Universität Potsdam.

Fukui, Naoki (1997): Attract and the A-over-A Principle, *UCI Working Papers in Linguistics* 3, 51–67.

Gallego, Ángel (2007): Phase Theory and Parametric Variation. PhD thesis, Universitat Autónoma de Barcelona, Barcelona.

Gallego, Ángel & Juan Uriagereka (2006): Sub-Extraction from Subjects. Ms., Universitat Autónoma de Barcelona.

Gärtner, Hans-Martin (2000): Are There V2 Relative Clauses in German?, *The Journal of Comparative Germanic Linguistics* 3, 97–141.

Gärtner, Hans-Martin & Jens Michaelis (2007): Some Remarks on Locality Conditions and Minimalist Grammars. In: U. Sauerland & H.-M. Gärtner, eds., *Interfaces + Recursion = Language?*. Mouton de Gruyter, Berlin, pp. 161–195.

Gazdar, Gerald (1981): Unbounded Dependencies and Coordinate Structure, *Linguistic Inquiry* 12, 155–184.

Gazdar, Gerald (1982): Phrase Structure Grammar. In: *The Nature of Syntactic Representation*. Reidel, Dordrecht, pp. 131–186.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag (1985): *Generalized Phrase Structure Grammar*. Blackwell, Oxford.

Georgi, Doreen (2010): A Local Derivation of Global Case Splits. Ms., Universität Leipzig.

Georgi, Doreen & Gereon Müller (2010): Noun Phrase Structure by Reprojection, *Syntax* 13(1), 1–36.

Georgi, Doreen, Fabian Heck & Gereon Müller (2009): Maraudage. Ms., Universität Leipzig. (Talk at the Repairs Workshop of the DGfS conference 2009, Osnabrück, and at the Potsdam/Leipzig Workshop on Movement and Morphology, Leucorea, April 2009).

Graf, Thomas (2008a): Ein modelltheoretischer Blick auf syntaktische Constraints. Ms., Universität Wien.

Graf, Thomas (2008b): The Formal Landscape of Syntactic Constraints. Ms., UCLA.

Grewendorf, Günther (1988): *Aspekte der deutschen Syntax*. Narr.

Grewendorf, Günther (1989): *Ergativity in German*. Foris, Dordrecht.

Grewendorf, Günther (2001): Multiple Wh-Fronting, *Linguistic Inquiry* 32, 87–122.

Grewendorf, Günther (2002): *Minimalistische Syntax*. Francke/UTB, Tübingen and Basel.

Grewendorf, Günther (2003): Improper Remnant Movement, *Gengo Kenkyo: The Journal of the Linguistic Society of Japan* 123, 47–94.

Grewendorf, Günther (2004): Die linke Bewegung von Restkategorien. Ms., Universität Frankfurt/Main.

Grewendorf, Günther & Joachim Sabel (1994): Long Scrambling and Incorporation, *Linguistic Inquiry* 25, 263–308.

Grewendorf, Günther & Joachim Sabel (1999): Scrambling in German and Japanese, *Natural Language and Linguistic Theory* 17, 1–65.

Grimshaw, Jane (1986): Subjacency and the S/S′ Parameter, *Linguistic Inquiry* 17, 364–369.

Grimshaw, Jane (1994): Heads and Optimality. Ms., Rutgers University. (Handout of talk, Universität Stuttgart).

Grimshaw, Jane (1997): Projection, Heads, and Optimality, *Linguistic Inquiry* 28, 373–422.

Grimshaw, Jane (1998): Constraints on Constraints in Optimality Theoretic Syntax. Ms., Rutgers University, New Brunswick, New Jersey.

Grimshaw, Jane & Vieri Samek-Lodovici (1998): Optimal Subjects and Subject Universals. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 193–219.

Grohmann, Kleanthes (1997): German Superiority, *Groninger Arbeiten zur Germanistischen Linguistik* 40, 97–107.

Grohmann, Kleanthes (1998): Syntactic Inquiries into Discourse Restrictions on Multiple Interrogatives, *Groninger Arbeiten zur Germanistischen Linguistik* 42, 1–60.

Grohmann, Kleanthes K. (2000): Prolific Peripheries: A Radical View From the Left. PhD thesis, University of Maryland.

Grohmann, Kleanthes K. (2003): Successive Cyclicity under (Anti-)Local Considerations, *Syntax* 6, 260–312.

Haegeman, Liliane (1995): *An Introduction to Government and Binding Theory*. 2 edn, Blackwell, Oxford.

Haider, Hubert (1983): Connectedness Effects in German, *Groninger Arbeiten zur Germanistischen Linguistik* 23, 82–119.

Haider, Hubert (1984): Topic, Focus, and V-Second, *Groninger Arbeiten zur German-*

istischen Linguistik 25, 72–120.

Haider, Hubert (1988): Theta-Tracking Systems – Evidence from German. In: L. Marácz & P. Muysken, eds., *Configurationality*. Foris, Dordrecht, pp. 185–206.

Haider, Hubert (1992): The Basic Branching Conjecture. Ms., Universität Stuttgart.

Haider, Hubert (1993a): *Deutsche Syntax – generativ*. Narr, Tübingen.

Haider, Hubert (1993b): ECP-Etüden: Anmerkungen zur Extraktion aus eingebetteten Verb-Zweit-Sätzen, *Linguistische Berichte* 145, 185–203.

Haider, Hubert (2000a): Branching and Discharge. In: P. Coopmans, M. Everaert & J. Grimshaw, eds., *Lexical Specification and Insertion*. Benjamins, Amsterdam, pp. 135–164.

Haider, Hubert (2000b): Towards a Superior Account of Superiority. In: U. Lutz, G. Müller & A. von Stechow, eds., *Wh-Scope Marking*. Benjamins, Amsterdam, pp. 231–248.

Haider, Hubert (2002): Superiority Revisited: Dutch, English, German, Icelandic Contrasts. A Representational Account. Ms., Universität Salzburg.

Haider, Hubert (2004): The Superiority Conspiracy: Four Constraints and a Processing Effect. In: A. Stepanov, G. Fanselow & R. Vogel, eds., *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 147–175.

Haider, Hubert (2006): Zufall und Notwendigkeit bei der germanischen VO/OV-Spaltung. Ms., Universität Salzburg.

Hale, Ken (1972): A New Perspective on American Indian Linguistics. In: A. Ortiz, ed., *New Perspectives on the Pueblos*. University of New Mexico Press, Albuquerque, pp. 87–103.

Hale, Ken & Samuel Jay Keyser (2002): *Prolegomenon to a Theory of Argument Structure*. MIT Press, Cambridge, Mass.

Halle, Morris (1997): Distributed Morphology: Impoverishment and Fission. In: B. Bruening, Y. Kang & M. McGinnis, eds., *Papers at the Interface*. Vol. 30, MITWPL, pp. 425–449.

Halle, Morris & Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. In: K. Hale & S. J. Keyser, eds., *The View from Building 20*. MIT Press, Cambridge, Mass., pp. 111–176.

Halle, Morris & Alec Marantz (1994): Some Key Features of Distributed Morphology. In: A. Carnie, H. Harley & T. Bures, eds., *Papers on Phonology and Morphology*. Vol. 21 of *MIT Working Papers in Linguistics*, MITWPL, Cambridge, Mass., pp. 275–288.

Harbert, Wayne & William Pet (1988): Movement and Adjunct Morphology in Arawak and Other Languages, *International Journal of American Linguistics* 54, 416–435.

Harley, Heidi (1995): Subjects, Events, and Licensing. PhD thesis, MIT.

Harley, Heidi & Elisabeth Ritter (2002): Person and Number in Pronouns: A Feature-

Geometric Analysis, *Language* 78, 482–526.

Hartmann, Jutta (2005): Wh-Movement and the Small Clause Analyses of the English 'There'-Construction, *Leiden Papers in Linguistics* 2.3, 93–106.

Hartmann, Katharina & Malte Zimmermann (2007): In Place – Out of Place? Focus in Hausa. In: K. Schwabe & S. Winkler, eds., *On Information Structure, Meaning and Form*. Benjamins, Amsterdam, pp. 365–403.

Hawkins, John (1999): Processing Complexity and Filler-Gap Dependencies Across Grammars, 75, 244–285.

Hazout, Ilan (2004): The Syntax of Existential Constructions, *Linguistic Inquiry* 35, 393–430.

Heck, Fabian (2004): A Theory of Pied Piping. PhD thesis, Universität Tübingen.

Heck, Fabian (2008): *On Pied-Piping. Wh-Movement and Beyond*. Vol. 98 of *Studies in Generative Grammar*, Mouton de Gruyter, Berlin.

Heck, Fabian (2009): On Certain Properties of Pied Piping, *Linguistic Inquiry* 40, 75–111.

Heck, Fabian (2010): Against Direct Recursion in Syntax. Ms., Universität Leipzig.

Heck, Fabian & Gereon Müller (2000): Successive Cyclicity, Long-Distance Superiority, and Local Optimization. In: R. Billerey & B. D. Lillehaugen, eds., *Proceedings of WCCFL*. Vol. 19, Cascadilla Press, Somerville, MA, pp. 218–231.

Heck, Fabian & Gereon Müller (2003): Derivational Optimization of Wh-Movement, *Linguistic Analysis* 33, 97–148. (Volume appeared 2007).

Heck, Fabian & Gereon Müller (2007): Extremely Local Optimization. In: E. Brainbridge & B. Agbayani, eds., *Proceedings of the 26th WECOL*. California State University, Fresno, pp. 170–183.

Heck, Fabian & Juan Cuartero (2011): Long-Distance Agreement in Relative Clauses. In: A. Alexiadou, T. Kiss & G. Müller, eds., *Local Modelling of Non-Local Dependencies in Syntax*. de Gruyter, Berlin.

Heck, Fabian & Malte Zimmermann (2004): DPs as Phases. Ms., Universität Leipzig and HU Berlin.

Heck, Fabian, Gereon Müller & Jochen Trommer (2008): A Phase-Based Approach to Scandinavian Definiteness Marking. In: C. B. Chang & H. J. Haynie, eds., *Proceedings of WCCFL 26*. Cascadilla Press, Somerville, Massachusetts, pp. 226–233.

Heim, Irene & Angelika Kratzer (1998): *Semantics in Generative Grammar*. Blackwell, Oxford.

Hiraiwa, Ken (2005): Dimensions of Symmetry in Syntax: Agreement and Clausal Architecture. PhD thesis, MIT, Cambridge, Mass.

Hole, Daniel (2008): Dativ, Bindung und Diathese. Habilitation Thesis, Humboldt-Universität Berlin.

Holmberg, Anders (1991): Head Scrambling. Handout of talk; GLOW 1991 (Leiden).

Holmberg, Anders (2002): Expletives and Agreement in Scandinavian Passives, *Jour-

nal of Comparative Germanic Linguistics 4, 85–128.

Holmberg, Anders (2005): Is There a Little Pro? Evidence from Finnish, *Linguistic Inquiry* 36, 533–564.

Holmberg, Anders & Christer Platzack (1995): *The Role of Inflection in the Syntax of the Scandinavian Languages*. Oxford University Press, Oxford.

Hooper, Joan & Sandra Thompson (1973): On the Applicability of Root Transformations, *Linguistic Inquiry* 4, 465–497.

Hornstein, Norbert (1995): *Logical Form*. Blackwell, Oxford.

Hornstein, Norbert (2001): *Move. A Minimalist Theory of Construal*. Blackwell, Oxford.

Hornstein, Norbert (2009): *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press, Cambridge.

Hornstein, Norbert & Amy Weinberg (1990): The Necessity of LF, *The Linguistic Review* 7, 129–167.

Hornstein, Norbert & Amy Weinberg (1995): The Empty Category Principle. In: G. Webelhuth, ed., *Government and Binding Theory and the Minimalist Program*. Blackwell, Oxford and Cambridge, pp. 241–296.

Hornstein, Norbert & Juan Uriagereka (2002): Reprojections. In: S. D. Epstein & D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, London, pp. 106–132.

Huang, Cheng-Teh James (1982): Logical Relations in Chinese and the Theory of Grammar. PhD thesis, MIT, Cambridge, Mass.

Huang, Cheng-Teh James (1995): Logical Form. In: G. Webelhuth, ed., *Government and Binding Theory and the Minimalist Program*. Blackwell, Oxford, pp. 125–175.

Iatridou, Sabine (1990): About Agr(P), *Linguistic Inquiry* 21, 551–577.

Jackendoff, Ray (1990): On Larson's Account of the Double Object Construction, *Linguistic Inquiry* 21, 427–454.

Jacobson, Paquline (1999): Towards a Variable-Free Semantics, *Linguistics and Philosophy* 22(2), 117–185.

Jaeggli, Osvaldo & Ken Safir (1989): The Null Subject Parameter and Parametric Theory. In: O. Jaeggli & K. Safir, eds., *The Null Subject Parameter*. Kluwer, Dordrecht, pp. 1–44.

Johnson, David & Shalom Lappin (1997): *Local Constraints vs. Economy*. CSLI Publications, Stanford University.

Johnson, Kyle (2003): Towards and Etiology of Adjunct Islands, *Nordlyd* 31(1), 187–215.

Jurka, Johannes (2008): Extraction out of Subjects in German. Ms., University of Maryland. (Talk at 36th Österreichische Linguistiktagung, Wien University, 2008).

Jurka, Johannes (2010): The Importance of Being a Complement. CED-Effects Revisited. PhD thesis, University of Maryland.

Kayne, Richard (1984): *Connectedness and Unambiguous Paths*. Foris, Dordrecht.

Kayne, Richard (1994): *The Antisymmetry of Syntax*. MIT Press, Cambridge, Mass.

Kayne, Richard (1998): Overt vs. Covert Movement, *Syntax* 1, 128–191.

Keine, Stefan (2008): Long-Distance Agreement und Zyklisches Agree. Ms., Universität Leipzig.

Keine, Stefan (2010): *Case and Agreement from Fringe to Core. Impoverishment Effects on Agree*. Linguistische Arbeiten, Mouton de Gruyter, Berlin.

Kim, Shin-Sook (2002): Intervention Effects are Focus Effects, *Japanese/Korean Linguistics* 10, 615–628. CSLI, Stanford.

Kim, Shin-Sook & Wolfgang Sternefeld (1997): Superiority vs. Crossover. Ms., Universität Tübingen.

Kiparsky, Paul (1973): 'Elsewhere' in Phonology. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle*. Academic Press, New York, pp. 93–106.

Kiparsky, Paul & Carol Kiparsky (1970): Fact. In: D. Steinberg & L. Jakobovits, eds., *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge University Press, Cambridge, pp. 345–369.

Kiss, Katalin (1986): Why Noun Complement Clauses are Barriers. In: J. Mascaró & M. Nespor, eds., *Grammar in Progress*. Foris, Dordrecht, pp. 265–277.

Kitahara, Hisatsugu (1993): Deducing 'Superiority' Effects from the Shortest Chain Requirement, *Harvard Working Papers in Linguistics* 3, 109–119.

Kitahara, Hisatsugu (1994): A Minimalist Analysis of Cross-Linguistically Variant CED Phenomena. In: *Proceedings of NELS*. Vol. 24, Amherst: GSLA, pp. 241–253.

Kitahara, Hisatsugu (1997): *Elementary Operations and Optimal Derivations*. MIT Press, Cambridge, Mass.

Kitahara, Hisatsugu (1999): Eliminating * as a Feature (of Traces). In: S. Epstein & N. Hornstein, eds., *Working Minimalism*. MIT Press, Cambridge, Mass., pp. 77–93.

Kiziak, Tanja (2007): Long Extraction or Parenthetical Insertion? Evidence from Judgement Studies. In: N. Dehé & Y. Kavalova, eds., *Parentheticals*. Benjamins, Amsterdam, pp. 121–144.

Kluender, Robert (2004): Are Subject Islands Subject to a Processing Account?. In: B. Schmeiser, V. Chand, A. Kelleher & A. Rodriguez, eds., *Proceedings of the 23th WCCFL*. Cascadilla Press, Somerville, Mass., pp. 101–125.

Kluender, Robert & Marta Kutas (1993): Subjacency as a Processing Phenomenon, *Language and Cognitive Processes* 8(4), 573–633.

Kobele, Greg (2006): Generating Copies. PhD thesis, UCLA, Los Angeles.

Kobele, Greg (2011): Deriving Reconstruction Asymmetries. In: A. Alexiadou, T. Kiss & G. Müller, eds., *Local Modelling of Non-Local Dependencies in Syntax*. Linguistische Arbeiten, De Gruyter, Berlin.

Koeneman, Olaf (2000): The Flexible Nature of Verb Movement. PhD thesis, Utrecht

University.

Koizumi, Masatoshi (1995): Phrase Structure in Minimalist Syntax. PhD thesis, MIT, Cambridge, Mass.

Koopman, Hilda (1995): On Verbs that Fail to Undergo V-Second, *Linguistic Inquiry* 26, 137–163.

Koopman, Hilda & Dominique Sportiche (1986): A Note on Long Extraction in Vata and the ECP, *Natural Language and Linguistic Theory* 4, 357–374.

Koster, Jan (1978): *Locality Principles in Syntax*. Foris, Dordrecht.

Koster, Jan (1986): The Relation Between Pro-Drop, Scrambling, and Verb Movements. Ms., Rijksuniversiteit Groningen.

Koster, Jan (1987): *Domains and Dynasties*. Foris, Dordrecht.

Koster, Jan (2000): Variable-Free Grammar. Ms., University of Groningen.

Kramer, Ruth (2007): The Amharic Definite Marker and the Syntax/PF Interface. Ms., University of California at Santa Cruz.

Kratzer, Angelika (1996): Severing the External Argument from its Verb. In: J. Rooryck & L. Zaring, eds., *Phrase Structure and the Lexicon*. Kluwer, Dordrecht, pp. 109–137.

Kroch, Anthony (1974): The Semantics of Scope in English. PhD thesis, MIT, Cambridge, Mass.

Kroch, Anthony (1989): Asymmetries in Long Distance Extraction in a Tree Adjoining Grammar. In: M. Baltin & A. Kroch, eds., *Alternative Conceptions of Phrase Structure*. University of Chicago Press, Chicago, pp. 66–98.

Kuno, Susumo (1973): Constraints on Internal Clauses and Sentential Subjects, *Linguistic Inquiry* 4, 363–385.

Kučerová, Ivona (2006): An Anti-Intervention Effect in Czech Splits: An Argument for Late Merge. In: R. Compton, M. Goledzinowska & U. Savchenko, eds., *Proceedings of FASL 15*. Michigan Slavic Publications, Ann Arbor, pp. 161–179.

Lahne, Antje (2007): A Multiple Specifier Approach to Left Peripheral Architecture. Ms., Universität Leipzig. To appear in *Linguistic Analysis*.

Lahne, Antje (2008): Specificity-driven Syntactic Derivation: A New View on Long-distance Agreement. Ms., Universität Leipzig.

Lahne, Antje (2009): Where There is Fire There is Smoke. Local Modelling of Successive-Cyclic Movement. PhD thesis, Universität Leipzig.

Lakoff, George (1971): On Generative Semantics. In: D. Steinberg & L. Jakobovits, eds., *Semantics*. Cambridge University Press, Cambridge, pp. 232–296.

Landau, Idan (2000): *Elements of Control*. Kluwer, Dordrecht.

Larson, Richard (1988): On the Double Object Construction, *Linguistic Inquiry* 19, 335–391.

Lasnik, Howard (1999): On Feature Strength: Three Minimalist Approaches to Overt Movement, *Linguistic Inquiry* 30, 197–217.

Lasnik, Howard & Mamoru Saito (1984): On the Nature of Proper Government, *Linguistic Inquiry* 15, 235–289.

Lasnik, Howard & Mamoru Saito (1992): *Move α*. MIT Press, Cambridge, Mass.

Lasnik, Howard & Myung-Kwan Park (2003): The EPP and the Subject Condition Under Sluicing, *Linguistic Inquiry* 34, 649–660.

Lebeaux, David (1988): Language Acquisition and the Form of the Grammar. PhD thesis, University of Massachusetts, Amherst.

Lechner, Winfried (2004): Extending and Reducing the MLC. In: A. Stepanov, G. Fanselow & R. Vogel, eds., *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 205–240.

Legate, Julie Anne (2005): Phases and Cyclic Agreement, *MITWPL* 49, 147–156. Perspectives on Phases.

Legendre, Géraldine, Paul Smolensky & Colin Wilson (1998): When is Less More? Faithfulness and Minimal Links in Wh-Chains. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 249–289.

Lenerz, Jürgen (1984): *Syntaktischer Wandel und Grammatiktheorie*. Niemeyer, Tübingen.

Leu, Thomas (2008): *What for* Internally, *Syntax* 11, 1–25.

Levine, Robert & Ivan Sag (2003): Wh-Nonmovement. Ms., Stanford University. To appear in *Gengo Kenkyu*.

Lewis, David (1972): General Semantics. In: D. Davidson & G. Harman, eds., *Semantics of Natural Language*. Reidel, Dordrecht, pp. 169–218.

Lindauer, Thomas (1995): *Genitivattribute: Eine morphosyntaktische Untersuchung zum deutschen DP/NP-System*. Niemeyer, Tübingen.

Longobardi, Giuseppe (1985): Connectedness and Island Constraints. In: J. Guéron, H.-G. Obenauer & J.-Y. Pollock, eds., *Grammatical Representation*. Foris, Dordrecht, pp. 169–185.

Lumsden, John (1992): Underspecification in Grammatical and Natural Gender, *Linguistic Inquiry* 23, 469–486.

Lutz, Uli (2001): Studien zu Extraktion und Projektion im Deutschen. PhD thesis, Universität Tübingen.

Mahajan, Anoop (2001): Word Order and Remnant VP Movement. UCLA, Ms.

Maling, Joan & Annie Zaenen (1982): A Phrase Structure Account of Scandinavian Extraction Phenomena. In: P. Jacobson & G. Pullum, eds., *The Nature of Syntactic Representation*. Reidel, Dordrecht, pp. 229–282.

Manzini, Rita (1983): On Control and Control Theory, *Linguistic Inquiry* 14, 421–446.

Manzini, Rita (1992): *Locality. A Theory and Some of Its Empirical Consequences*. MIT Press, Cambridge, Mass.

Manzini, Rita (2009): Minimality Effects Without Head Movement. Ms., University of Florence.

Marušič, Franc (2005): On Non-Simultaneous Phases. PhD thesis, Stony Brook University.

Matushansky, Ora (2005): Going Through a Phase, *MITWPL* 49, 157–181. Perspectives on Phases.

Matushansky, Ora (2006): Head-Movement in Linguistic Theory, *Linguistic Inquiry* 37, 69–109.

McCarthy, John (1999): Sympathy and Phonological Opacity, *Phonology* 16:3, 331–399.

McCloskey, James (1979): *Transformational Syntax and Model Theoretic Semantics. A Case Study in Modern Irish*. Reidel, Dordrecht.

McCloskey, James (2002): Resumptives, Successive Cyclicity, and the Locality of Operations. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 184–226.

Medeiros, David (2008): Optimal Growth in Phrase Structure, *Biolinguistics* 2, 152–195.

Meinunger, André (2004): Verb Position, Verbal Mood and the Anchoring (Potential) of Sentences. In: H. Lohnstein & S. Trissler, eds., *The Syntax and Semantics of the Left Periphery*. Mouton de Gruyter, Berlin, pp. 313–341.

Merchant, Jason (2001): *The Syntax of Silence - Sluicing, Islands, and the Theory of Ellipsis*. Oxford University Press, Oxford.

Merchant, Jason (2004): Fragments and Ellipsis, *Linguistics and Philosophy* 27, 661–738.

Moro, Andrea (1997): *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge University Press, Cambridge.

Müller, Gereon (1989): Barrieren und Inkorporation. Master's thesis, Universität Konstanz.

Müller, Gereon (1995): *A-bar Syntax*. Mouton/de Gruyter, Berlin.

Müller, Gereon (1998): *Incomplete Category Fronting*. Kluwer, Dordrecht.

Müller, Gereon (1999): Imperfect Checking, *The Linguistic Review* 16, 359–404.

Müller, Gereon (2000): *Elemente der optimalitätstheoretischen Syntax*. Stauffenburg, Tübingen.

Müller, Gereon (2001): Order Preservation, Parallel Movement, and the Emergence of the Unmarked. In: G. Legendre, J. Grimshaw & S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 279–313.

Müller, Gereon (2002): Free Word Order, Morphological Case, and Sympathy Theory. In: G. Fanselow & C. Féry, eds., *Resolving Conflicts in Grammars: Optimality Theory in Syntax, Morphology, and Phonology*. Buske, Hamburg, pp. 9–48. Special issue of *Linguistische Berichte*.

Müller, Gereon (2003): Local vs. Global Optimization in Syntax: A Case Study. In: J. Spenader, A. Eriksson & Ö. Dahl, eds., *Variation within Optimality Theory. Proceedings of the Stockholm Workshop*. Stockholm University, Department of Linguistics, pp. 82–91.

Müller, Gereon (2004a): Phrase Impenetrability and Wh-Intervention. In: A. Stepanov, G. Fanselow & R. Vogel, eds., *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 289–325.

Müller, Gereon (2004b): Verb-Second as vP-First, *Journal of Comparative Germanic Linguistics* 7, 179–234.

Müller, Gereon (2006): Pro-Drop and Impoverishment. In: P. Brandt & E. Fuß, eds., *Form, Structure, and Grammar. A Festschrift Presented to Günther Grewendorf on Occasion of his 60th Birthday*. Akademie Verlag, Berlin, pp. 93–115.

Müller, Gereon (2007): Towards a Relativized Concept of Cyclic Linearization. In: U. Sauerland & H.-M. Gärtner, eds., *Interfaces + Recursion = Language?*. Mouton de Gruyter, Berlin, pp. 61–114.

Müller, Gereon (2008): *Inflectional Morphology in a Minimalist Grammar*. Lecture notes, Universität Leipzig. (www.uni-leipzig.de/∼muellerg/mu230a.pdf).

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. In: K. K. Grohmann, ed., *Explorations of Phase Theory. Features and Arguments*. Mouton de Gruyter, Berlin, pp. 269–308.

Müller, Gereon (2010a): Movement from Verb-Second Clauses Revisited. In: T. Hanneforth & G. Fanselow, eds., *Language and Logos. Festschrift for Peter Staudacher on his 70th Birthday*. Akademieverlag, Berlin.

Müller, Gereon (2010b): On Deriving CED Effects from the PIC, *Linguistic Inquiry* 41, 35–82.

Müller, Gereon & Wolfgang Sternefeld (1993): Improper Movement and Unambiguous Binding, *Linguistic Inquiry* 24, 461–507.

Müller, Gereon & Wolfgang Sternefeld (1996): A-bar Chain Formation and Economy of Derivation, *Linguistic Inquiry* 27, 480–511.

Müller, Gereon & Wolfgang Sternefeld (2001): The Rise of Competition in Syntax: A Synopsis. In: G. Müller & W. Sternefeld, eds., *Competition in Syntax*. Mouton/de Gruyter, Berlin, pp. 1–68.

Müller, Stefan (2005): Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung, *Linguistische Berichte* 203, 297–330.

Müller, Stefan (2007): *Head-Driven Phrase Structure Grammar: Eine Einführung*. Stauffenburg, Tübingen.

Nakamura, Masanori (1998): Reference Set, Minimal Link Condition, and Parametrization. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 291–313.

Neeleman, Ad & Chriszta Szendrői (2006): Radical Pro-Drop and the Morphology of Pronouns, *Linguistic Inquiry* 38, 671–714.

Neeleman, Ad & Hans van de Koot (2010): A Local Encoding of Syntactic Dependencies and its Consequences for the Theory of Movement, *Syntax* 13, 331–372.

Nevins, Andrew (2004): Derivations Without the Activity Condition. In: M. McGinnis & N. Richards, eds., *Proceedings of the EPP/Phase Workshop*. MITWPL, Cambridge, Mass.

Nissenbaum, Jon (2000a): Covert Movement and Parasitic Gaps. In: M. Hirotani, A. Coetzee, N. Hall & J.-Y. Kim, eds., *Proceedings of NELS 30*. GLSA, Amherst, Mass, pp. 542–555.

Nissenbaum, Jon (2000b): Investigations of Covert Phrase Movement. PhD thesis, MIT, Cambridge, Mass.

Noonan, Máire (1988): Superiority Effects: How do Antecedent Government, Lexical Government and V/2 Interact?, *McGill Working Papers in Linguistics* pp. 192–214.

Nunes, Jairo (2004): *Linearization of Chains and Sideward Movement*. MIT Press, Cambridge, Mass.

Nunes, Jairo & Juan Uriagereka (2000): Cyclicity and Extraction Domains, *Syntax* 3, 20–43.

Ott, Dennis (2011a): Local Instability: The Syntax of Split Topics. PhD thesis, Harvard University.

Ott, Dennis (2011b): A Note on Free Relative Clauses in the Theory of Phases, *Linguistic Inquiry* 42, 183–192.

Partee, Barbara, Alice ter Meulen & Robert Wall (1993): *Mathematical Methods in Linguistics*. Kluwer, Dordrecht.

Perlmutter, David & Scott Soames (1979): *Syntactic Argumentation and the Structure of English*. The University of California Press, Berkeley.

Pesetsky, David (1982): Paths and Categories. PhD thesis, MIT, Cambridge, Mass.

Pesetsky, David (1987): Wh-in-Situ: Movement and Unselective Binding. In: E. Reuland & A. ter Meulen, eds., *The Representation of (In)Definiteness*. MIT Press, Cambridge, Mass, pp. 98–129.

Pesetsky, David (1989): Language-Particular Processes and the Earliness Principle. Ms., MIT, Cambridge, Mass.

Pesetsky, David (1998): Some Optimality Principles of Sentence Pronunciation. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 337–383.

Pesetsky, David (2000): *Phrasal Movement and Its Kin*. MIT Press, Cambridge, Mass.

Pesetsky, David & Esther Torrego (2001): T-to-C Movement: Causes and Consequences. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 355–426.

Pesetsky, David & Esther Torrego (2006): Probes, Goals and Syntactic Categories.

Ms., MIT. Lingbuzz/000321.

Pesetsky, David & Esther Torrego (2007): The syntax of valuation and the interpretability of features. In: S. Karimi, V. Samiian & W. K. Wilkins, eds., *Phrasal and clausal architecture: Syntactic derivation and interpretation*. Benjamins, Amsterdam, pp. 262–294.

Polinsky, Maria (2003): Non-Canonical Agreement is Canonical, *Transactions of the Philological Society* 101, 279–312.

Polinsky, Maria & Eric Potsdam (2001): Long-Distance Agreement and Topic in Tsez, *Natural Language and Linguistic Theory* 19, 583–646.

Pollard, Carl J. & Ivan A. Sag (1994): *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Pollock, Jean-Yves (1989): Verb Movement, Universal Grammar, and the Structure of IP, *Linguistic Inquiry* 20, 365–424.

Postal, Paul (1972): Some Rules That Are Not Successive Cyclic, *Linguistic Inquiry* pp. 211–222.

Prince, Alan & Paul Smolensky (2004): *Optimality Theory. Constraint Interaction in Generative Grammar*. Blackwell, Oxford.

Pullum, Geoffrey (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.

Rackowski, Andrea & Norvin Richards (2005): Phase Edge and Extraction, *Linguistic Inquiry* 36, 565–599.

Ramchand, Gillian (2003): First Phase Syntax. Book ms., Oxford University.

Reinhart, Tanya (1983): *Anaphora and Semantic Interpretation*. Croom Helm, London.

Reinhart, Tanya & Eric Reuland (1993): Reflexivity, *Linguistic Inquiry* 24, 657–720.

Reis, Marga (1985): Satzeinleitende Strukturen im Deutschen. In: W. Abraham, ed., *Erklärende Syntax des Deutschen*. Narr, Tübingen, pp. 271–311.

Reis, Marga (1996): Extractions from Verb-Second Clauses in German?. In: U. Lutz & J. Pafel, eds., *On Extraction and Extraposition in German*. Benjamins, Amsterdam, pp. 45–88.

Reis, Marga & Inger Rosengren (1992): What do Wh-Imperatives Tell Us About Wh-Movement?, *Natural Language and Linguistic Theory* 10, 79–118.

Reuland, Eric (2001): Primitives of Binding, *Linguistic Inquiry* 32, 439–492.

Rezáč, Milan (2004): Elements of Cyclic Syntax: Agree and Merge. PhD thesis, University of Toronto.

Richards, Marc (2004): Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax. PhD thesis, University of Cambridge, Cambridge, UK.

Richards, Marc (2008): Defective Agree, Case Alternations, and the Prominence of Person. In: M. Richards & A. Malchukov, eds., *Scales*. Vol. 86 of *Linguistische*

Arbeitsberichte, Universität Leipzig, pp. 137–161.

Richards, Marc (2009): Movement-Related Morphology and Defective Phases: Germanic Participle Agreement Revisited. Ms., Universität Leipzig. (Talk at the Potsdam/Leipzig Workshop on Movement and Morphology, Leucorea, April 2009).

Richards, Marc (2010): Deletability of Edge Features. Ms., Universität Leipzig.

Richards, Marc (2011a): Deriving the Edge: What's in a Phase?, *Syntax* 14, 74–95.

Richards, Marc (2011b): Probing the Past: On Reconciling Long-Distance Agreement with PIC. In: A. Alexiadou, T. Kiss & G. Müller, eds., *Local Modelling of Non-Local Dependencies in Syntax*. Linguistische Arbeiten, De Gruyter, Berlin.

Richards, Norvin (1998): The Principle of Minimal Compliance, *Linguistic Inquiry* 29, 599–629.

Richards, Norvin (2001): *Movement in Language*. Oxford University Press, Oxford.

Riemsdijk, Henk van (1978): *A Case Study in Syntactic Markedness: The Binding Nature of Prepositional Phrases*. Foris, Dordrecht.

Riemsdijk, Henk van (1989): Movement and Regeneration. In: P. Benincá, ed., *Dialectal Variation and the Theory of Grammar*. Foris, Dordrecht, pp. 105–136.

Riemsdijk, Henk van & Edwin Williams (1986): *Introduction to the Theory of Grammar*. MIT Press, Cambridge, Mass.

Rizzi, Luigi (1982): *Issues in Italian Syntax*. Foris, Dordrecht.

Rizzi, Luigi (1986): Null Objects in Italian and the Theory of 'pro', *Linguistic Inquiry* 17, 501–557.

Rizzi, Luigi (1990): *Relativized Minimality*. MIT Press, Cambridge, Mass.

Rizzi, Luigi (1996): Residual Verb-Second and the Wh-Criterion. In: A. Belletti & L. Rizzi, eds., *Parameters and Functional Heads*. Oxford University Press, Oxford, pp. 63–89.

Rizzi, Luigi (1997): The Fine Structure of the Left Periphery. In: L. Haegeman, ed., *Elements of Grammar*. Kluwer, Dordrecht.

Rizzi, Luigi (2001): Relativized Minimality Effects. In: M. Baltin & C. Collins, eds., *The Handbook of Contemporary Syntactic Theory*. Blackwell, Oxford, pp. 89–110.

Rizzi, Luigi (2004): Locality and Left Periphery. In: L. Rizzi, ed., *The Structure of CP and IP. The Cartography of Syntactic Structures, vol. 2*. Oxford University Press.

Rizzi, Luigi (2006): Criterial Freezing, EPP, and Asymmetries. Ms., Siena University.

Rizzi, Luigi (2007): On Some Properties of Criterial Freezing, *CISCL Working Papers on Language and Cognition* 1, 145–158.

Roberts, Ian (1991): Excorporation and Minimality, *Linguistic Inquiry* 22, 209–218.

Roberts, Ian (1993): *Verbs and Diachronic Syntax*. Kluwer, Dordrecht.

Roberts, Ian (2001): Head Movement. In: M. Baltin & C. Collins, eds., *The Handbook of Contemporary Syntactic Theory*. Blackwell, Oxford, pp. 113–147.

Roberts, Ian (2009): On the Probable Existence of Head Movement and the Definite Non-Existence of the Head Movement Constraint. Ms., University of Cambridge.

Roberts, Ian (2010): *Agreement and Head Movement: Clitics, Incorporation and Defective Goals*. MIT Press, Cambridge, Mass.

Roberts, Ian & Anna Roussou (2002): The Extended Projection Principle as a Condition for the Tense-Dependency. In: P. Svenonius, ed., *Subjects, Expletives, and the EPP*. Benjamins, Amsterdam, pp. 125–156.

Rogers, James (2003): Syntactic Structures as Multi-Dimensional Trees, *Research on Language and Computation* 1, 265–305.

Rohrbacher, Bernhard (1999): *Morphology-Driven Syntax: A Theory of V to I Raising and Pro-Drop*. Benjamins, Amsterdam.

Ross, John (1967): Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, Mass.

Ross, John (1969): Guess Who?. In: R. Binnick, A. Davison, G. Green & J. Morgan, eds., *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society, Chicago, pp. 252–286.

Rouveret, Alain, ed. (2011): *Resumptive Pronouns at the Interfaces*. Vol. 5 of *Language Faculty and Beyond*, Benjamins, Amsterdam.

Rudin, Catherine (1988): On Multiple Questions and Multiple WH Fronting, *Natural Language and Linguistic Theory* 6, 445–501.

Sabel, Joachim (1996): *Restrukturierung und Lokalität*. Akademie-Verlag, Berlin.

Sabel, Joachim (2000): Partial Wh-Movement and the Typology of Wh-Questions. In: U. Lutz, G. Müller & A. von Stechow, eds., *Wh-Scope Marking*. Benjamins, Amsterdam, pp. 409–446.

Sabel, Joachim (2002): A Minimalist Analysis of Syntactic Islands, *The Linguistic Review* 19, 271–308.

Sabel, Joachim (2011): Derivationally Bound Anaphors and Pronouns at the C/I Interface. In: A. Alexiadou, T. Kiss & G. Müller, eds., *Local Modelling of Non-Local Dependencies in Syntax*. Linguistische Arbeiten, De Gruyter, Berlin.

Sag, Ivan & Thomas Wasow (1999): *Syntactic Theory. A Formal Introduction*. CSLI Publications, Stanford University.

Sag, Ivan, Philip Hofmeister & Neal Snider (2008): Processing Complexity in Subjacency Violations: The Complex Noun Phrase Constraint. Ms., Stanford University. To appear in *Proceedings of CLS*.

Sag, Ivan, Philip Hofmeister, Inbal Arnon, Neal Snider & Florian Jaeger (2008): Processing Accounts for Superiority Effects. Ms., Stanford University.

Salzmann, Martin (2006): Resumptive Prolepsis. A Study in Indirect A′-Dependencies. PhD thesis, Universiteit Leiden.

Sauerland, Uli (1995): Review of 'A-bar Syntax' by Gereon Müller, *Glot International* 1(8), 14–15.

Sauerland, Uli (1999): Erasability and Interpretation, *Syntax* 3, 161–188.

Schäfer, Florian (2007): On the Nature of Anticausative Morphology. PhD thesis,

Universität Stuttgart.

Schmellentin, Claudia (2006): *PP-Extraktionen. Eine Untersuchung zum Verhältnis von Grammatik und Pragmatik*. Niemeyer, Tübingen.

Schneider-Zioga, Patricia (2005): Wh-agreement as an intervention effect. Ms., University of California, Irvine.

Sells, Peter (1984): Syntax and Semantics of Resumptive Pronouns. PhD thesis, University of Massachusetts, Amherst.

Sells, Peter (2001): *Structure, Alignment and Optimality in Swedish*. Stanford Monographs in Linguistics, CSLI Publications, Palo Alto.

Sells, Peter (2006): Using Subsumption Rather than Equality in Functional Control. In: M. Butt & T. King, eds., *Proceedings of LFG-06*. CSLI Publications, Universität Konstanz.

Shin, Soo Song (1988): On Long-Distance Movement in German, *Linguistische Berichte* 115, 229–254.

Silverstein, Michael (1976): Hierarchy of Features and Ergativity. In: R. Dixon, ed., *Grammatical Categories in Australian Languages*. Australian Institute of Aboriginal Studies, Canberra, pp. 112–171.

Smolensky, Paul & Geraldine Legendre (2006): *The Harmonic Mind*. MIT Press, Cambridge, Mass.

Sportiche, Dominique (1989): Le Mouvement Syntaxique: Contraintes et Paramètres, *Langages* 95, 35–80.

Stabler, Edward (1996): Acquiring and Parsing Languages with Movement. Book ms., UCLA.

Stabler, Edward (1998): Acquiring Languages with Movement, *Syntax* 1, 72–97.

Starke, Michal (2001): Move Dissolves Into Merge: A Theory of Locality. PhD thesis, University of Geneva.

Starke, Michal (2006): Nanosyntax Class Lectures. Ms., University of Tromsø.

Starke, Michal (2009): Nanosyntax: A Short Primer to a New Approach to Language, *Nordlyd* 36, 1–6.

Staudacher, Peter (1987): Extraktion aus deutschen Verbz-Zweit-Komplementen. Ms., Universität Regensburg.

Staudacher, Peter (1990): Long Movement from Verb-Second-Complements in German. In: G. Grewendorf & W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 319–339.

Stechow, Arnim von (1992): Kompositionsprinzipien und grammatische Struktur. In: P. Suchsland, ed., *Biologische und soziale Grundlagen der Sprache*. Niemeyer, Tübingen, pp. 175–248.

Stechow, Arnim von (2000): Partial Wh-Movement, Scope Marking, and Transparent Logical Form. In: U. Lutz, G. Müller & A. von Stechow, eds., *Wh-Scope Marking*. Benjamins, Amsterdam, pp. 447–478.

Stechow, Arnim von (2005): LF in einem Phasenmodell. Ms., Universität Tübingen.

Stechow, Arnim von & Wolfgang Sternefeld (1988): *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, Opladen.

Stepanov, Artur (2007): The End of CED? Minimalism and Extraction Domains, *Syntax* 10, 80–126.

Sternefeld, Wolfgang (1985): Deutsch ohne grammatische Funktionen: Ein Beitrag zur Rektions- und Bindungstheorie, *Linguistische Berichte* 99, 394–439.

Sternefeld, Wolfgang (1989): V-Movement, Extraction from V/2 Clauses, and the ECP, *Working Papers in Scandinavian Syntax* 44, 119–140.

Sternefeld, Wolfgang (1991a): Chain Formation, Reanalysis, and the Economy of Levels. In: H. Haider & K. Netter, eds., *Derivation and Representation in the Theory of Grammar*. Kluwer, Dordrecht, pp. 75–143.

Sternefeld, Wolfgang (1991b): *Syntaktische Grenzen. Chomskys Barrierentheorie und ihre Weiterentwicklungen*. Westdeutscher Verlag, Opladen.

Sternefeld, Wolfgang (1992): Transformationstypologie und strukturelle Hierarchie. Ms., Universität Tübingen.

Sternefeld, Wolfgang (1996a): Comparing Reference Sets. In: C. Wilder, H.-M. Gärtner & M. Bierwisch, eds., *The Role of Economy Principles in Linguistic Theory*. Akademie Verlag, Berlin, pp. 81–114.

Sternefeld, Wolfgang (1996b): A Minimalist Semantics for Questions. Ms., Universität Tübingen.

Sternefeld, Wolfgang (2006): *Syntax*. Stauffenburg, Tübingen. Two volumes.

Stjepanović, Sandra & Shoichi Takahashi (2001): Eliminating the Phase Impenetrability Condition. Ms., Kanda University of International Studies.

Stowell, Tim (1981): Origins of Phrase Structure. PhD thesis, MIT, Cambridge, Mass.

Stowell, Tim (1985): Null Operators and the Theory of Proper Government. Ms. UCLA.

Svenonius, Peter (1994): C-Selection as Feature-Checking, *Studia Linguistica* 48, 133–155.

Svenonius, Peter (2001): Impersonal Passives: A Phase-Based Analysis. In: *Proceedings of the 18th Scandinavian Conference in Linguistics*. Travaux de l'Institut de Linguistique de Lund, Lund.

Svenonius, Peter (2004): On the Edge. In: D. Adger, C. de Cat & G. Tsoulas, eds., *Peripheries. Syntactic Edges and their Effects*. Kluwer, Dordrecht, pp. 261–287.

Szabolcsi, Anna (1994): The Noun Phrase. In: F. Kiefer, ed., *Syntax and Semantics, vol 27: The Structure of Hungarian*. Academic Press, San Diego, pp. 179–274.

Szabolcsi, Anna & Frans Zwarts (1993): Weak Islands and an Algebraic Semantics for Scope Taking, *Natural Language Semantics* 1, 235–285.

Szabolcsi, Anna & Marcel den Dikken (2003): Islands. In: L. Cheng & R. Sybesma, eds., *The Second GLOT International State-of-the-Article Book*. Mouton de

Gruyter, Berlin, pp. 213–240.

Takahashi, Daiko (1994): Minimality of Movement. PhD thesis, University of Connecticut.

Takano, Yuji (1994): Unbound Traces and Indeterminacy of Derivation. In: M. Nakamura, ed., *Current Topics in English and Japanese*. Hituzi Syobo, Tokyo, pp. 229–253.

Tappe, Thilo (1981): Wer glaubst du hat recht?. In: M. Kohrt & J. Lenerz, eds., *Sprache: Formen und Strukturen*. Niemeyer, Tübingen, pp. 203–212.

Taraldsen, Tarald (2008): Colonnata and the Syntax/Lexicon Interface. Ms., University of Tromsø. (Talk at WoTM 4, Großbothen, June 2008).

Thiersch, Craig (1978): Topics in German Syntax. PhD thesis, MIT, Cambridge, Mass.

Thiersch, Craig (1985): VP and Scrambling in the German Mittelfeld. Ms., University of Tilburg.

Torrence, Harold (2005): Asking Questions Quietly: The Syntax of Silent Wh-Words in Wolof. Ms., University of Kansas, Lawrence.

Travis, Lisa (1984): Parameters and Effects of Word Order Variation. PhD thesis, MIT, Cambridge, Mass.

Truswell, Robert (2007): Locality of Wh-Movement and the Individuation of Events. PhD thesis, University College London.

Unger, Christina (2010): A Computational Approach to the Syntax of Displacement and the Semantics of Scope. PhD thesis, Universiteit Utrecht, LOT.

Uriagereka, Juan (1999): Multiple Spell-Out. In: S. Epstein & N. Hornstein, eds., *Working Minimalism*. MIT Press, Cambridge, Mass., pp. 251–282.

Vicente, Luis (2007): The Syntax of Heads and Phrases. PhD thesis, LOT: Netherlands Graduate School of Linguistics, Universiteit Leiden.

Vikner, Sten (1990): Verb Movement and the Licensing of NP-positions in the Germanic Languages. PhD thesis, University of Geneva.

Vikner, Sten (1995): *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford University Press, New York and Oxford.

Vikner, Sten (1997): V-to-I Movement and Inflection for Person in All Tenses. In: L. Haegeman, ed., *The New Comparative Syntax*. Longman, London, pp. 189–213.

Vikner, Sten (2001a): V-to-I Movement and Do-Insertion in Optimality Theory. In: G. Legendre, J. Grimshaw & S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 424–464.

Vikner, Sten (2001b): Verb Movement Variation in Germanic and Optimality Theory. Habilitation thesis, Universität Tübingen.

Wagner, Michael (2008): Focus and Recursion. Ms., Cornell University.

Webelhuth, Gert (1992): *Principles and Parameters of Syntactic Saturation*. Oxford University Press, Oxford.

Wexler, Ken & Peter Culicover (1980): *Formal Principles of Language Acquisition*. MIT Press, Cambridge, Mass.

Williams, Edwin (1974): Rule Ordering in Syntax. PhD thesis, MIT, Cambridge, Mass.

Williams, Edwin (1994): *Thematic Structure in Syntax*. MIT Press, Cambridge, Mass.

Williams, Edwin (1997): Blocking and Anaphora, *Linguistic Inquiry* 28, 577–628.

Williams, Edwin (2003): *Representation Theory*. MIT Press, Cambridge, Mass.

Williams, Edwin (2006): The Subject-Predicate Theory of 'There', *Linguistic Inquiry* 37, 648–651.

Wiltschko, Martina (1997): Scrambling, D-linking and Superiority in German, *Groninger Arbeiten zur Germanistischen Linguistik* 41, 107–142.

Woolford, Ellen (2007): Aspect Splits as Contextual Faithfulness. Ms., University of Massachusetts, Amherst.

Wunderlich, Dieter (1996): Minimalist Morphology: The Role of Paradigms. In: G. Booij & J. van Marle, eds., *Yearbook of Morphology 1995*. Kluwer, Dordrecht, pp. 93–114.

Wunderlich, Dieter (1997a): Cause and the Structure of Verbs, *Linguistic Inquiry* 27, 27–68.

Wunderlich, Dieter (1997b): A Minimalist Model of Inflectional Morphology. In: C. Wilder, H.-M. Gärtner & M. Bierwisch, eds., *The Role of Economy Principles in Linguistic Theory*. Akademie Verlag, Berlin, pp. 267–298.

Wurmbrand, Susanne (2001): *Infinitives. Restructuring and Clause Structure*. Mouton de Gruyter, Berlin.

Wurmbrand, Susanne (2008): Word Order and Scope in German. Ms., University of Connecticut.

Zaenen, Annie & Jessie Pinkham (1976): The Discovery of Another Island, *Linguistic Inquiry* 7, 652–664.

Zeller, Jochen (2005): On Clitic Left Dislocation in Zulu, *Frankfurter Afrikanistische Blätter* 18.

Zwart, Jan-Wouter (1993): Dutch Syntax. A Minimalist Approach. PhD thesis, Rijksuniversiteit Groningen.

Index