# 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 [•F•] can only affect the category bearing the [F] feature that is closer to [•F•].

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 acyclic 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 effects 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 $\alpha$, a syntactic operation may not target a position that is included within another XP $\beta$ that is dominated by $\alpha$.

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. $\underbrace{\overbrace{[_{\text{YP}} \ldots \text{Y}}^{\text{SCC}}\ [_{\underline{\text{XP}}} \ldots [_{\text{X}'}\ \text{X}}_{\text{PIC}}\ [_{\text{WP}} \ldots \text{W}\ [_{\underline{\text{UP}}} \ldots \text{U} \ldots\ ]]]]]$

b. $\underbrace{\overbrace{[_{\underline{\text{ZP}}} \ldots \text{Z}}^{\text{SCC}}\ [_{\text{YP}} \ldots \text{Y}\ [_{\underline{\text{XP}}} \ldots [_{\text{X}'}\ \text{X}}_{\text{PIC}}\ [_{\text{WP}} \ldots \text{W}\ [_{\underline{\text{UP}}} \ldots \text{U} \ldots\ ]]]]]$

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 included in the minimal 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 3.2–3.4); 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-c) 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 3.5, 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 $\theta$-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 $\theta$- (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 selectional (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 specifier (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 $\alpha_n$. 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 $\alpha$ 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 selectional 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 domains 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 $\alpha$ can move if it is a phase (CP, vP, and DP are phases).
c. A goal $\alpha$ is the closest one to a probe if there is no distinct goal $\beta$ such that for some X (X a head or a maximal projection[11]), X c-commands $\alpha$ but not $\beta$.
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 $\alpha$ is a lexical category, $\alpha$ L-marks $\beta$ iff $\beta$ agrees with the head of $\gamma$ that is $\theta$-governed by $\alpha$.

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) [$_{\mathrm{CP}}$ Who do you [$_{\mathrm{vP}}$ think [$_{\mathrm{CP}}$ that we should [$_{\mathrm{vP}}$ 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 targetting it; thus, there is nothing that could stop further movement of the *wh*-phrase from the subject clause via the matrix Specv position, to the matrix SpecC position in (22).

---

[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.

(22)
a. [$C_{[+wh]}$ [[C [v *who* ]] v ]]
b. [$C_{[+wh]}$ [[C [$v_y$ *who*$_y$ ]] v ]] (Agree v-DP)
c. [$C_{[+wh]}$ [[C [*who*$_y$ $v_y$ ~~*who*$_y$~~ ]] v ]] (Move DP)
d. [$C_{[+wh]}$ [[C [*who* v ~~*who*~~ ]] v ]] (Forget Agree)
e. [$C_{[+wh]}$ [[$C_z$ [*who*$_z$ v ~~*who*~~ ]] v ]] (Agree C-DP)
f. [$C_{[+wh]}$ [[*who*$_z$ $C_z$ [~~*who*$_z$~~ v ~~*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 left edge of 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

[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.

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; 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.

(23) *Consequences of the LCA*:

a. A head precedes its complement ($\beta$).

b. A specifier ($\alpha$) must formally qualify as an adjunct. It is unique and precdes its head.

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

(24) $[_{\text{XP}}\ \alpha\ [_{\text{XP}}\ \text{X}\ \beta\ ]]$

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 $\alpha$ c-commands $\beta$, then $\alpha$ precedes $\beta$.

b. Induction step: If $\gamma$ precedes $\beta$ and $\gamma$ dominates $\alpha$, then $\alpha$ precedes $\beta$.

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*:

---

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

In any well-formed constituent structure tree, for any nodes x and y, if x precedes 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 incompat-

[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.

ible with yet another basic tenet of the phase-based approach to syntax developed in 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

---

[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.

attaching it to vP, yielding (31). In this derivation, the constraint that a minimal, bare $X^0$ 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 prob-

[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.

lematic in view of languages like German, where certain kinds of undisputable non-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 distinguishes 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, acyclic 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 movement applies in the first case, and not accessible *yet* when movement applies in the second. 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 that such a radically

---

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

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), Rizzi (2006), 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 $\alpha_1$ that are distinct from and do not contain any $\alpha_i$.
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 $\Gamma$.

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 $\alpha_i$.

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 ]

Aduncts (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 targetted 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 fully a 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 monograph). First, in the guise of Shortest Move, the analysis involves a trans-derivational constraint. This is incompatible with the meta-requirement formulated in (11-a) (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 the 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 make 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 involing 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 ] neudeřila [$_{DP_3}$ žádná t$_1$ ] Petra$_2$
girl$_{nom}$ hit no$_{nom}$ Petr$_{acc}$
'No girl hit Petr.'

b. [$_{DP_1}$ Holka ] neudeřila 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 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 approach 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 grammar, 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 *perfomance*, 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 varying degrees of experimental backing by behavioural and neuro-imaging 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 Structure 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), Maling & Zaenen (1982), and Gazdar, Klein, Pullum & Sag (1985)), the displacement property of natural languages is not modelled by movment transformations. Rather, displacement is accounted for by postulating SLASH features – category-valued features 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*:
$\text{X} \rightarrow W, \text{XP} \Rightarrow \text{X} \rightarrow W, \text{XP[+NULL]}$

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 is present 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 movemement, extra assumptions are required. For instance, it has been suggested that adjuncts may also enter a (generalized version of) subcategorization lists. For instance, 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 note 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.

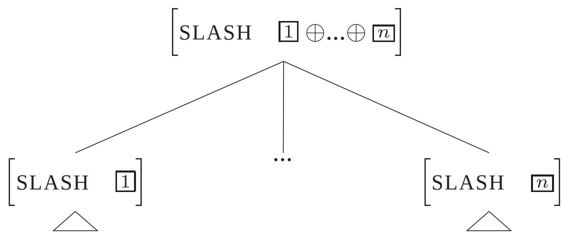
(60) *The lexical item 'loves' with a* SLASH *feature*:

$$\left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{HEAD} & \left[\textit{verb}\right] \\ \text{SPR} & \langle \boxed{3} \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{SLASH} & \langle \boxed{4} \rangle \end{array}\right] \\ \text{ARG-ST} & \left\langle \boxed{3}\ \overset{NP_i}{\left[\text{AGR}\ \ \textit{3sing}\right]}, \boxed{4}NP_j \right\rangle \\ \text{SEM} & \left[\begin{array}{ll} \text{MODE} & \text{prop} \\ \text{INDEX} & \text{s} \\ \text{RESTR} & \left\langle \left[\begin{array}{ll} \text{RELN} & \text{love} \\ \text{LOVER} & \text{i} \\ \text{LOVED} & \text{j} \end{array}\right] \right\rangle \end{array}\right] \end{array}\right]$$

loves

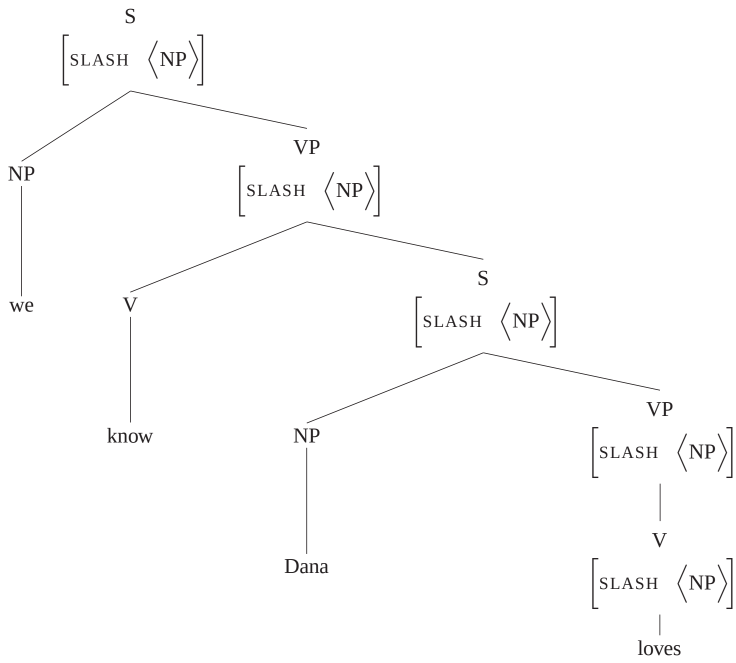
*Middle* Turning next to the middle, the role of the Foot Feature Principle in GPSG is taken over by the *Gap Principle* in Sag & Wasow (1999), see (61).

(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:




Here is an example of how the Gap Principle works.

(62)



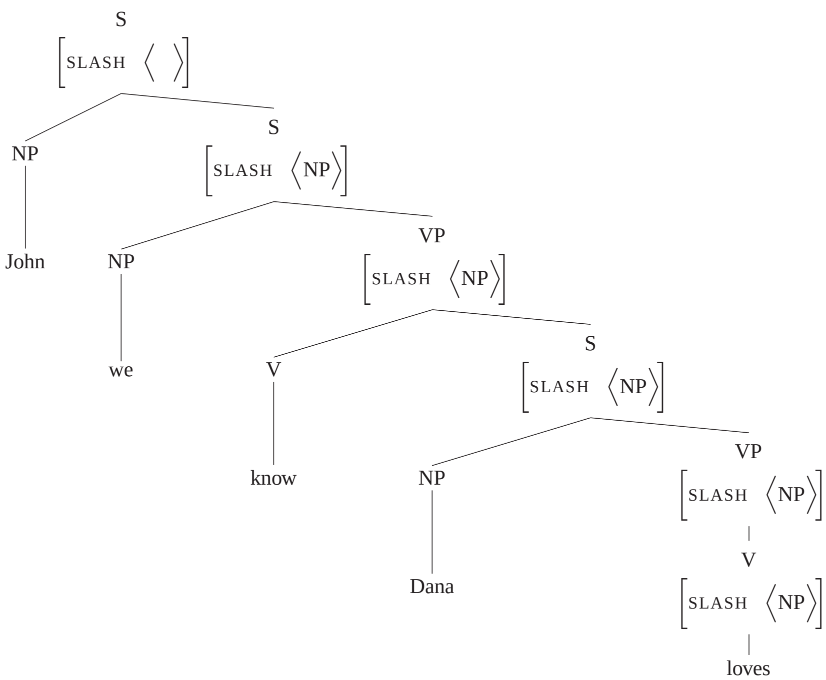

*Top* As in GPSG, there is a rule that expands a clausal category to a filler and a slashed head daughter. Sag & Wasow (1999) assume the Head Filler Rule in (63).

(63) *Head Filler Rule*:

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

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 constructions 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 covered (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 constraints 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 Nonfinal 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):
$\left[\text{INHER} \mid \text{SLASH} \quad \textit{empty-set}\right] < \left[\text{INHER} \mid \text{SLASH} \quad \textit{nonempty-set}\right]$

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 approach to locality constraints on displacement to a performance-based approach, grounded in the hypothesis that restrictions on the human sentence processor can derive 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 processing, pragmatic, and discourse effects that can be manipulated in such a way as to ameliorate the severe unacceptability exhibited by [examples involving 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 argument 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 that were 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).