LINGUISTISCHE ARBEITS BERICHTE

Syntactic Buffers

A Local–Derivational Approach to Improper Movement,
Remnant Movement, and Resumptive Movement

Gereon Müller
Universität Leipzig

91

Institut für Linguistik
Universität Leipzig
2014

Contents

Chapter 1

Introduction

1. Overview

The present monograph sets out to achieve a fairly modest goal: It tries to solve in a systematic way a problem which arises in a strictly local, derivational (phase-based) approach to syntax in cases where it looks as though information must be available in a certain domain that should not be available at this point,

* Research for this monograph was supported by a DFG grant for the project “Local Modelling of Non-Local Dependencies in Syntax” (MU 1444/8-1, Fabian Heck & Gereon Müller, principal investigators). For comments, discussion, and other kinds of help with the present work, I am grateful to Klaus Abels, David Adger, Anke Assmann, Josef Bayer, Rajesh Bhatt, Petr Biskup, Aaron Doliana, Robert Frank, Doreen Georgi, Günther Grewendorf, Jeremy Hartmann, Fabian Heck, Kyle Johnson, Stefan Keine, Tibor Kiss, Timo Klein, Hans van de Koot, Antje Lahne, Stefan Müller, Andreas Pankau, Tom Roeper, Martin Salzmann, Halldor Sigurðsson, Peggy Speas, Yuji Takano, Jochen Trommer, Philipp Weisser, Edwin Williams, and Joanna Zaleska, as well as to audiences at Universität Konstanz (Workshop on Structure-Building, April 2012), at Universität Leipzig (April 2012; October 2013; April 2014), at Universität Stuttgart (July 2012), at Universität Frankfurt/Main (Workshop on Remnant Movement, June 2013), at FU Berlin (Workshop on Progress in Linguistics, August 2013), and at the University of Massachusetts, Amherst (March 2014). An earlier, shorter version of chapter 2 appeared in *Lingua* (2014); a much shorter version of the material in chapter 3 will appear in the proceedings of the Frankfurt/Main Workshop on Remnant Movement; and an abridged version of chapter 4 focussing solely on German appeared in a special volume of *Linguistische Berichte* (2014). I would also like to thank the anonymous reviewers of these articles. This monograph was completed in its essentials while I was a visiting scholar (“syntax guru”) at the University of Massachusetts, Amherst, in March/April, 2014, and I am grateful to the UMass linguistics department (and to Messrs Johnson, Bhatt, and Cratoni in particular) for providing an excellent, inspiring working environment.

due to strict locality. More specifically, I will look at instances of movement in syntax where in order to determine whether displacement is possible at a given stage of the derivation, information must be taken into account that shows up at an earlier stage and should not be accessible anymore. I will argue that this problem arises with three types of movement (improper movement, remnant movement, and resumptive movement) in German and other languages;[1] and I would like to suggest a unified solution that relies on the notion of a *buffer* associated with the moved item on which contextual information of earlier, more deeply embedded and, by now, inaccessible domains is temporarily stored. The information on a buffer of a moved item is minimal throughout: It merely consists of a list of symbols that changes throughout the derivation, and that never grows too big because symbols are not only added to the buffer; symbols are also constantly deleted from the buffer as the derivation proceeds.

The background of the present study is provided by phase-based minimalist syntax (see Chomsky (2001; 2008)), a local–derivational approach to grammar: Syntactic structures are generated bottom-up, by alternating operations like Merge, Move (or internal and external Merge), and Agree, and the accessible window of a derivation is quite small throughout – it is standardly assumed to be confined to the structure included in either the current phase or the edge (specifier) domain (plus the head) of the previous phrase, due to the Phase Impenetrability Condition (PIC); see Chomsky (2001).

(1) *Phase Impenetrability Condition* (PIC; Chomsky (2000; 2001)):
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.

There are two ways to conceive of the PIC. One is to view it as simply a locality constraint on a par with the Subjacency Condition (see Chomsky (1977)), the Condition on Extraction Domain (CED, see Huang (1982)), or Riemsdijk's (1978) Head Constraint (alternatively, Koster's (1978) Bounding Condition), which the PIC superficially resembles most (see Abels (2012a)). In that case, one might expect that its effects can sometimes be overridden (see,

[1] Throughout this monograph, "resumptive movement" does not refer to movement of a resumptive pronoun; rather, it stands for movement that is resumptive in the sense that it leaves a pronoun behind.

e.g., Kayne's (1982) Connectedness approach, or Richards's (1998) Principle of Minimal Compliance); that it might not hold for all syntactic items in the same way (see, e.g., argument/adjunct asymmetries as derived in Lasnik & Saito (1984) and Chomsky (1986), or the role of D-linking or referentiality (Pesetsky (1987), Cinque (1990)); or that it might be subject to cross-linguistic variation (as argued by Rizzi (1982) for the Subjacency Condition), and more deeply embedded material may become selectively available after all. The other possibility is to consider it a general restriction on the working of grammar that leaves no room for exceptions, parametrization, etc. This has always been Chomsky's view, and it underlies his hypothesis that the PIC is essentially functionally motivated, as a third factor principle that contributes to efficient computation. On this view, once a phase is completed, the material that is c-commanded by the phase head is "spelled out" (i.e., gone for good), and thereby becomes inaccessible for any further syntactic operations for very deep reasons. It is this latter view that I adopt throughout this monograph.

As a consequence, all long-distance dependencies – i.e., all dependencies spanning more than one (X' domain of a) phase – must be modelled locally. Thus, unbounded wh-movement must be assumed to be composed of a series of smaller movement steps to intermediate phase edges, and comparable local analyses postulating a decomposition of seemingly non-local syntactic operations into sequences of smaller steps must been given for other non-local phenomena, like long-distance reflexivization, non-local case assignment, and long-distance agreement. It seems fair to conclude that these kinds of analyses have generally been quite successful, in the sense that simple, elegant, and, most importantly, empirically well-supported analyses have been given for a variety of non-local phenomena in the literature (see Alexiadou et al. (2012) for a recent overview).[2]

[2] A question that arises in this context is what consequences the phase-based approach has for compositional semantic interpretation. Various possibilities arise. First, it could be that the PIC, while strictly blocking syntactic access, still permits semantic access to spelled-out domains. This would seem to be the most conservative assumption. Second, semantic interpretation may also have to apply cyclically, in parallel with the syntactic derivation; see Kobele (2006; 2012). Third, semantic interpretation may apply completely seperately, without recourse to the syntactic derivation (see Unger (2010) for the general viability of such an approach). In what follows, I will abstract away from this issue: There may well lurk serious problems for semantic interpretation here, but these are orthogonal to the genuinely *syntactic* problems raised by a phase-based ap-

However, there is a class of problems for a phase-based approach (more generally, any local–derivational syntactic approach) which has not received a general, unified account so far, but which must eventually be solved if such an approach is to prove viable in the long term. These problems are defined by syntactic configurations where it looks as though information of some syntactic domain A must be used in the current syntactic domain B even though A is not accessible at this point. There are in fact two subclasses of this kind of problem: First, the case may arise where it seems that information from a syntactic domain A must be used in a syntactic domain B even though A is *not accessible yet* because the material of A has not yet entered the derivation; this can be referred to as a *look-ahead* problem. Second, it may be the case that information from domain A must be used in domain B even though A is *not accessible anymore* because it is too deeply embedded (i.e., part of the c-command domain of a lower phase head which has undergone spell-out); stretching algorithm terminology a bit, this can be referred to as a *backtracking* problem. The look-ahead and backtracking problems are schematically depicted in (2-a) and (2-b), respectively, with B as the currently accessible syntactic domain.

(2) a. *Look-Ahead*:

$$\underbrace{[_{\text{XP}}\ldots[_{\text{X}'}\ \text{X}\ldots}_{\text{A}}\ [_{\text{YP}}\ldots[_{\text{Y}'}\ \text{Y}\ldots\ \underbrace{[_{\text{ZP}}\ldots[_{\text{Z}'}\ \text{Z}\ldots[_{\text{WP}}\ldots[_{\text{W}'}\ \text{W}}_{\text{B}}\ldots]]]]]]]]$$

b. *Backtracking*:

$$\underbrace{[_{\text{XP}}\ldots[_{\text{X}'}\ \text{X}\ldots[_{\text{YP}}\ldots[_{\text{Y}'}\ \text{Y}}_{\text{B}}\ \ldots[_{\text{ZP}}\ldots[_{\text{Z}'}\ \text{Z}\ \underbrace{\ldots[_{\text{WP}}\ldots[_{\text{W}'}\ \text{W}}_{\text{A}}\ldots]]]]]]]]$$

In what follows, I will be exclusively concerned with the subclass of problems falling under *backtracking*.[3] A precondition for any attempt to make information from an earlier, more deeply embedded domain A available in the current domain B in a local approach to syntax is that there is an operation that trans-

proach which I will be concerned with in the following chapters, and which I now turn to in the main text.

[3] See Heck & Müller (2000b) and Müller (2011) for two versions of an approach to the first problem – viz., look-ahead – as it arises with a local reformulation of Chomsky's (2001) statement that edge features triggering intermediate movement steps can be added to a phase head *only if this has an effect on output* (which is clearly an instance of look-ahead).

ports the information in a successive-cyclic way across the domains. At least for the purposes of the present investigation, I will follow Hornstein (2001; 2009) in assuming that the only such operation is syntactic *movement*.[4] Assuming this, one might think that this is basically all there is to say about the modelling of backtracking configurations as in (2-b) in a local–derivational approach: Some item α has moved from A to B, and this is why information from A becomes accessible in B. Some of the αs that will have to undergo movement might be somewhat more abstract, and thus require careful empirical motivation (see, e.g., Hornstein (2001), Boeckx, Hornstein & Nunes (2010) on control, or Fischer (2006) on reflexivization), but as such the locality problem would be solved by simple movement.

However, postulating movement alone will not suffice for the three construction types that I will consider in the present monograph, which have all proven recalcitrant from a local–derivational point of view: improper movement, remnant movement, and resumptive movement. The reason is that, although one can make a case that they all involve movement of some item α (which is fairly uncontroversial in the first case, well-motivated in the second case, and, as I will argue, unavoidable in a local approach in the third case) from an embedded domain A to the current domain B, the A-information that is needed in B is not in and of itself located on α (either inherently, as a lexical property, or as a consequence of Agree, via standard assumptions about feature valuation), but rather comes from the *syntactic context* of α in A. Thus, what must be accessed in B (in order to determine whether movement of α

---

[4] This excludes *cyclic Agree* analyses as an option. Cyclic Agree has been suggested for instances of long-distance agreement. On this view, what looks like long-distance agreement across a phase is actually to be decomposed into a series of shorter agreement steps, such that first, the embedded verb agrees with the embedded long-distance agreement-trigger, and eventually the matrix verb agrees with the embedded verb (see Butt (1995; 2008), Legate (2005), Keine (2008), Lahne (2012), and Preminger (2009)). A fundamental conceptual question raised by these analyses is that one and the same feature needs to act as a probe and a goal; this assumption invariably has implausible ramifications for the concept of Agree. The obvious alternative to cyclic Agree is movement. See Polinsky & Potsdam (2001), Polinsky (2003), and Chandra (2005; 2007) for movement-based approaches to long-distance agreement in which the agreement controller (i.e., the DP) moves to a position accessible by the matrix verb, thereby feeding agreement; and Börjesson & Müller (2014) for a movement-based approach to long-distance agreement in which the agreement target (i.e., the matrix verb) is base-generated in the embedded clause, bringing about agreement there, and moves, by reprojection, to the matrix clause only later, thereby counter-bleeding agreement.

is legitimate) is not, say, a case feature or a ϕ-feature or a movement-related feature (such as [wh]) of some DP α, but rather contextual information of the following type:[5]

(3) *Contextual information that may be needed on a moved item*:

- information that specifies what kinds of phase edges α has passed on its way from A to B
- information that specifies whether some item has moved out of α in A, and whether this latter item has already reached its criterial position in A
- information that specifies whether a copy has been made of α, and whether α has encountered a barrier on its way from A to B

I will argue in chapter 2 that the first kind of information from an earlier, embedded domain A in (3) must be accessible in a higher domain B in order to determine whether movement counts as improper or not; this will, inter alia, account for the difference between legitimate clause-bound scrambling to Specv and illegitimate long-distance scrambling to Specv in German; cf. (4-ab).

(4) a. dass das Buch$_1$ keiner t$_1$ liest
that the book$_{acc}$ no-one$_{nom}$ reads

b. *dass Karl das Buch$_1$ glaubt [$_{\text{CP}}$ dass keiner t$_1$ liest]
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$ reads

In chapter 3, I turn to remnant movement configurations (more generally, α-over-β configurations, where both α and β undergo movement), and I show that the second type of information from an earlier domain A in (3) must be available in the current domain B in order to distinguish between well-formed cases of remnant movement as in (5-a) and ill-formed cases exhibiting freezing effects as in (5-b).

(5) a. [$_{\text{VP}_2}$ t$_1$ Gelesen] hat das Buch$_1$ keiner t$_2$
read has the book no-one

[5] That said, in local–derivational approaches where case features or ϕ-features radically disappear on an item after checking or valuation (rather than, say, becoming inactive), such information will not be present on a moved item either; see Stabler (1996; 1997), Kobele (2006), Graf (2013), Deal (2014), among others.

b. *Was$_1$ denkst du [$_{\mathrm{VP}_2}$ t$_1$ gelesen] hat keiner t$_2$?
what think you read has no-one

Finally, in chapter 4 I address a resumption strategy in relative clauses in German, and I argue that the properties of constructions of the type in (6) (with the resumptive pronoun set in italics) can be accounted for in a local–derivational approach if (a) resumption involves movement (essentially because there is no other way to transport information from an embedded domain B to a higher domain A), and (b) the third type of information mentioned in (3) from a lower domain A is available in the current domain B.

(6) ein Buch [$_{\mathrm{CP}}$ Op$_1$ [$_{\mathrm{C}}$ wo] [$_{\mathrm{TP}}$ ich einen Mann getroffen habe [$_{\mathrm{CP}}$
a book where I a man$_{acc}$ met have
der *es*$_1$ gelesen hat]]]
who it read has

I would like to suggest that information of the type in (3) is placed on a buffer. Since movement is a precondition for transporting information from one domain to another, the buffer that temporarily stores (and, in many cases, subsequently gets rid of) earlier contextual information should plausibly be related to movement. I would like to propose that the locus of this storage is the movement-related feature of the moved item (e.g., [wh] for wh-phrases, [top] for moved topics, [rel] for relative operators); more precisely, the value of such a feature. Thus, syntactic buffers are queue (-like) lists that constantly change throughout the derivation but – and this is the single most important assumption underlying the present approach – must qualify as legitimate (essentially: respect the functional sequence of heads, f-seq) in criterial positions. The constraint demanding this (which I will call *Williams Cycle*, for reasons that will become clear in chapter 2) will be decisive in all three kinds of constructions; i.e., a *unified* local–derivational analysis of these three recalcitrant phenomena is possible.

As a general background to the following investigation, I will assume an approach to movement as it is developed in Müller (2011).[6] This approach is

[6] This is mainly to have a frame of reference; I firmly believe that most of what follows in chapters 2-4, and in particular the decisive concept of a syntactic buffer, can be upheld in other local–derivational, phase-based approaches.

characterized by four assumptions. First, all phrases are phases, not just vP and CP (plus possibly DP). On the one hand, I take it to be conceptually attractive if local domains are as small as possible, given the motivation for phases in terms of computational efficiency. On the other hand, as argued at length in Müller (2011), empirical evidence from the distribution of island effects can also be shown to support such a step. Nevertheless, one should bear in mind that reducing locality domains even further than originally envisaged under Chomsky's concept of phases, from CP/vP to all XPs, will perhaps amplify the problems raised for a local–derivational by improper movement, remnant movement, and resumption, but it will not signficantly change them. (For instance, if information from within V′ is not accessible in a higher clause, information from within the next higher v′ generally will also not be accessible.)

Second, I assume that all syntactic operations are driven by designated features: There are structure-building features (rendered as [•F•]) that trigger internal and external Merge, and there are probe features (rendered as [∗F∗]) that trigger Agree.

Third, intermediate movement steps to phase edges are brought about by edge features (rendered as [•X•]). Edge features are neither inherent properties of phase heads (as suggested in Chomsky (2008)), nor are they "flavoured", i.e., versions of the structure-building features in criterial positions (as proposed by McCloskey (2002), Abels (2012b), Georgi (2014), and Deal (2014)). Rather, they are category-neutral and movement type-neutral all-purpose structure-building features that can be generated on phase heads under certain conditions (to which I will turn in chapter 2; also see footnote 3).

Finally, I would like to contend that a strictly local–derivational approach forces the conclusion that there are no traces (no copies either, and also no occurrences in a multidominance approach), at least not as objects that syntactic constraints (like the Subjacency Condition or the Empty Category Principle, see Chomsky (1977; 1981), Lasnik & Saito (1984; 1992)) could refer to. Attributing such a role to traces (copies, occurrences) invariably presupposes a non-local approach: In a non-local approach, a constraint on traces (copies, occurrences) may lead to different results than postulating the respective constraint on the movement operation as such (because changes may have affected the context of the position from which movement takes place; in fact, this is how traces are motivated in work like Fiengo (1977) (based on the interaction of movement with the Proper Binding Condition), Freidin (1978) (based on the interaction of movement with the Wh-Island Condition), and Lasnik & Saito (1984) (based on the interaction of movement with the Empty Category Principle (ECP))); in a local approach adopting (something like) the PIC, such contextual differences cannot arise. This leaves, as the sole possible remaining motivation for postulating traces (copies, occurrences), principles of semantic interpretation (on which see footnote 2 above).[7]

Given these assumptions, and anticipating some of the reasonings in the following chapters, some possible buffers as they arise in syntactic derivations look as in (7). In (7-a), DP is a wh-phrase for which movement steps through VP and vP phase edges in the present clause, and through TP and CP phase edges of an embedded clause, have been recorded on its buffer (there may have been earlier movement steps, but these are not registered on the buffer anymore). In contrast, in (7-b), DP is a wh-phrase for which movement steps trough VP, vP, TP and CP phase edges (in that order) are recorded on its buffer (again, there may have been earlier steps). (7-c) illustrates a VP undergoing topicalization; the buffer indicates that it has moved through vP, TP and CP phase edges, and that another item that bears the index 1 has been extracted from it prior to its movement to Specv (so VP is a remnant category). Fourth, in (7-d) there is a relative operator from which a copy (ultimately, the resumptive pronoun) that bears index 2 has been split off before movement to VP and vP phase edges. Finally, (7-e) is a relative operator that has just been merged, and not yet subjected to any other kind of operation; it has an empty buffer. All these buffers are tolerated as such in the derivation, but as will become clear below, only the one in (7-b) (plus, vacuously and therefore irrelevantly, the one in (7-e)) would qualify as a legitimate syntactic object in a criterial position in which an inherent movement-inducing feature of a head (such as [•wh•] on C) is satisfied by the moved item.

(7) *Some buffers*

a. $\text{DP}_{[\text{wh}:\boxed{\text{vVCT}}]}$

b. $\text{DP}_{[\text{wh}:\boxed{\text{CTvV}}]}$

---

[7] All that said, traces will often be added in derivations and representations given in the following chapters, but this is purely expository purposes.

c. $\text{VP}_{[\text{top:}\boxed{\text{CTv1}}]}$
d. $\text{DP}_{[\text{rel:}\boxed{\text{vV}\bullet 2\bullet}]}$
e. $\text{DP}_{[\text{rel:}\boxed{}]}$

In a sense, then, the concept of syntactic buffers introduced in the present work can be viewed as being the opposite of the concept of SLASH-feature percolation proposed in Gazdar (1981; 1982); Gazdar et al. (1985): In the latter approach, properties of the moved item are registered on the syntactic context; in the present approach, properties of the syntactic context are registered on the moved item.

The material in this monograph is organized as follows. In the remainder of the introduction, I illustrate, and roughly motivate, local–derivational approaches to syntax, of which phase-based approaches are instances, and distinguish them from alternatives. After that, chapter 2 is devoted to improper movement, chapter 3 addresses remnant movement, and chapter 4 is concerned with resumptive movement.

2. Background

2.1. Two Dichotomies

Current theories of syntax can roughly be grouped into four classes according to two dichotomies: An approach can be *derivational* (such that syntactic structures are generated by structure-building and structure-manipulating operations) or *declarative* (such that syntactic structures are viewed as potential objects that have to comply with the grammatical constraints of a language); and an approach can be *local* (such that all syntactic dependencies are confined to local domains, e.g., subtrees, and syntactic constraints have access only to these subtrees) or *non-local* (such that syntactic dependencies can span larger portions of structure, potentially the whole sentence, and syntactic constraints also have access to these large domains).

A cross-classification yields four models of grammar. Current *local–derivational* approaches include most prominently the Minimalist Program (Chomsky (2001; 2008)), where syntactic structures are generated bottom-up, by alternating operations like Merge, Move, and Agree, and the accessible window of a derivation is quite small throughout (it is standardly assumed to be confined to the minimal phase).

Generalized Phrase Structure Grammar (GPSG) (Gazdar (1981), Gazdar et al. (1985)) and Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag (1994), Sag & Wasow (1999), Müller, St. (2007)) belong to the class of *local–declarative* approaches: Dependencies and constraints are evaluated in small subtrees (essentially, mother-daughter relations), and syntactic structures are assumed to be licensed by constraint satisfaction, rather than being generated successively by syntactic operations.

Next, there are *non-local–derivational* models, like the classical Principles-and-Parameters theory as developed in Chomsky (1981; 1986) and Chomsky & Lasnik (1993), or the classical transformational approach ('standard theory') developed in Chomsky (1965). In Principles-and-Parameters theory, syntactic structures are generated by generalized rules of the phrase-structure component (X-bar theory) and operations like Move-α or (even more generally) Affect-α (Lasnik & Saito (1984; 1992)); by assumption, syntactic dependencies can in principle span large domains (even though, e.g., movement is, in practice, required to stop in all intermediate clause-initial positions by locality constraints), and constraints can in principle access the entire structure of a given, arbitrarily large syntactic object.

Finally, Lexical Functional Grammar (LFG) (Bresnan (2001)) is an approach that is *non-local—declarative*: Syntactic structures are licensed (rather than generated), and dependencies and constraints that restrict them have access to the full structure (or, more precisely, structures, given the central assumption that sentences are simultaneously represented at several parallel levels of representation). Similarly, optimality-theoretic syntax is standardly taken to be a non-local–declarative model (an instance of "harmonic parallelism", in Prince & Smolensky's (2004) terms); see, e.g., Grimshaw (1997) and Legendre et al. (1998), among many others).

Against this background, the question arises what potential empirical and conceptual evidence for or against a particular model looks like.

2.2. Local vs. Non-Local Approaches

As for the difference between local and non-local approaches to syntax, empirical arguments for the former are provided by the existence of (morphological, syntactic or semantic) reflexes of seemingly non-local dependencies like movement (see, e.g., Sag & Wasow (1999), Fox (2000), Bouma et al. (2001), McCloskey (2002), Asudeh (2004), Huybregts (2009), Assmann et al.

(2010), Lahne (2009), Georgi (2014), and references cited in all these works). These reflexes of movement initially favour a local modelling of non-local dependencies because they suggest a partitioning of the structure affected by movement into subparts (either by assuming series of small, successive-cyclic movement steps, or by assuming SLASH features that are percolated locally), and the availability of the relevant information (viz., that some domain has been affected by movement) for other operations (e.g., insertion of special morphological exponents, application of certain syntactic processes, semantic reconstruction, etc.) Reflexes of movement can be captured in non-local approaches to movement dependencies like the standard LFG accounts (Bresnan (2001), Dalrymple (2001, ch. 14)), earlier accounts that envisage unbounded movement transformations (Chomsky (1965; 1975), Ross (1967), Bresnan (1976a;b)), or *path*-based approaches as they have been developed in Principles-and-Parameters theory as an alternative to successive-cyclic movement (Kayne (1982), Pesetsky (1985), Longobardi (1985), and Koster (1987)), but only at a cost.[8]

Further empirical arguments for a local approach to syntax are given in Heck & Müller (2000a; 2003) on the basis of syntactic repair operations. For instance, it is argued in Heck & Müller (2003) that multiple wh-movement in sluicing constructions in German, which otherwise only moves one wh-

[8] Dalrymple (2001, ch. 14) develops a non-local LFG approach to movement that addresses this issue. Here, non-local movement dependencies can be stated as identity relations between two grammatical functions; what qualifies as a legitimate identity relation in a non-local dependency is then encoded as a *regular expression* in the phrase-structure component. This analysis does not break up a movement dependency into smaller parts (as with successive-cyclic movement steps or SLASH feature percolation), and, as such, it does not imply any *record*, or *track-keeping device*, of a non-local movement dependency in the syntactic structure that shows up between the displaced item and its base position. However, in view of the existence of (morphological) reflexes of movement, Dalrymple (2001) proposes that a track-keeping device (a feature [LDD], for 'long-distance dependency') can be *added* to phrase structures after all, so as to provide a point of reference for a (morphological) reflex of movement. This solution may be viewed as satisfactory from a purely technical point of view; but it seems clear that it is inferior to local modellings of the phenomenon: The sole purpose of the [LDD] feature is to permit accounts of (morphological) reflexes of movement; the feature does not play any role in bringing about, or restricting, movement dependencies per se.
The same conclusions can be drawn for path-based approaches; as McCloskey (1988, 30) notes, here the movement path is "not formally marked in any way".

phrase to the specifier of an interrogative C, is a repair strategy that is only permitted as a last resort so as to satisfy recoverability; cf. (8-a) (with multiple wh-movement) vs. (8-b) (where only one wh-phase moves, as in standard interrogative clauses, and non-realization of the second wh-phrase as a result of sluicing – conceived of as TP deletion – leads to a fatal recoverability problem under the relevant reading).

(8) a. Irgend jemand hat irgend etwas geerbt aber Karl weiss nicht
someone has something inherited but Karl knows not
mehr [$_{\text{CP}}$ wer$_1$ was$_2$ [$_{\text{TP}}$ ~~t_1 t_2 geerbt~~ ~~hat~~]]
more who what inherited has

b. *Irgend jemand hat irgend etwas geerbt aber Karl weiss nicht
someone has something inherited but Karl knows not
mehr [$_{\text{CP}}$ wer$_1$ [$_{\text{TP}}$ ~~t_1 was$_2$ geerbt~~ ~~hat~~]]
more who what inherited has

Crucially, multiple sluicing as a repair operation can only apply if the wh-phrases are clause-mates: A wh-phrase in an embedded clause can never undergo successive-cyclic movement to a matrix clause that hosts another wh-phase (see (9-a)), even though successive-cyclic wh-movement is an option as such in (simple) sluicing constructions (see (9-b)).

(9) a. *Irgend jemand hat behauptet dass Maria irgend etwas geerbt
someone has claimed that Maria something inherited
hat aber Karl weiss nicht mehr [$_{\text{CP}_3}$ wer$_1$ was$_2$ [$_{\text{TP}}$ ~~t_1 behauptet~~
has but Karl knows not more who what claimed
~~hat~~ [$_{\text{CP}_4}$ ~~dass~~ [$_{\text{TP}}$ ~~Maria t_2 geerbt~~ ~~hat~~]]]]
has that Maria inherited has

b. Maria hat behauptet dass sie irgend etwas geerbt hat aber
Maria has claimed that she something inherited has but
Karl weiss nicht mehr [$_{\text{CP}_3}$ was$_1$ [$_{\text{TP}}$ ~~Maria behauptet hat~~ [$_{\text{CP}_4}$ ~~t'_1~~
Karl knows not more what Maria claimed has
~~dass sie t_1 geerbt~~ ~~hat~~]]]
that she inherited has

The unavailability of multiple sluicing with non-clause-bound wh-movement in German shows that the decision whether a recoverability-driven repair operation (i.e., multiple wh-movement) is possible or not cannot be made on the basis of information contained in the whole sentence, but must be made on the basis of local information that is accessible in the clause where the second wh-phrase is externally merged: Thus, in (9-a), in the embedded CP_4 where DP_2 is base-generated, the recoverability problem does not yet arise (because it is the TP of CP_3 that is subject to deletion), and given that look-ahead is not an option, the derivation fails to carry out intermediate movement steps with DP_2 in the same way in which it fails to do so in regular multiple wh-questions where the lower wh-phrase shows up in a separate clause; see (10-ab).

(10) a. Wer$_1$ hat t$_1$ behauptet [$_{CP}$ dass Maria was$_2$ geerbt hat] ?
who$_{nom}$ has claimed that Maria$_{nom}$ what$_{acc}$ inherited has

b. *Wer$_1$ was$_2$ hat t$_1$ behauptet [$_{CP}$ t$'_2$ dass Maria t$_2$ geerbt hat] ?
who$_{nom}$ what$_{acc}$ has claimed that Maria$_{nom}$ inherited has

Finally, when CP_3 is reached in (9-a) and the information is in fact present that would trigger multiple wh-movement (i.e., movement also of the embedded wh-phrase), such movement is blocked by locality (e.g., the PIC).

In addition, conceptual arguments for a local modelling of non-local dependencies have been advanced. These arguments typically center around notions like complexity (see Gazdar (1981) on preserving context-freeness by avoiding transformations as tools to model non-local dependencies, and Chomsky (2001; 2005; 2008) on introducing phases as a means of bringing about efficient computation).

Another conceptual argument for a local (as opposed to a non-local) modelling of non-local dependencies that is perhaps more straightforwardly relevant comes from learning theory (see Heck & Müller (2010)): In a local approach, the set of possible grammars that the language learner needs to consider is reduced (see, e.g., Chomsky (1972), Sternefeld (2006)). The argument goes as follows. Let T_1 be a theory according to which every grammar of a natural language obeys the constraint that a dependency may not cross more than one clause boundary. Next, let T_2 be a theory according to which arbitrarily many clause boundaries may be crossed by syntactic dependencies. If one compares T_1 and T_2, it turns out that, ceteris paribus, the set of possible grammars of T_2 is a superset of the set of possible grammars of T_1. The reason is that T_2 also (but, crucially, not exclusively) contains grammars that generate only dependencies which are more local in the sense that they cross at most one clause boundary. This consideration may suggest that a local reanalysis of non-local dependencies in syntax may push the development of the theory further into the direction of explanative adequacy.[9]

2.3. Derivational vs. Declarative Approaches

Turning next to the second fundamental distinction, viz., that between derivational and declarative approaches to syntax, it can be noted that central arguments for the former are typically based on *opacity* phenomena, i.e., *counter-bleeding* and *counter-feeding* (Chomsky (1951; 1975), Kiparsky (1973)). In derivational terms, *counter-feeding* describes rule interactions where a rule A would feed another rule B (i.e., create the context for B to apply) but does not actually do so because it applies too late (B's chance to apply has already passed): Thus, A counter-feeds B. *Counter-bleeding* captures rule interactions where a rule A would bleed another rule B (i.e., destroy the context in which B can apply) but, again, does not actually do so because it applies too late (B has already applied): A counter-bleeds B. From a derivational perspective, opacity is entirely unproblematic: Operations that induce counter-feeding or counter-bleeding apply late in the derivation (put another way, operations that are counter-fed and counter-bled apply early). However, from a declarative perspective that does without postulating sequences of operations, counter-feeding and counter-bleeding raise potential problems: Rule interaction in counter-feeding and counter-bleeding environments is opaque because it cannot be determined by simply looking at the respective output representations why some operation (or process) B has not applied even though its context of application would seem to be present (*counter-feeding*), and why some operation B has applied even though its context of application would not seem to be present (*counter-bleeding*). Thus, a declarative approach to syntax faces initial problems of underapplication and overapplication in the area of opaque interactions.

A classical instance of counter-feeding in syntax involves *wanna*-contraction in English. This operation is possible in control contexts but impossible in exceptional case marking (ECM) contexts; see (11-a) vs. (11-b).

---

[9] In this context, also compare Lightfoot (1994) on the hypothesis of 'degree-0 learnability' that restricts parameter learning to matrix clauses, hence, to local dependencies.

(11) a. Who do you want to/wanna meet ?
b. Who do you want to/*wanna meet Mary ?

A derivational analysis is straightforward (Bresnan (1972), Pullum (1979), Arregi & Nevins (2012)): There are two relevant operations, viz. (i) *wanna*-contraction (which contracts *want* and *to*, yielding *wanna*, under strict phonological adjacency) and (ii) wh-movement to the specifier of an interrogative C node. Suppose next that *wanna*-contraction necessarily precedes wh-movement.[10] On this view, in (11-a), *want* and *to* are adjacent to begin with; so *wanna*-contraction can apply without problems. In (11-b), however, the ECM-subject *who* initially shows up between *want* and *to*. Wh-movement of *who* would result in a feeding of *wanna*-contraction, but it doesn't because (by assumption) this operation applies too late. Thus, (11-b) instantiates *counter-feeding*.[11]

A standard argument for counter-bleeding in syntax involves reconstruction for Principle A of the binding theory. The core observation is that reflexive pronouns can undergo movement, and they can occur in phrases that undergo movement. After the movement operation, they should therefore not be able to satisfy Principle A of the binding theory since they are not bound by their antecedents anymore; but they can evidently satisfy this constraint in the relevant examples. Relevant examples that show this are given in (12-a) (with scrambling of the reflexive to a pre-subject position in German), (12-b) (with topicalization of a reflexive in English), and (12-c) (with topicalization of an

[10] Such an order would follow immediately if derivational order is governed by the Strict Cycle Condition (see page 30 below for a formulation), and cyclic domains are sufficiently small (as suggested by McCawley (1984; 1998); also see footnote 12 below): *Wanna*-contraction affects the matrix VP, but wh-movement of the subject of the infinitive affects the matrix CP, which is clearly more inclusive.

[11] Note that this reasoning presupposes that at least one of the following statements holds: (i) There is no structural subject in control infinitives. (ii) There is an empty subject in control infinitives (e.g., PRO, as in Chomsky (1981)), which does not block phonological adjacency. (iii) If there is an overt subject in control infinitives, and this is deleted by Equi NP deletion (or some other rule), then this latter rule must apply early, so that it can feed *wanna*-contraction. (iv) If the movement theory of control (MTC) is adopted (Hornstein (2001), Boeckx, Hornstein & Nunes (2010)), control movement to the matrix θ-position must take place early, so that it can feed *wanna*-contraction.

XP containing a reflexive in German).

(12) a. dass [$_{DP_1}$ sich] der Fritz$_1$ gestern im Spiegel gesehen hat
that REFL the Fritz yesterday in the mirror seen has
b. [$_{DP_1}$ Himself] John$_1$ does not really like
c. [$_{DP}$ Bücher über sich$_1$] hat er$_1$ keine gelesen
books about REFL has he none read

Principle A requires a reflexive pronoun to be bound within its minimal clause, the constraints triggering movement require the reflexive to show up in a displaced position outside the c-command domain of its binder, and the well-formedness of the examples in (12) can then be taken to indicate that Principle A can be checked, and satisfied, before subsequent movement would bleed this operation.[12]

As a third case exhibiting opaque interaction in syntax, consider the analysis of the prohibition against movement of ergative DPs given in Assmann et al. (2012). The generalization to be derived is that in many ergative systems, an ergative DP cannot undergo A-bar movement (wh-movement, topicaliztion, relativization, focus movement), in contrast to an absolutive DP. Crucially, no comparable restriction holds in accusative systems, where accusative and nominative DPs are both mobile in principle. This can be derived as follows. First, for locality reasons (the PIC, given that all phrases qualify as phases), ergative and accusative DPs must undergo an intermediate movement step to SpecT in the course of A-bar movement. Second, T assigns (absolutive/nominative) case to an internal argument DP in ergative systems, and to an external argument DP in accusative systems, as an instance of Agree; v assigns structural case to an external argument in ergative systems (ergative), and to an internal argument in accusative systems (accusative) (see Murasugi (1992)), again as an instance of Agree (Chomsky (2001)). Third, movement of some DP α to SpecT makes assignment of absolutive/nominative to the remaining DP β impossible because SpecT counts as closer to T than the vP-internal domain.[13] Thus, if an inter-

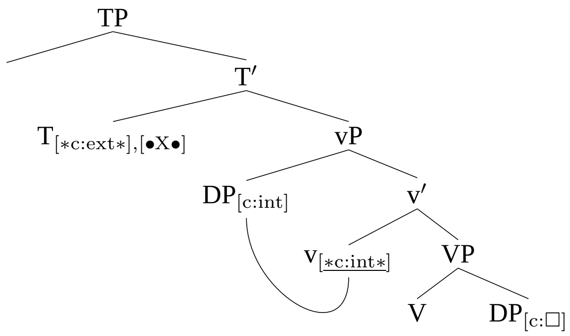
[12] Again, the required order would follow from the Strict Cycle Condition (with satisfaction of Principle A taking place in a smaller domain than movement), but eventually a bit more would have to be said, given that there is evidence that Principle A can be satisfied at any point in the derivation; see Barss (1986), Belletti & Rizzi (1988).

[13] This is a simplification as closeness actually does not play a role in this analysis. Rather, a

mediate movement step of α occurs *before* absolutive/nominative assignment by T, such assignment is blocked, and ungrammaticality results because β remains without structural case, as an instance of *bleeding*. In contrast, if an intermediate step of α occurs *after* absolutive/nominative assignment by T, it comes too late to create problems, and there is *counter-bleeding*. In Assmann et al. (2012), it is argued that there is independent evidence that ergative DPs move early (before Agree operations have taken place), whereas accusative DPs move late (after Agree operations have taken place; see Müller (2009)), yielding a bleeding effect in the first case, and a counter-bleeding effect in the second. This is shown in a bit more detail in the derivations in (13) (illegitimtae movement of an ergative DP to SpecT) and (14) (legitimate movement of an accusative DP to SpecT). Note that [c:int] is internal structural case assigned by v (ergative/accusative), whereas [c:ext] is external structural case assigned by T (absolutive/nominative); [c:□] is a case feature of a DP that is not yet valued.

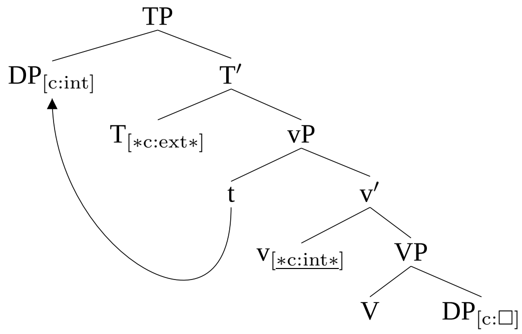
(13) *Illegitimate movement of ergative DPs*

a. Structure after T is merged

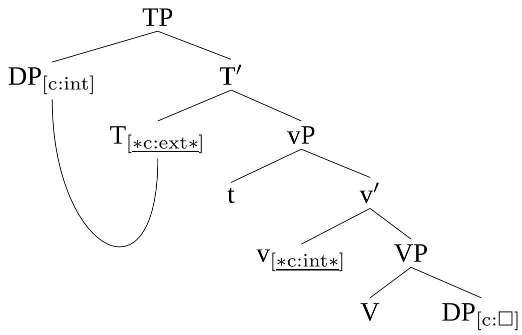


general Specifier-Head Bias is postulated according to which Agree with a specifier is, ceteris paribus, preferred to Agree with an item in the c-command domain; note that this assumption is exactly the opposite of what is stated in Béjar & Řezáč (2009).

b. Movement of DP_{erg} takes place before case assignment by T



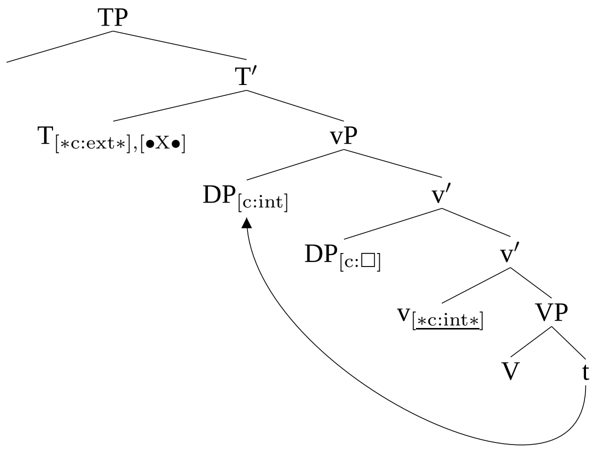
c. T's case feature is absorbed by the ergative DP



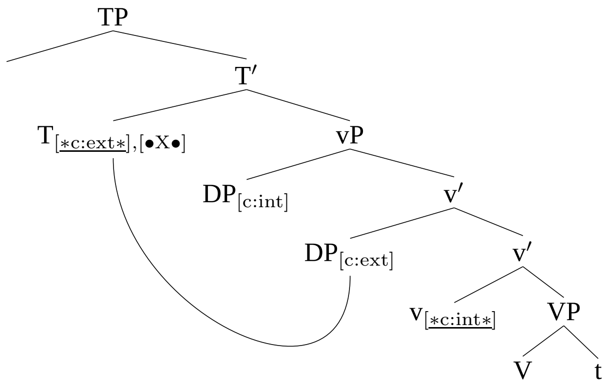
In the derivation in (13) one can see that an intermediate step of movement of an ergative DP "marauds" the case feature that T could otherwise assign to the internal argument DP; the latter remains without case, and ungrammaticality arises.

(14) *Legitimate movement of accusative DPs*

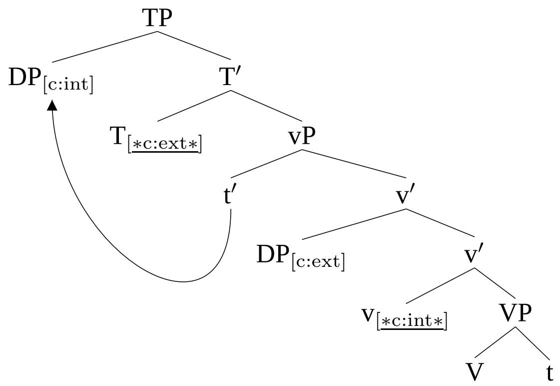
a. Structure after T is merged



b. Case feature assignment by T takes place before movement of the accusative DP



c. Movement of the accusative DP comes too late to bleed case-assignment to the external argument



Thus, if an accusative DP moves after Agree operations effected by T have taken place whereas an ergative DP moves before Agree operations effected by T have taken place, it is correctly predicted that there is a counter-bleeding effect in the former case, and a bleeding effect in the latter. Importantly, this difference cannot be detected by just looking at the output representations on the TP cycle (even if they are enriched with devices like traces): The accusative DP in SpecT in (14-c) *does* occupy the preferred position for case valuation with T, in the same way that the ergative DP in SpecT in (13-c) does. Consequently, (14) poses a serious problem from a declarative perspective: The resulting two TP structures in (13) and (14) *are identical* in the relevant respects, with an ergative/accusative DP in SpecT that should uniformly block case assignment by T to an item that still needs case but does so only in one case.

Some opaque interactions in syntax can be captured in declarative approaches by enriching representations of linguistic expressions with abstract material that encodes (what would otherwise qualify as) earlier derivational steps (e.g., by assuming traces/copies, and various type of empty pronouns, like pro and PRO), and by then postulating (more complex) constraints that refer to the enriched structure. Thus, counter-feeding of *wanna*-contraction by wh-movement can be captured representationally if movement leaves a trace (or copy) in the original position, and the *wanna*-contraction rule is formulated in a more abstract way, such that traces (copies) can block phonological adjacency (whereas an empty subject pronoun PRO in control infinitives for some reason cannot do so). However, in other cases, deriving opacity effects in syntax by enriching representations is not as straightforward because it does not suffice to postulate abstract elements; there also have to be *special constraints* or *mechanisms* that refer to them, and these typically do not qualify as simple and elegant, and thus lead to an inferior analysis. More specifically, the problem is that the *effects* of constraint interaction are integrated into the definition of a single constraint, which makes this constraint extremely implausible (see Grimshaw (1998), Chomsky (2001; 2008)). The counter-bleeding effect with reflexivization and movement is a case in point. As shown by Barss (1984), this effect can be derived in a declarative approach that does not rely on an order of syntactic operations by assuming a version of Principle A as in (15-a) (with binding replaced by chain-binding), and postulating the (doubly) disjunctive notion of chain-binding as in (15-b), which bascically undoes movement for the purposes of binding theory evaluation; thus, the effects of an interaction of operations (movement counter-bleeds reflexivization/Principle A) are not derived from the application of the operations themselves, but are stipulated as part of a more complex constraint.[14]

(15) a. *Principle A*:
At S-structure, an anaphor is chain-bound in its binding domain.

b. *Chain-Binding*:
α chain-binds β iff (a), (b), and (c) hold:
(i) α and β are co-indexed.
(ii) α occupies an A-position.
(iii) α c-commands β, or α c-commands a trace of γ, where γ = β or γ dominates β.

Concerning the third case discussed above – viz., the interaction of case assignment by T and movement of some case-bearing DP to SpecT –, the situation is even worse. It is completely unclear how the counter-bleeding effect with

[14] Barss (1986) ultimately adopts a concept of *chain accessibility sequences*, which extends the chain binding proposal sketched here; but this analysis suffers from the same shortcomings and is in fact even more complex.

nominative assignment by T in contexts with accusative movement to SpecT could be accounted for in a declarative approach by enriching representations and postulating more complex constraints.

These three examples may suffice as illustrations of what form opacity arguments for a derivational approach to syntax typically take. It should be emphasized that the arguments considered here are by no means conclusive: It is of course possible to come up with a completely different account of *wanna*-contraction that is compatible with a declarative analysis without requiring postulation of a different behaviour of two types of empty categories (see Pullum (1997)); or to devise a non-structural theory of reflexivization in which movement is not expected to potentially give rise to bleeding effects (see, e.g., Pollard & Sag (1992), Reinhart & Reuland (1993), Büring (2005)); or to account for the ban on ergative movement (and the concurrent lack of a comparable ban on accusative movement) in a different way that does not involve an interaction of movement and case assignent (see Campana (1992), Stiebels (2006), Coon et al. (2011)). These considerations notwithstanding, I take the derivational approach to syntax to be well supported by opacity arguments, and will presuppose it in what follows.

Taken together, the reasoning in the last two subjections suggests a local–derivational approach to syntax; and the phase-based model of Chomsky (2001) is just such an approach. Against this background, I will address improper movement, remnant movement, and resumptive movement constructions in German in the following three chapters. In each case, I will illustrate that there are backtracking problems under a phase-based approach, and I will argue that these problems can be solved in a simple way by postulating syntactic buffers.

Chapter 2

Improper Movement

1. Introduction

Different movement types can be distinguished by the different landing sites (or 'criterial positions', in Rizzi's (2007) terms) that they target. For instance, at least for present purposes and against the background of a clause structure consisting of CP, TP, vP, and VP, it can be assumed that scrambling in languages like German or Dutch targets a Specv position; the same may go for object shift in the Scandinavian languages. EPP-driven raising to subject in English ends up in a SpecT position. Wh-movement targets a SpecC position; and so on. When one considers locality restrictions on the various movement types, an interesting generalization emerges. It seems that there is a correlation between the position targetted by a movement type (low vs. high) and the distance over which it can apply (short vs. long): Movement types that have landing sites which are low in the clausal structure (e.g., SpecT, Specv) typically cannot apply long-distance; and movement types that have landing sites which are high in the clausal structure (e.g., SpecC) typically can apply long-distance. Thus, (1-ab) shows that scrambling in German is clause-bound; in contrast to, e.g., wh-movement or topicalization in the same language, a CP boundary cannot be crossed.

(1) a. dass das Buch$_1$ keiner t$_1$ liest
that the book$_{acc}$ no-one$_{nom}$ reads

b. *dass Karl das Buch$_1$ glaubt [$_{CP}$ dass keiner t$_1$ liest]
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$ reads

The same goes for object shift; see the Icelandic examples in (2-ab) (from Vikner (2005)).

(2) a. Ég veit [$_{CP}$ af verju þau seldu bókina$_1$ ekki t$_1$]
I know why they sold books$_{acc}$ not

b. *Ég veit bókina$_1$ [$_{CP}$ af verju þau seldu ekki t$_1$]
I know books$_{acc}$ why they sold not

Fronting of unstressed pronouns in German is also an operation that targets a TP-internal position in the clause, and it may not apply long-distance; see (3-ab).

(3) a. dass es$_1$ Fritz t$_1$ gelesen hat
that it$_{acc}$ Fritz$_{nom}$ read has

b. *dass ich es$_1$ glaube [$_{CP}$ dass Fritz t$_1$ gelesen hat]
that I$_{nom}$ it$_{acc}$ think that Fritz$_{nom}$ read has

The prohibition against non-clause-bound raising in English ('super-raising') is illustrated in (4).

(4) a. Mary$_1$ seems [$_{TP}$ t$_1$ to like John]
b. *Mary$_1$ seems [$_{CP}$ that t$_1$ likes John]

(5-ab) shows that whereas clitic movement in Italian does not have to be maximally local (it may target a matrix verb in restructuring infinitive constructions, as an instance of 'clitic climbing'), it can never cross a CP boundary (in non-restructuring environments).

(5) a. Mario lo$_1$ vuole [$_{TP}$ leggere t$_1$]
Mario it wants to read

b. *Mario lo$_1$ odia [$_{CP}$ C [$_{TP}$ leggere t$_1$]]
Mario it hates to read

Finally, extraposition in English may selectively violate certain island constraints (e.g., it may take place from subject DPs), but it cannot cross a CP (see Ross's (1967) Upward Boundedness Constraint, also known as the Right

Roof Constraint); cf. (6-ab).[1] This conforms to the above generalization if it is assumed that extraposition targets a low position in the clause.

(6) a. [$_{\text{DP}}$ A review t_1] will appear [$_{\text{PP}_1}$ of his new book]
b. *John always maintains [$_{\text{CP}}$ that [$_{\text{DP}}$ a review t_1] will appear shortly] whenever he is asked about it [$_{\text{PP}_1}$ of his new book]

The generalization correlating the height of the landing site and the possible length of the displacement path is standardly accounted for by a conspiracy of two constraints: a locality constraint and a constraint against improper movement. Thus, first, there is a locality constraint that permits extraction from a CP only via SpecC. For present purposes at least, this role can be played by the Phase Impenetrability Condition (PIC; Chomsky (2001), also see (1) of chapter 1), according to which only specifiers and the head of a phase are accessible to operations outside the phase (given that CP is a phase, and phrasal movement cannot target C). This precludes skipping the embedded SpecC position in (1-b), (3-b), (4-b), (5-b), and (6-b). Second, there is a constraint on improper movement according to which movement to a TP-internal position may precede movement to SpecC so as to permit (7-a) (where raising is followed by wh-movement), or indeed (7-b) (given that subjects are merged in Specv and then undergo EPP-driven movement to SpecT); but not vice versa: Movement from SpecC to a TP-internal position is blocked. This asymmetry can be taken to reflect the hierarchy of the target positions in the tree.

(7) a. [$_{\text{CP}}$ Who$_1$ C [$_{\text{TP}}$ t'_1 T seems t_1 to like John]] ?
b. [$_{\text{CP}}$ Who$_1$ C [$_{\text{TP}}$ t'_1 T [$_{\text{vP}}$ t_1 likes John]]] ?

In the following section, I will briefly discuss a number of proposals of how to formally capture this constraint against improper movement; and I will show that none of them meets all the requirements imposed by three general potential problems that I will assume to restrict the space for analyses: (a) the generality problem, (b) the locality problem, and (c) the promiscuity problem.

[1] Also see McCloskey (1999), Overfelt (2013) for a slightly different formulation of the constraint – such that extraposition may not leave the vP – that nevertheless maintains the basic insight.

2. Existing Analyses

2.1. Principle C

According to the highly influential account developed in May (1979) and adopted in Chomsky (1981), improper movement emerges as an instance of a Principle C effect. The account relies on two assumptions. First, locally A-bar bound traces qualify as a certain kind of trace that special constraints may hold for, viz., as *variables*; a trace is locally A-bar bound if its immediate chain antecedent is in an A-bar position, such as SpecC. And second, variables (in this technical sense) obey Principle C of the Binding Theory: They must not be bound from an A-position. On this view, a derivation of a super-raising construction as in (4-b) where an intermediate trace is established in SpecC (as required by a locality constraint like the PIC) is excluded by Principle C; see (8).

(8) *Mary$_1$ seems [$_{CP}$ t$'_1$ that t$_1$ likes John]

The initial trace t$_1$ qualifies as a variable (it is locally A-bar bound by the intermediate trace t$'_1$); however, t$_1$ is illegitimately also A-bound from the matrix SpecT position (an A-position).

To extend this account to other cases of improper movement, the respective movement types must be assumed to end up in A-positions, and the initial traces must also uniformly qualify as variables; see Fanselow (1990) for such an account of the clause-boundedness of scrambling in German.

2.2. Unambiguous Binding

In Müller & Sternefeld (1993), it is argued that a more general approach to improper movement is required because (a) scrambling in German does not exhibit all the typical properties of A-movement – it licenses parasitic gaps, it does not lead to new licensing options for reflexives and reciprocals, and so on; and (b) there are asymmetries between uncontroversial A-bar movement types as well, e.g., topicalization vs. wh-movement in German. To account for movement type asymmetries in general, Müller & Sternefeld (1993) rely on two assumptions. First, different movement types are defined by targetting different landing sites. Second, a Principle of Unambiguous Binding (PUB) makes use of these differences in landing sites; see (9).

(9) *Principle of Unambiguous Binding* (PUB):
A variable that is α-bound must be β-free in the domain of the head of its chain (where α and β refer to different types of positions).

Variables are defined as before, as locally A-bar bound traces. On this view, the ill-formed cases of improper movement in (1-b)–(6-b) are excluded by the PUB: Locality considerations require the use of SpecC as an intermediate escape hatch here, but doing so (a) ensures that the original trace t_1 in the base position qualifies as a variable, subject to the PUB, and (b) inevitably leads to a PUB violation because a variable t_1 is then ambiguously bound, by t_1' in a SpecC position, and by the head of the chain itself in the final target position – a SpecV/Specv position in the case of illegimate long-distance scrambling and object shift, a SpecT position in the case of illegitimate super-raising, a right-adjunction position in the case of illegitimate long-distance extraposition (see Müller (1996) for an analysis along these lines), and so on. In contrast, a sequence of A-movement followed by A-bar movement (as in (7)) is correctly predicted to be unproblematic because the original trace does not qualify as a variable (as in the approach based on Principle C).

2.3. The Williams Cycle

A third kind of constraint blocking improper movement goes back to Williams (1974); it has been further developed in Williams (2003). Versions of the constraint have been adopted by Sternefeld (1992), Grewendorf (2003; 2004), Abels (2008), Neeleman & van de Koot (2010), Bader (2011), and Keine (2014), among others. The basic idea is that movement to (or, more generally, rule application in) a specific domain in an embedded clause may be followed by movement to (or rule application in) the *same* kind of domain, or a *higher* domain, in the matrix clause, but not to (or in) a *lower* kind of domain in the matrix clause. As for the central notion of *syntactic domain* relevant here, Williams (1974) distinguishes between the following nested domains in a clause: $S' > S > \text{Pred} > \text{VP}$. Thus, once an item has undergone movement to, say, the Pred domain, any subsequent movement operation applying to this item can only target the Pred, S or S' domains; if an item has been moved to the S domain, a following movement operation applying to the same item can only go to S or S', and so on. This way, the generalization introduced at the beginning of the present chapter is implemented in a very direct way, essentially as a syntactic primitive. This constraint can be viewed as a specific version of the Strict Cycle Condition (see Chomsky (1973)); in line with this, I will here and henceforth refer to it as the "Williams Cycle".[2] The Williams Cycle is formulated as a *Generalized Ban on Improper Movement* (GBOIM) in Williams (2003, 72).[3]

(10) *Generalized Ban on Improper Movement* (GBOIM; Williams (2003)): Given a [...] clausal structure $X_1 > ... > X_n$ (where X_i takes X_{i+1}P as its complement), a movement operation that spans a matrix and an embedded clause cannot move an element from X_j in the embedded clause to X_i in the matrix, where i $<$ j.

In approaches that rely on some version of the Williams Cycle, improper movement as in (1-b)–(6-b) can in principle be accounted for; in particular, movement from SpecC to SpecV, Specv, or SpecT can be blocked because movement to a higher kind of domain in the embedded clause is followed by movement

---

[2] Williams (1974) does not give the constraint a name; but "Williams Cycle" is the label that the constraint was given in Chomsky's 1974 MIT class lectures (Edwin Williams, p.c.). – Note that the Williams Cycle is both more restrictive (in some areas) and potentially less restrictive (in others) than the Strict Cycle Condition. Consider the following version of the Strict Cycle Condition (a minimally updated version of Chomsky's original definition; see Müller (2011) for this specific formulation).

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

The Williams Cycle is more restrictive than the SCC in the sense that, for any given moved item δ, subsequent movement of δ may only go to a higher domain *that is of the same type or of a higher type*; in contrast, the SCC only requires subsequent δ movement to target some higher domain. On the other hand, in contrast to the SCC, the Williams Cycle (in the form in which it is presented in the main text) says nothing about the order of operations affecting *different items*; though see the original formulations in Williams (1974; 2003), which are somewhat more general in this respect.

[3] In Williams's (2003) system, the GBOIM is actually a theorem that follows directly from Williams' (more basic) Level Embedding Conjecture (LEC), which states that operations that take place at one level cannot take place again at a higher, more comprehensive level, where other operations defining that latter level apply; the levels that Williams envisages include FS (Focus Structure), SS (Surface Structure), CS (Case Structure), and TS (Theta Structure) (see Williams (2003, 23) for a fuller list). Since the LEC presupposes an organization of grammar that is radically different from more established standard derivational approaches, and since it does not seem to make radically different predictions empirically, I abstract away from it throughout this chapter.

to a lower kind of domain in the matrix clause. There is a proviso, though. The fatal first movement step to the embedded SpecC position that is required by locality is not inherently feature-driven; SpecC is not a ‘criterial position’ here. Thus, if the Williams Cycle is assumed to only hold for ‘criterial’ movement operations (see Abels (2008), for instance), improper movement in (1-b)–(6-b) is in fact not predicted to be impossible per se, and additional assumptions are called for to exclude the ill-formed derivations; see Abels (2012b) for one specific proposal. On the other hand, if intermediate non-criterial movement steps (that take place without inherent features of the host demanding them) *do* qualify as relevant for the Williams Cycle, then problems will arise as soon as one assumes that there are more intermediate landing sites required by locality than just SpecC. To wit, assuming the PIC, if vP is also a phase, the *intermediate* movement step from the embedded SpecC position to the matrix Specv position in the well-formed example in (11-a) instantiating long-distance wh-movement in German is wrongly excluded by the Williams Cycle in the same way that the *criterial* movement step from the embedded SpecC position the matrix Specv position in the ill-formed example in (11-b) showing that long-distance scrambling is impossible in German is excluded. ((11-b) = (1-b), with the intermediate traces added that are required by the PIC if vP and CP are phases.)

(11) a. Welches Buch$_1$ hat [$_{\mathrm{vP}}$ t$_1'''$ Karl gemeint [$_{\mathrm{CP}}$ t$_1''$ dass [$_{\mathrm{vP}}$ t$_1'$
which book$_{acc}$ has Karl meant that
jeder t$_1$ lesen möge]]] ?
everyone read should

b. *dass Karl [$_{\mathrm{vP}}$ das Buch$_1$ glaubt [$_{\mathrm{CP}}$ t$_1''$ dass [$_{\mathrm{vP}}$ t$_1'$
that Karl$_{nom}$ the book$_{acc}$ thinks that
keiner t$_1$ liest]]]
no-one$_{nom}$ reads

For now, I leave it at that. I will come back to this issue (it forms part of what I call the promiscuity problem).

Also note that that the symbol $<$ in $i < j$ in (10) is not to be interpreted as applying to numeric indices (in which case *i* and *j* would have to change places), but as a reversal of the embedding relation $>$.

2.4. *The Activity Condition*

In Chomsky (2000; 2001) and subsequent related work, an Activity Condition is adopted for syntactic operations: To be eligible for movement, an item must have an active (i.e., unchecked uninterpretable) feature sought by the movement-inducing head. This assumption provides a simple account of the ban on super-raising in English (see (4-b), repeated as (12)). In these constructions, the moved DP has its ϕ- and case features checked in the lower TP, by the embedded finite T; thus, the DP cannot be attracted by matrix T because it is not active anymore at this point.

(12) *Mary$_1$ seems [$_{\text{CP}}$ t$'_1$ that t$_1$ likes John]

The simplicity of the approach notwithstanding, it can be observed that conceptual and empirical problems have been noted with the Activity Condition (see Nevins (2004); also Bošković (2007) for critical discussion). One empirical argument raised by Nevins is that the Activity Condition is at variance with the existence of non-nominative subjects in SpecT (in languages like Icelandic) that have their ϕ- and case features checked independently (and earlier in the derivation).

2.5. *Feature Splitting*

An approach that is specifically designed to replace Chomsky's approach in terms of the Activity Condition is the Feature Splitting analysis developed in Obata & Epstein (2011). This approach is based on the following three assumptions. First, the PIC forces long-distance movement via SpecC. Second, uninterpretable features (like case features) are not permitted in the edge domain of a phase head (C) once the phase head's complement has undergone spell-out. (This is based on Richards's (2007) argument to this effect; also see Chomsky (2008)). Third and finally, in view of the second assumption, an operation of *feature splitting* must take place if a wh-subject is to undergo movement: The case/ϕ-features undergo movement to SpecT (under Agree with T, which has inherited the relevant probe features from C); and the wh- (or Q-) feature undergoes a separate (but, by assumption, simultaneous) movement step to SpecC. The derivation of a wh-subject question in English on the basis of these assumptions (and against the background of the *copy theory* of movement (re-) introduced in Chomsky (1993)) is illustrated in (13).

(13) $[_{\text{CP}}$ Who$_{\text{[wh]}}$ C $[_{\text{TP}}$ who$_{[\phi],\text{[case]}}$ T $[_{\text{vP}}$ who$_{\text{[wh]},[\phi],\text{[case]}}$ left]]] ?

The feature splitting approach covers super-raising without further ado. In cases like (14), matrix T does not find a matching goal: The copy in the lower SpecT position has undergone spell-out already, and the copy in the lower SpecC position does not have ϕ- and case features anymore.

(14) *Who seems $[_{\text{CP}}$ who$_{\text{[wh]}}$ C $[_{\text{TP}}$ who$_{[\phi],\text{[case]}}$ $[_{\text{T}}$ will] $[_{\text{vP}}$ who$_{\text{[wh]},[\phi],\text{[case]}}$ leave]]] ?

This analysis can be generalized to cases where the super-raised item is not a wh-phrase, as in (4-b)/(12): Irrespective of how an intermediate movement step of the (non-wh) DP to the embedded SpecC position (as required by the PIC) can be effected, it is clear that because of the assumption that case and ϕ-features cannot show up in SpecC, feature splitting must apply, and the DP in SpecC is not accessible to attraction by a higher T head anymore.[4]

2.6. Problems With the Existing Analyses

Closer inspection reveals that independently of potential individual shortcomings as they have been noted above, none of the accounts of improper movement just discussed can be maintained under a strictly derivational, local approach to displacement in which syntactic structure is generated bottom-up, by successive application of structure-building operations (such as internal or external Merge), and only very small parts of the structure are accessible at any given point in the derivation (cf. Chomsky (2000; 2001; 2008)). In particular, none of the existing accounts of improper movement manages to avoid all three separate problems that may arise with improper movement analyses from this perspective: (a) the generality problem, (b) the locality problem, and (c) the promiscuity problem. I discuss the three problems in turn.

2.6.1. Generality

The PUB-based and Williams Cycle-based accounts are general in the sense that all kinds of improper movement in (1-b)-(6-b) can be derived. In contrast, the Principle C account fails as soon as one of the instances of improper

[4] Also see Adger (2003, 388), who postulates that “only wh-features are visible in the specifier of CP”.

movement to a criterial position listed above can be shown to qualify as A-bar movement (as argued, e.g., in Müller & Sternefeld (1993) for scrambling, and in Müller (1996) for extraposition). Even more obviously, the Activity Condition-based and Feature Splitting-based accounts developed in Chomsky (2000) and Obata & Epstein (2011), respectively, are confined to super-raising, and cannot be generalized to other cases of improper movement (like long-distance scrambling in German) in any obvious way. In these other contexts, there is, by assumption, some head in the upper clause that attracts some item from the lower clause (i.e., that shares some feature with such an item) in a way that no other head (in the lower clause) does. So, independently of what the exact nature of the movement-related feature is that is involved in scrambling, pronoun movement, clitic climbing, object shift, and extraposition (if there is any such feature to begin with), it seems clear that such a feature could neither be rendered inactive in the embedded clause (because these features must be optional on the heads on which they occur, and, by assumption, therefore do not show up in the embedded clause if long-distance movement is to be triggered), as would be required under the Activity Condition-based analysis; nor could such a feature obligatorily have to be split off the item that undergoes movement to SpecC, and be checked in the TP domain (because T cannot check these features, and because it is unclear why these features should behave like case features on moved items with respect to interpretability, rather than like wh-features), as would be required under the feature splitting analysis. This consideration then only leaves PUB-based accounts and Williams Cycle-based accounts as serious contenders for a local–derivational implementation of the improper movement restriction.

2.6.2. Locality

Except for, possibly, the Activity Condition analysis and the Feature Splitting analysis, all the above accounts of improper movement require scanning large amounts of syntactic structure. Thus, the Principle C account must simultaneously take into account the base position of the moved item (which contains the trace that will ultimately give rise to a violation of Principle C); the position of the intermediate trace in the embedded SpecC position (which is relevant for determining whether the trace in base position must obey Principle C); and the position of the moved item in the final landing site.[5]

More importantly (given the Principle C account's lack of generality), the PUB-based account and the Williams Cycle-based accounts also face a locality problem. In the PUB-based account, to determine whether a trace is ambiguously bound, potentially large domains of syntactic structure must be checked that contain the initial trace, the moved item in the final landing site, and any intervening intermediate traces. Similarly, in Williams Cycle-based accounts, large pieces of structure must be considered: Under the formulation in (10), this is evident because the restriction explicitly holds for "a movement that *spans a matrix and an embedded clause*" (my emphasis). The same goes for other approaches relying on the Williams Cycle; see, e.g., Abels' (2008) inherently non-local notion of "affectedness" of large syntactic domains (cf. the appendix below). Thus, it can be concluded that the accounts of improper movement that circumvent a generality problem all face a locality problem: They are incompatible with a local–derivational approach to structure-building that permits only a very small amount of accessible syntactic structure at any step of the derivation (given the PIC).[6] More specifically, there is a backtracking problem (see chapter 1): To determine whether a movement step to, say, Specv counts as proper or improper, information must be available that specifies what kinds of phase edges the moved item has gone through (so as to determine whether the item has originated in the same clause or in an embedded clause).

2.6.3. Promiscuity

The third problem with existing approaches to improper movement arises under the assumption that many more intermediate positions are accessed in the course of successive-cyclic movement under current locality considerations than just SpecC (which used to be the standard assumption up to Chomsky

---

[5] By extension, this reasoning implies that Principle C and other binding conditions should be abandoned in general in local–derivational approaches to syntax, i.e., also for overt categories. See Fischer (2006) for an approach to binding condititions that complies with this requirement (but cannot be extended to improper movement in any obvious way); also see Reuland (2001).

[6] The second version of the Williams Cycle-based approach developed in Bader (2011, ch. 4-5) is an exception in this respect. In this analysis, locality can be maintained by postulating that the phrase-structural makeup of a moved item inherently fully mirrors the phrase structure through which it moves, and stipulating simultaneous spell-out operations of parallel features on the moved item and the clausal spine.

(1986)).[7] Given the PIC and the assumption that CP, vP, and DP are phases, intermediate movement steps to Specv, SpecC, and SpecD are required for all movement types without necessarily giving rise to improper movement effects. Things get only worse if *all* intervening XPs must be crossed via intermediate movement steps to SpecX in the course of movement; see Sportiche (1989), Takahashi (1994), Agbayani (1998); Chomsky (2005; 2008), Bošković (2002), Boeckx (2003), Boeckx & Grohmann (2007), Stroik (2009), Putnam (2009), and Müller (2011), among many others. Assuming either many or all intervening XPs to require and permit intermediate escape hatches, it is clear that the intermediate landing sites are highly *promiscuous* – they simply must not care what kind of ultimate target position a moved item will end up in.

This calls into question both the PUB-based account and Williams Cycle-based accounts of improper movement. A PUB-based account would predict virtually all movement to be improper: A wh-object moving via Specv to a clause-bound SpecC position would create an ambiguously bound initial trace in the same way that scrambling from SpecC to Specv does. Similarly, Williams Cycle-based accounts would make wrong predictions: Local movement of a wh-object to SpecC via Specv would still be unproblematic (in contrast to what would be the case under a PUB-based account); however, as noted above, well-formed long-distance wh-movement to a matrix SpecC position via first an embedded Specv position, then an embedded SpecC position and finally a matrix Specv position would wrongly be excluded in the same way that long-distance scrambling via first an embedded Specv position and then an embedded SpecC position is correctly excluded as improper; recall the two constructions in (11-a) (legitimate long-distance wh-movement) and (11-b) (illegitimate long-distance scrambling), which are repeated here as (15-ab).

(15) a. Welches Buch$_1$ hat [$_{vP}$ t_1''' Karl gemeint [$_{CP}$ t_1'' dass [$_{vP}$ t_1'
which book$_{acc}$ has Karl meant that
jeder t_1 lesen möge]]] ?
everyone read should

b. *dass Karl [$_{vP}$ das Buch$_1$ glaubt [$_{CP}$ t_1'' dass [$_{vP}$ t_1'
that Karl$_{nom}$ the book$_{acc}$ thinks that

---

[7] A version of this problem is also mentioned in Neeleman & van de Koot (2010, 346-347) and Bader (2011, ch. 5).

keiner t_1 liest]]]
no-one$_{nom}$ reads

Thus, it seems that if massive intermediate movement steps to promiscuous escape hatches are assumed, a dilemma is unavoidable for a PUB-based account and for Williams Cycle-based accounts: Either it is postulated that only criterial positions (final landing sites of movement) count for improper movement. Then it is unclear how, e.g., long-distance scrambling via SpecC can be excluded (where the intermediate SpecC landing site is not a criterial position); more generally, none of the improper movement effects in (1-b)–(6-b) can be derived anymore. Or it is assumed that all positions (including all non-criterial intermediate positions) count for improper movement. Then it is unclear how, e.g., long-distance wh-movement via matrix Specv can be permitted (given that long-distance scrambling targetting the same position needs to be ruled out). In a nutshell, given promiscuous intermediate movement steps, the accounts of improper movement that handle the generality problem are either not restrictive enough anymore, or they are much too restrictive. This implies that either additional assumptions must be made to save these accounts, or that they must be abandoned, and replaced by something completely different.[8]

[8] Abels (2012b) pursues the first strategy. He adopts a weak version of the Williams Cycle where intermediate traces in non-criterial positions are simply ignored, and then invokes an additional system of "flavoured" edge features for intermediate movement steps that mimick the ultimate features giving rise to criterial movement. The analysis works such that for each phase head requiring an intermediate movement step, it is stipulated (possibly from language to language) which kind of flavoured edge features it can be equipped with. If, e.g., C cannot have a flavoured case/ϕ edge feature but can have a flavoured wh-edge feature, wh-movement can apply long-distance whereas raising cannot; if the restrictions on flavoured edge features are reversed on C in a language, super-raising is possible whereas long-distance wh-movement is not; and so on. Abels (2012a;b) adduces potentially interesting evidence from Tagalog to support such an approach. However, I will not consider this approach in any more detail in what follows because it denies that there is any inherent systematicity to the effects in (1-b)–(6-b); i.e., on this view, the generalization formulated at the beginning of the present chapter (according to which movement types with low landing sites tend not to apply long-distance) simply does not exist – and this despite the fact that the flavoured edge features of Abels (2012b) are assumed to accompany, rather than replace, the Williams Cycle. Furthermore, note that these kinds of flavoured edge features cannot by themselves provide a comprehensive account of *all* relevant instances of improper movement (e.g., if two inherently feature-driven movement operations ending up in criterial positions are combined, flavoured edge features as such cannot rule out the combination as improper). Finally, it is worth pointing out that non-promiscuous, flavoured edge features are incompatible with the

2.6.4. Interim Conclusion

In view of all this, my goal in what follows is to provide a *local reformulation* of the Williams Cycle as a core component of the theory of improper movement that is compatible with a strictly derivational approach, with extremely small accessible domains throughout (where each phrase is a phase), and that meets the requirements imposed not only by the locality problem, but also by the remaining two problems just discussed: It has to be general (covering all the cases in (1-b)–(6-b)); and it has to be compatible with the promiscuity of edge features.[9]

3. A Local–Derivational Approach to the Williams Cycle

3.1. Background: Edge Features and Successive-Cyclic Movement

Following Chomsky (2000; 2001; 2008) and much related work, I assume that intermediate movement steps are brought about by edge features. Since the generation and discharge of edge features will be instrumental in accounting for improper movement effects by a reformulated Williams Cycle to be developed below, some clarifications about edge features and the role that they play in derivations are called for at this point.

The basic question is whether edge feature insertion is assumed to be freely available or severely constrained. A version of the first option is pursued in Chomsky (2008; 2013), where phase heads are simply assumed to have an "edge property" that allows them to generate any number of specifiers; this is extensionally equivalent to assuming that edge feature insertion is freely available throughout. The second option is adopted in Chomsky (2000; 2001), where constraints on edge feature insertion are specified. It seems clear that if edge feature insertion is free (or if phase heads have an edge property), and edge-features are category-neutral, no restrictions on improper movement can be imposed in the domain of edge features. I assume that edge feature in-

---

assumption of a crash-proof syntax (see Frampton & Gutmann (2002)) because most choices of edge features on phase heads will be incompatible with the item that in fact needs to undergo the intermediate movement step.

[9] Abels (2008) remarks that his version of the Williams Cycle "cannot be understood directly as a constraint on derivations", and makes "no attempt to reformulate [it] in derivational terms". Essentially, this is what I set out to do in what follows.

sertion is not free. Independent evidence for this assumption is provided in Müller (2011), where it is shown that constraints on edge feature insertion play a decisive role in deriving Minimal Link Condition (MLC) and Condition on Extraction Domain (CED) effects from the PIC. In what follows, I will adopt a simplified version of this approach as a general background for a theory of improper movement; as noted in chapter 1, this is due to the need to have *some* sufficiently explicit frame of reference within which a local version of the Williams Cycle based on the generation and discharge of edge features can be formulated.

Thus, suppose first that all phrases are phases, and that the PIC holds. Second, all syntactic operations are driven by designated features: In particular, there are structure-building features ([•F•]) that trigger internal and external Merge operations (movement and base-concatenation, respectively); for instance, wh-movement is triggered by [•wh•] on C. Third, edge feature insertion as it is required for effecting intermediate movement steps (given the second assumption) is restricted by the Edge Feature Condition in (16) (this is a modification of what is proposed in Chomsky (2000; 2001)).

(16) *Edge Feature Condition*:
An edge feature [•X•] can be assigned to the head π of a phase only if (a) π is still active and (b) this has an effect on outcome.

In the present context, it does not matter how *activity* of a phase head is determined (but I will return to this issue in chapter 4); how the requirement of having *an effect on outcome* is to be interpreted in a local way, without look-ahead; and how MLC and CED effects follow from the PIC under these assumptions; see Müller (2011) for all this. What is important, though, is that edge features play a crucial role in licencing intermediate movement steps to phase edges, that edge feature insertion is restricted, and that this approach to edge features has interesting consequences for deriving locality constraints on movement – thus, edge features have independent empirical motivation.

However, from a minimalist perspective, there is a basic problem with the very existence of inserted edge features in syntactic derivations: Edge feature insertion violates the Inclusiveness Condition (see, e.g., Chomsky (2001; 2005; 2008)), according to which material that is not originally part of the numeration cannot be introduced into syntactic derivations in the course of the derivation.

3.2. Assumptions

I would like to suggest that the solution to the problem with the Inclusiveness Condition is that edge features are not in fact inserted in the course of the derivation; rather, they involve material that is already present (also cf. Lahne (2009)). For concreteness, suppose that edge features are *defective copies* of categorial features of phase heads that have been assigned a structure-building property: The original categorial information is stripped off but retained in some form on the feature. The edge features thus generated successively value movement-related features of moved items passing through the specifier positions of the phase heads where the respective edge feature originates, thereby creating lists that record aspects of the derivational history of movement – i.e., syntactic *buffers*. Such information is maintained for a while in derivations, but is deleted as soon as information of the same type is encountered. Finally, when a criterial target position is reached, the *functional sequence* (f-seq) of heads (see Starke (2001)) must be respected on such lists; this is the Williams Cycle. Thus, in a nutshell, the categorial information of the domains that a moved item passes through is picked up and registered on it, potentially yielding temporary improper movement configurations, but since the relevant information can be lost again, such temporary improper movement configurations can in principle be fixed before the criterial position is reached.

More specifically, I will make the following assumptions about the mechanics of edge feature generation and discharge. First, an edge feature is a defective copy of the categorial feature of a phase head accompanied by a structure-building instruction ([• •]; cf. Lahne (2009)). The copy mechanism is given in (17-a) (with γ a variable over category labels), and it is illustrated for some phase heads in (17-b).

(17) a. $[\gamma] \rightarrow [\gamma]$, $[\bullet X_{\gamma} \bullet]$
b. $[V] \rightarrow [V]$, $[\bullet X_{V} \bullet]$; $[v] \rightarrow [v]$, $[\bullet X_{v} \bullet]$; $[T] \rightarrow [T]$, $[\bullet X_{T} \bullet]$; $[C] \rightarrow [C]$, $[\bullet X_{C} \bullet]$; ...

As shown in (17), the original content of the feature is lost in the course of defective copying; this makes the feature usable (i.e., there is no instruction to merge an item with the exact same categorial feature as that of the phase head, which one would expect under non-defective copying). However, crucial aspects of the original information (viz., the categorial feature of the phase head) remain intact so as to make it possible to trace (recent) steps of the derivation:

The categorial information is still there as part of the structure-building edge feature, but it does not by itself restrict the nature of the merge operation that the edge feature effects. Accordingly, the Edge Feature Condition needs to be revised as in (18) (compare (16)).

(18) *Edge Feature Condition (revised)*:
An edge feature [$\bullet$X$_\gamma$$\bullet$] can be generated by defective copying of the categorial feature of a head γ of a phase only if (a) γ is active and (b) this has an effect on outcome.

Second, movement-related features on moved items (i.e., the β features that are attracted by phase heads with corresponding structure-building features [$\bullet\beta\bullet$]) have *lists* as values; as will become clear below, these lists act as syntactic *buffers*. This is shown for the features that I assume to be involved in scrambling, wh-movement, topicalization, relativization, and EPP-driven movement to SpecT in (19), with $\Box$ representing an initially empty list.[10]

(19)
a. [Σ:$\Box$] (scrambling)
b. [wh:$\Box$] (wh-movement)
c. [top:$\Box$] (topicaliztion)
d. [rel:$\Box$] (relativization)
e. [EPP:$\Box$] (raising to SpecT)

Third, edge feature discharge involves valuation of the movement-related feature of the moved item by the (defective) categorial information on the phase head, so as to ensure complete matching of the two items. Categorial information is successively added on top of the list, which thus acts as a buffer

[10] Whereas [wh:$\Box$] and [rel:$\Box$] features for wh-movement and relativization, respectively, can be assumed to obligatorily show up on an item, [Σ:$\Box$] and [top:$\Box$] features for scrambling and topicalization must be optional on items. Similarly, the movement-inducing features [$\bullet$wh$\bullet$] and [$\bullet$rel$\bullet$] show up obligatorily on interrogative and relative C, respectively (in languages of the English or German type), whereas [$\bullet\Sigma\bullet$] is optional on v/V in scrambling languages, and [$\bullet$top$\bullet$] is obligatory on C in German verb-second clauses, and optional (on C or some additional functional head of the left periphery) in English. As for the nature and justification of of a feature like [($\bullet$)Σ($\bullet$)] for scrambling, see Grewendorf & Sabel (1999) and Sauerland (1999), among others. Also, for present purposes I will remain uncommitted as to what ultimately underlies the EPP feature postulated here, and whether it is related to case, categorial label, or something else yet.

storing minimal aspects of the derivation. This is shown in (20) for an abstract derivation in which a wh-phrase undergoes successive-cyclic movement via all intervening phase edges to the embedded SpecC position (which is not its ultimate target position because, by assumption, C is declarative here and lacks [•wh•]).[11]

(20)
a. Merge(V:[•X_V•], DP:[wh:□]) $\rightarrow$ V DP:[wh:$\boxed{\text{V}}$]
b. Merge(v:[•X_v•], DP:[wh:$\boxed{\text{V}}$]) $\rightarrow$ v DP:[wh:$\boxed{\text{vV}}$]
c. Merge(T:[•X_T•], DP:[wh:$\boxed{\text{vV}}$]) $\rightarrow$ T DP:[wh:$\boxed{\text{TvV}}$]
d. Merge(C:[•X_C•], DP:[wh:$\boxed{\text{TvV}}$]) $\rightarrow$ C DP:[wh:$\boxed{\text{CTvV}}$]

Fourth, when identical categorial information is added at the top, the original information is deleted at the bottom.[12] Such a deletion of parts of the derivational record in feature value lists is depicted in (21), which continues the derivation in (20) into the matrix clause; prior $\boxed{\text{V}}$ information is deleted once the wh-phrase encounters a new SpecV position; and prior $\boxed{\text{v}}$ information is deleted when it reaches a new Specv position.

(21)
a. Merge(V:[•X_V•], DP:[wh:$\boxed{\text{CTvV}}$]) $\rightarrow$ V DP:[wh:$\boxed{\text{VCTv}\cancel{\text{V}}}$]
b. Merge(v:[•X_v•], DP:[wh:$\boxed{\text{VCTv}\cancel{\text{V}}}$]) $\rightarrow$ v DP:[wh:$\boxed{\text{vVCT}\cancel{\text{vV}}}$]

The operations of (recursive) valuation of a movement-related feature and of deletion in feature lists are formulated more generally in (22).

(22)
a. *Valuation*:
Merge(Y:[•X_γ•], Z:[F:$\boxed{\delta_1 \text{ ... } \delta_n}$]) $\rightarrow$ Y Z:[F:$\boxed{\gamma\ \delta_1 \text{ ... } \delta_n}$]
where F is a movement-related feature, δ_1, ..., δ_n is a (possibly

[11] Note that linearization is not indicated in this and the following abstract representations capturing the result of Merge; the same goes for the projection status of the category that the moved item is merged with.

[12] Thus, the list that acts as the value of a movement-related feature is (essentially, see footnote 13 below) a *queue* (i.e., a first-in/first-out list), not a *stack* (i.e., a first-in/last-out list). The system outlined here is derivational, and information gets lost during the derivation. Effectively, the proposed deletion mechanism instantiates a ban on recursion in feature value lists, possibly motivated by economy considerations; cf. also Heck (2010) and Arsenijević & Hinzen (2012), who argue that direct recursion is not available in syntax.

empty) list of (category, possibly other) symbols, and γ is the category label of Y.

b. *Deletion*:
Y Z:[F: $\boxed{\gamma\ \delta_1\ ...\ \delta_i\ \gamma\ \delta_j\ ...\ \delta_n}$] $\rightarrow$ Y Z:[F: $\boxed{\gamma\ \delta_1\ ...\ \delta_i\ \delta_j\ ...\ \delta_n}$]
where F is a movement-related feature, δ_1, ..., δ_n is a (possibly empty) list of (catgory, possibly other) symbols, and γ is the category label of Y.[13]

Finally, when a moved item has reached its target position, it discharges the movement-related structure-building feature of the head. This feature (which is inherently present on a phase head and not generated in accordance with the Edge Feature Condition) must also carry the categorial information of the head it is associated with, e.g., [$\bullet$wh$_{\mathrm{C}}\bullet$]; this ensures deletion of the earlier $\boxed{\mathrm{C}}$ information in the case at hand; see (23) (where (23-a) continues where (21-b) left off).

(23) a. Merge(T:[$\bullet$X$_{\mathrm{T}}\bullet$], DP:[wh: $\boxed{\mathrm{vVCT}\cancel{\mathrm{vV}}}$])
$\rightarrow$ T DP:[wh: $\boxed{\mathrm{TvVC}\cancel{\mathrm{TvV}}}$]

b. Merge(C:[$\bullet$wh$_{\mathrm{C}}\bullet$], DP:[wh: $\boxed{\mathrm{TvVC}\cancel{\mathrm{TV}}}$])
$\rightarrow$ C DP:[wh: $\boxed{\mathrm{CTvV}\cancel{\mathrm{CTvV}}}$] ($\surd$f-seq)

Crucially, at this point (i.e., in the criterial position), it is determined whether the information recording the intermediate landing sites conforms to the functional sequence (f-seq) independently established in syntactic structures (e.g., C$\succ$T$\succ$v$\succ$V); in the case at hand, it does. This is the new version of the

[13] Two remarks. First, (22-b) presupposes that deletion of categorial information in feature lists is to be taken literally. Still, since nothing hinges on this issue, I continue to render deleted γ as $\cancel{\gamma}$ in the derivations that follow in this chapter, so as to maximize perspicuity. Second, (22-b) is formulated in a slightly more liberal way that permits deletion under identity in a non-final position (i.e., not necessarily at the bottom of the list). This difference play will no role for the core phenomena to be discussed below, given that f-seqs have a uniform structure and all deletion will take place at the bottom of the list; but things are different under the approach to cross-linguistic variation developed in section 4 (cf. in particular the derivations in (40) and (41) below). Also, an extension of the present proposal to remnant movement and resumption in the next two chapters, while not actually requiring it in the vast majority of constructions, will nevertheless suggest a slightly more liberal concept as more natural (see particularly the discussion of (29) in chapter 3 below).

Williams Cycle, which can be formulated as in (24).

(24) *Williams Cycle*:
Information on a list of a movement-related feature β must conform to f-seq when β is checked by an inherent structure-building feature [$\bullet\beta_{\pi}\bullet$] of a phase head π (i.e., in criterial positions).

Before moving on, let us see how the present system fares with respect to the Inclusiveness Condition. First, defective copying of categorial features does not violate the Inclusiveness Condition; copying as such is generally taken to be unproblematic from this perspective, and defective copying removes rather than adds information. In addition, the current approach envisages the assignment of a structure-building property ([$\bullet$ $\bullet$]) to a feature resulting from copying; again I take this to be innocuous since it does not add a feature but rather interprets a given feature in one of very few possible ways (structure-building, probe, inert). Finally, valuation of movement-related features by successively copying categorial information to the list does not add anything to the derivation that was not present before; in this respect, valuation of movement-related features behaves exactly like all other kinds of feature valuation that are standardly assumed, even if the feature value is slightly more complex, and changes repeatedly throughout the derivation. All in all, I would like to contend that the current approach is close to optimal as regards the reconciliation of restricted edge feature generation with the Inclusiveness Condition.

With the new Williams Cycle-based system in place, let me go through some sample derivations distinguishing proper from improper movement.

3.3. Clause-Bound Wh-Movement

Consider first a simple case of clause-bound wh-movement, as in (25) in English.

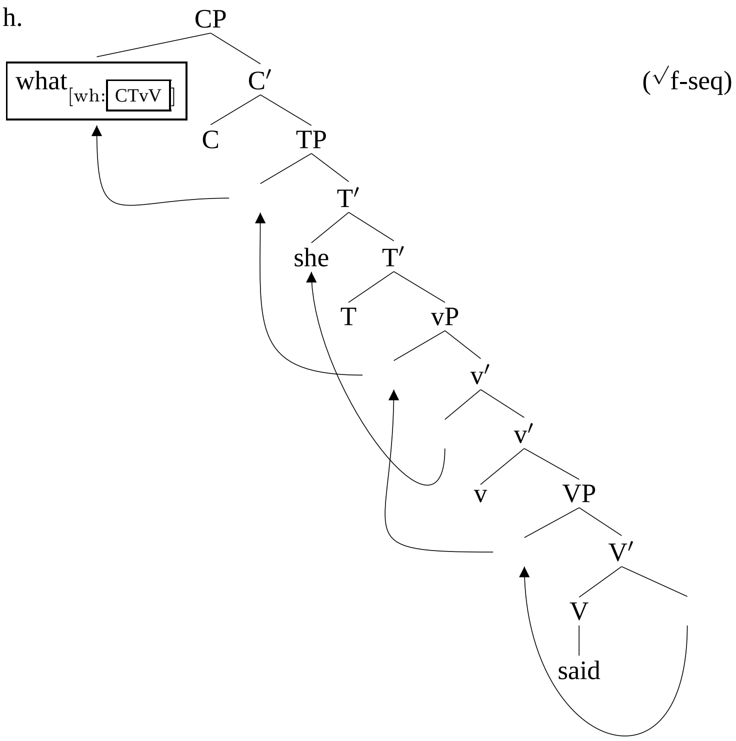
(25) (I wonder) [$_{\text{CP}}$ what$_2$ C [$_{\text{TP}}$ she$_1$ T [$_{\text{vP}}$ t$_1$ v [$_{\text{VP}}$ said t$_2$]]]]

The derivation is shown in (26). On the VP level, the wh-object *what* first needs to undergo movement to SpecV because of the PIC, given that first-merged items (i.e., complements) are not yet part of the edge domain of a phase. An edge feature [$\bullet X_V \bullet$] can be generated by defective copying here in accordance with the Edge Feature Condition in (18) (see (26-a)), and edge feature discharge triggers movement of DP$_{\text{wh:}\square}$ to SpecV, valuing the movement-related feature by adding categorial V information to the (initially empty) list: $\mathrm{DP}_{\mathrm{wh}:\boxed{\mathrm{V}}}$; see (26-b).[14] Next, on the vP level, a new edge feature [$\bullet X_v \bullet$] is generated (see (26-c)) and discharged by movement of $\mathrm{DP}_{\mathrm{wh}:\boxed{\mathrm{V}}}$ to Specv, which further values the wh-feature on the moved item: $\mathrm{DP}_{\mathrm{wh}:\boxed{\mathrm{vV}}}$; see (26-d). The pattern is repeated on the TP level, where [$\bullet X_T \bullet$] is first generated (see (26-e)) and then discharged by movement of DP to SpecT, thereby valuing its wh-feature with the newly encountered syntactic context; see $\mathrm{DP}_{\mathrm{wh}:\boxed{\mathrm{TvV}}}$ in (26-f) (in addition, subject raising to SpecT takes place, triggered by an inherent [$\bullet$EPP$\bullet$] feature). Finally, on the CP level, the inherent movement-inducing feature [$\bullet \mathrm{wh}_C \bullet$] on the interrogative C head (see (26-g)) triggers movement of $\mathrm{DP}_{\mathrm{wh}:\boxed{\mathrm{TvV}}}$ to SpecC, valuing the wh-feature by adding the categorial information and thereby producing $\mathrm{DP}_{\mathrm{wh}:\boxed{\mathrm{CTvV}}}$; cf. (26-h). Since an inherent structure-building feature has been checked at this point (which is signalled by a box around moved items in trees, here and in what follows), the Williams Cycle in (24) demands matching of the categorial information on the list that is the value of the moved DP with f-seq; since the former conforms to the latter, the derivation is legitimate (of course, the same goes for EPP-driven subject movement in (25-f), which is also criterial).

(26)
a. $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said $]_{[\bullet X_V \bullet]}$ $\mathrm{what}_{[\mathrm{wh}:\square]}$]
b. $[_{\mathrm{VP}}$ $\mathrm{what}_{[\mathrm{wh}:\boxed{\mathrm{V}}]}$ $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said]]]
c. $[_{\mathrm{v}'}$ she $[_{\mathrm{v}'}$ $\mathrm{v}_{[\bullet X_v \bullet]}$ $[_{\mathrm{VP}}$ $\mathrm{what}_{[\mathrm{wh}:\boxed{\mathrm{V}}]}$ $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said]]]]]
d. $[_{\mathrm{vP}}$ $\mathrm{what}_{[\mathrm{wh}:\boxed{\mathrm{vV}}]}$ $[_{\mathrm{v}'}$ she $[_{\mathrm{v}'}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said]]]]]]
e. $[_{\mathrm{T}'}$ $\mathrm{T}_{[\bullet X_T \bullet],[\bullet \mathrm{epp} \bullet]}$ $[_{\mathrm{vP}}$ $\mathrm{what}_{[\mathrm{wh}:\boxed{\mathrm{vV}}]}$ $[_{\mathrm{v}'}$ she $[_{\mathrm{v}'}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said]]]]]]]
f. $[_{\mathrm{TP}}$ $\mathrm{what}_{[\mathrm{wh}:\boxed{\mathrm{TvV}}]}$ $[_{\mathrm{T}'}$ she T $[_{\mathrm{vP}}$ $[_{\mathrm{v}'}$ $[_{\mathrm{v}'}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said]]]]]]]]
g. $[_{\mathrm{C}'}$ $\mathrm{C}_{[\bullet \mathrm{wh} \bullet]}$ $[_{\mathrm{TP}}$ $\mathrm{what}_{[\mathrm{wh}:\boxed{\mathrm{TvV}}]}$ $[_{\mathrm{T}'}$ she T $[_{\mathrm{vP}}$ $[_{\mathrm{v}'}$ $[_{\mathrm{v}'}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V}'}$ $[_{\mathrm{V}}$ said]]]]]]]]]

[14] Note that this is at variance with the assumption that extremely local movement is precluded; it implies that a strict anti-locality requirement on movement cannot hold, pace Bošković (1997), Abels (2003), and Grohmann (2003), among others.

h.



3.4. Long-Distance Wh-Movement

Consider next the case of long-distance wh-movement from a declarative CP, as in (27) in English.

(27) What$_2$ do you think [$_{CP}$ C [$_{TP}$ she$_1$ T [$_{vP}$ t$_1$ v [$_{VP}$ said t$_2$]]]] ?

The derivational steps in the embedded CP are almost exactly as in (26); see (28). The only relevant difference to (26) is that movement on the CP level is required not by an inherent structure-building feature of C (because there is no such feature), but by an edge feature [$\bullet X_C \bullet$] that is generated in accordance with the revised Edge Feature Condition in (18), by defective copying of the categorial feature of the phase head. As a consequence, valuation of the wh-feature in (28-h) does not activate the Williams Cycle: Movement has not yet targetted a criterial position.

(28) a. $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said $]_{[\bullet\mathrm{X_V}\bullet]}$ what$_{[\mathrm{wh}:\square]}$]
b. $[_{\mathrm{VP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{V}}]}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]
c. $[_{\mathrm{v'}}$ she $[_{\mathrm{v'}}$ v$_{[\bullet\mathrm{X_v}\bullet]}$ $[_{\mathrm{VP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{V}}]}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]]]
d. $[_{\mathrm{vP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{vV}}]}$ $[_{\mathrm{v'}}$ she $[_{\mathrm{v'}}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]]]]
e. $[_{\mathrm{T'}}$ T$_{[\bullet\mathrm{X_T}\bullet],[\bullet\mathrm{epp}\bullet]}$ $[_{\mathrm{vP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{vV}}]}$ $[_{\mathrm{v'}}$ she $[_{\mathrm{v'}}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]]]]]
f. $[_{\mathrm{TP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{TvV}}]}$ $[_{\mathrm{T'}}$ she T $[_{\mathrm{vP}}$ $[_{\mathrm{v'}}$ $[_{\mathrm{v'}}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]]]]]]
g. $[_{\mathrm{C'}}$ C$_{[\bullet\mathrm{X_C}\bullet]}$ $[_{\mathrm{TP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{TvV}}]}$ $[_{\mathrm{T'}}$ she T $[_{\mathrm{vP}}$ $[_{\mathrm{v'}}$ $[_{\mathrm{v'}}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]]]]]]]
h. $[_{\mathrm{CP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{CTvV}}]}$ $[_{\mathrm{C'}}$ C $[_{\mathrm{TP}}$ she $[_{\mathrm{T'}}$ T $[_{\mathrm{vP}}$ $[_{\mathrm{v'}}$ $[_{\mathrm{v'}}$ v $[_{\mathrm{VP}}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ said]]]]]]]]]]

On the matrix VP, vP, and TP levels, edge feature generation, edge feature discharge, and wh-valuation on the moved item proceed as in the embedded domain, but there is an interesting difference: Movement to matrix SpecV in (28-j) adds the symbol V at the top of the wh-feature list of the moved DP, and concurrently *deletes* this categorial information at the bottom (in accordance with (22-b)); movement to matrix Specv in (28-l) adds v on top of the list and deletes v at the bottom; and movement to matrix SpecT in (28-n) does the same with T. In all three cases, a feature list results that does not conform to f-seq (viz., $\boxed{\mathrm{VCTv}}$, $\boxed{\mathrm{vVCT}}$, and $\boxed{\mathrm{TvVC}}$), respectively; but since all three movement steps are triggered by edge features generated in order to comply with the PIC rather than by inherent structure-building features of a phase head, this is unproblematic from the perspective of the Williams Cycle, which is satisfied vacuously in these contexts. In contrast, the final movement step to SpecC in (28-p) *is* triggered by an inherent movement-inducing feature of the matrix interrogative C head, viz., [•wh$_{\mathrm{C}}$•], so the Williams Cycle will spring into action and demand a correspondence of f-seq and the wh-feature list present on the moved DP. However, movement to SpecC has resulted in adding C to the feature list, and the lower C symbol is then deleted; thus, the temporary improper movement configuration ceases to exist in the criterial position. Therefore, the Williams Cycle is (non-vacuously) satisfied by the final movement step.

(27) i. $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ think$]_{[\bullet\mathrm{X_V}\bullet]}$ $[_{\mathrm{CP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{CTvV}}]}$ $[_{\mathrm{C'}}$ C $[_{\mathrm{TP}}$ she said]]]]
j. $[_{\mathrm{VP}}$ what$_{[\mathrm{wh}:\boxed{\mathrm{VCTv}\cancel{\mathrm{V}}}]}$ $[_{\mathrm{V'}}$ $[_{\mathrm{V}}$ think] $[_{\mathrm{CP}}$ $[_{\mathrm{C'}}$ C $[_{\mathrm{TP}}$ she said]]]]]

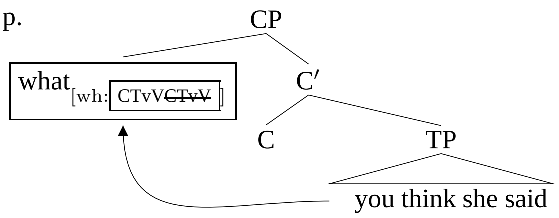
k. $[_{\text{v}'}$ you $[_{\text{v}'}$ v$_{[\bullet\text{X}_\text{v}\bullet]}$ $[_{\text{VP}}$ what$_{[\text{wh:}\boxed{\text{VCTv}\cancel{\text{V}}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ think$]$ $[_{\text{CP}}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]]]

l. $[_{\text{vP}}$ what$_{[\text{wh:}\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ $[_{\text{v}'}$ you $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ think$]$ $[_{\text{CP}}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]]]]]

m. $[_{\text{T}'}$ T$_{[\bullet\text{X}_\text{T}\bullet],[\bullet\text{epp}\bullet]}$ $[_{\text{vP}}$ what$_{[\text{wh:}\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ $[_{\text{v}'}$ you $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]

n. $[_{\text{TP}}$ what$_{[\text{wh:}\boxed{\text{TvVC}\cancel{\text{TvV}}}]}$ $[_{\text{T}'}$ you $[_{\text{T}'}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]]

o. $[_{\text{C}'}$ C$_{[\bullet\text{wh}\bullet]}$ $[_{\text{TP}}$ what$_{[\text{wh:}\boxed{\text{TvVC}\cancel{\text{TvV}}}]}$ $[_{\text{T}'}$ you $[_{\text{T}'}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]]]]

p.


($\sqrt{}$f-seq)

Whereas the Williams Cycle thus predicts wh-movement to be able to apply non-locally, in a successive-cyclic manner, predictions are quite different for movement types which target a lower position in the clause, like scrambling. I turn to this in the next section.

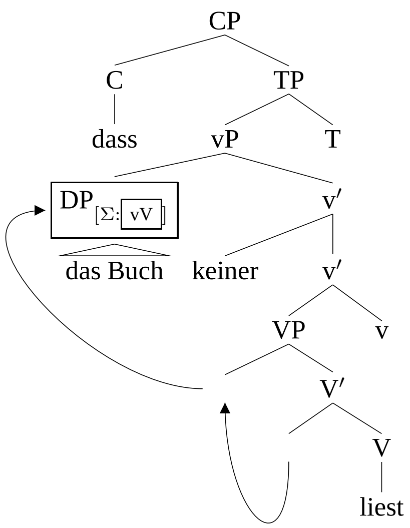
3.5. *Clause-Bound Scrambling*

Suppose, as before, that scrambling (in German) targets SpecV or Specv, and involves optional structure-building ([$\bullet\Sigma\bullet$]) and movement-related ([Σ:$\square$]) features on the attracting V or v head and the moved item, respectively. In (29) (= (1-a)), scrambling must target a Specv position, with the subject DP *keiner* (‘no-one’) staying in situ.

(29) dass das Buch$_1$ keiner t_1 liest
that the book$_{acc}$ no-one-nom reads

The derivation is straightforward, and shown in (30). Extremely local movement to SpecV (which is also string-vacuous, given the SOV nature of German) takes place at first (see (30-b)). This movement step is brought about by an edge feature $[\bullet X_V \bullet]$ generated on V in accordance with the Edge Feature Condition, and it values the list of the Σ-feature on the object DP with the symbol V. Since, by assumption, the next higher v head already bears the structure-building feature $[\bullet\Sigma\bullet]$, a criterial position is reached in the next step, and movement of the object DP stops here (see (30-d)). The Williams Cycle is therefore checked at this point, and since the sequence $\boxed{\text{vV}}$ conforms to f-seq, the derivation can legitimately continue to the TP and CP levels, as in (30-ef).

(30)
a. $[_{V'}$ $[_{DP}$ das Buch $]_{[\Sigma:\square]}$ $[_{V}$ liest $]_{[\bullet X_V \bullet]}$]
b. $[_{VP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{V}]}$ $[_{V'}$ $[_{V}$ liest]]]
c. $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{V}]}$ $[_{V'}$ $[_{V}$ liest]]]] $v_{[\bullet\Sigma\bullet]}$]
d. $[_{vP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]]
e. $[_{TP}$ $[_{vP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]] T]
f.



(√f-seq)

3.6. *The Ban on Long-Distance Scrambling*

Things are different with illegitimate long-distance scrambling in German; cf. (31) (= (1-b)).

(31) *dass Karl das Buch glaubt [$_{CP}$ dass keiner t_1 liest]
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$ reads

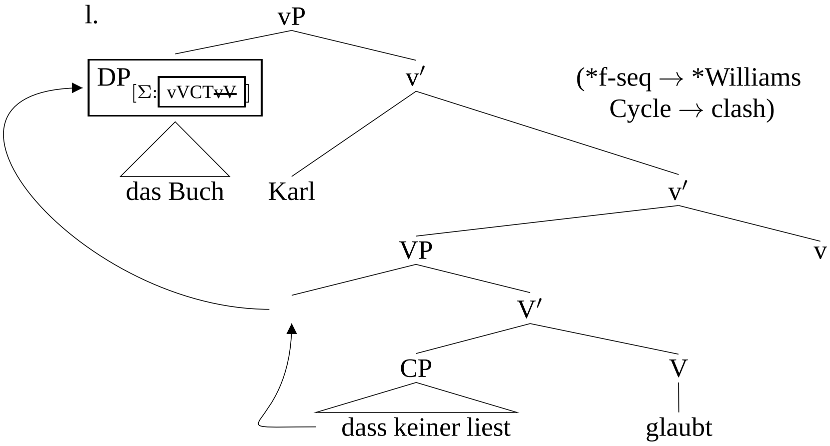
In the embedded clause, the derivation proceeds in basically the same way as the derivation of well-formed long-distance wh-movement in steps a.–h. of (28). The only relevant difference (lexical choices, absence of the EPP, and linearization aside) is that the movement-related feature on the object DP that gets valued successively by categorial information associated with the domains that it passes through is now [Σ:□], and not [wh:□] anymore. These steps are illustrated in (32-a)–(32-h). The list $\boxed{\text{CTvV}}$ resulting at the CP level conforms to f-seq, but this is immaterial since the embedded SpecC is not yet a criterial position (movement to SpecC is triggered by an edge feature generated in the derivation, rather than by an inherent feature of C).

(32) a. [$_{V'}$ [$_{DP}$ das Buch]$_{[\Sigma:\Box]}$ [$_{V}$ liest]$_{[\bullet X_V \bullet]}$]
b. [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{V}]}$ [$_{V'}$ [$_{V}$ liest]]]
c. [$_{v'}$ keiner [$_{v'}$ [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{V}]}$ [$_{V'}$ [$_{V}$ liest]]] v$_{[\bullet X_v \bullet]}$]]
d. [$_{vP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{vV}]}$ [$_{v'}$ keiner [$_{v'}$ [$_{VP}$ [$_{V'}$ [$_{V}$ liest]]] v]]]
e. [$_{T'}$ [$_{vP}$ keiner [$_{v'}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{vV}]}$ [$_{v'}$ [$_{VP}$ [$_{V'}$ [$_{V}$ liest]]] v]]] T$_{[\bullet X_T \bullet]}$]
f. [$_{TP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{TvV}]}$ [$_{T'}$ [$_{vP}$ keiner [$_{v'}$ [$_{v'}$ [$_{VP}$ [$_{V'}$ [$_{V}$ liest]]] v]]] T]]
g. [$_{C'}$ [$_{C}$ dass]$_{[\bullet X_C \bullet]}$ [$_{TP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{TvV}]}$ [$_{T'}$ [$_{vP}$ keiner liest v] T]]]
h. [$_{CP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{CTvV}]}$ [$_{C'}$ [$_{C}$ dass] [$_{TP}$ [$_{T'}$ [$_{vP}$ keiner liest v] T]]]]

The subsequent edge feature-driven intermediate movement step in the matrix VP domain is also as in the case of long-distance wh-movement in (28); see (32-j), which gives rise to a Σ-feature list $\boxed{\text{VCTv}}$ that is at variance with f-seq but per se unproblematic because the Williams Cycle is vacuously fulfilled in a non-criterial landing site. However, the movement step to the matrix Specv position, though *structurally* similar to that in the legitimate derivation in (28), is fatal; see (32-l): Movement to Specv gives rise to a list $\boxed{\text{vVCT}}$ on the long-distance scrambled DP's [Σ] feature, and since this last movement step is triggered by an inherent (albeit optional) structure-building feature [$\bullet\Sigma_v\bullet$]

on v, rather than by an edge feature [$\bullet X_v \bullet$], the mismatch between f-seq and the feature list on the moved DP leads to a violation of the Williams Cycle. The derivation underlying (31) also includes further steps: CP extraposition; Merge of T; optional EPP-driven movement of the matrix subject DP to SpecT (cf. Grewendorf (1989)); and Merge of C. However, these steps can be ignored here: The derivation crashes after the scrambled item has reached its criterial position in the matrix vP, and the Williams Cycle is violated.

(31)
i. [$_{V'}$ [$_{CP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{CTvV}}]}$ [$_{C'}$ [$_{C}$ dass] [$_{TP}$ keiner liest v T] [$_{V}$ glaubt]$_{[\bullet X_V \bullet]}$]]]
j. [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{VCTv}\cancel{\text{V}}}]}$ [$_{V'}$ [$_{CP}$ [$_{C'}$ [$_{C}$ dass] [$_{TP}$ keiner liest v T] [$_{V}$ glaubt]]]]]
k. [$_{v'}$ Karl [$_{v'}$ [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{VCTv}\cancel{\text{V}}}]}$ [$_{V'}$ [$_{CP}$ dass keiner liest] [$_{V}$ glaubt]]] v$_{[\bullet\Sigma\bullet]}$]]
l.



Note that the same consequence arises if long-distance scrambling targets SpecV rather than Specv (which might be an option yielding the same string (31) given that subjects raise to SpecT only optionally in German). The only difference would be a fatal (f-seq-violating) value $\boxed{\text{VCTv}}$ of Σ on DP instead of $\boxed{\text{vVCT}}$.

Furthermore, the present analysis also predicts that a wh-phrase that undergoes long-distance wh-movement cannot be fed by intermediate, feature-driven long-distance scrambling to, say, Specv (as opposed to using Specv as an escape hatch provided by an edge feature). Of course, the question arises as to how the two options (which yield identical strings *and* identical structural representations throughout) can be distinguished. It has often been proposed that the absence of (strong) superiority effects with clause-bound wh-movement in German, and the absence of weak crossover effects with clause-bound wh-movement in German, can be traced back to the option of an intermediate scrambling operation because scrambling is independently known to be able to circumvent these effects; see Fanselow (1996) and Grohmann (1997) for superiority effects, and Grewendorf (1988) for weak crossover effects. In the present approach, which recognizes promiscuous escape hatches and thus cannot, e.g., simply equate the Specv position with a scrambling position, this implies that checking of [$\bullet\Sigma_{\mathrm{v}}\bullet$] gives rise to certain properties, like absence of weak crossover effects and absence of superiority effects, whereas checking of a pure edge feature [$\bullet\mathrm{X}_{\mathrm{v}}\bullet$] (in the same position), or checking of [$\bullet\mathrm{wh}_{\mathrm{C}}\bullet$] does not.[15]

Thus, on this view, clause-bound wh-movement in (33-a) does not induce a superiority effect, and clause-bound wh-movement in (33-b) does not trigger a weak crossover effect (for most speakers), because the wh-phrase that is in the criterial SpecC position on the surface (*was*$_2$ ('what') in (33-a), and *wen*$_1$ ('whom') in (33-b)), has undergone an intermediate movement step to a Specv position in the same clause by virtue of an optional inherent feature [$\bullet\Sigma_{\mathrm{v}}\bullet$] on v (and a matching movement-related [Σ]-feature on the DP); and not by virtue of [$\bullet\mathrm{X}_{\mathrm{v}}\bullet$] on v.

(33) a. (Ich weiß nicht) [$_{\mathrm{CP}}$ was$_2$ C wer$_1$ t$_2$ gesagt hat]
I know not what$_{acc}$ who$_{nom}$ said has

b. [$_{\mathrm{CP}}$ Wen$_1$ mag seine$_1$ Mutter t$_1$ nicht] ?
whom likes his mother not

[15] The question of why exactly scrambling – conceived of as checking of [$\bullet\Sigma_{\mathrm{v/V}}\bullet$] – has these consequences is immaterial in the present context. Still, for weak crossover effects, one may assume that checking of [$\bullet\Sigma_{\mathrm{v/V}}\bullet$] can provide (what used to be called) an A-binder for a pronoun that needs to be interpreted as a bound variable (see Heim & Kratzer (1998)); and for superiority effects, one may postulate – as is in fact done by Fanselow and Grohmann – that scrambling can systematically avoid MLC effects.

Given that discharge of a movement-inducing (edge or inherent) feature on the VP and vP levels in (33-ab) involves a valuation of *both* movement-related features on the affected DP (viz., [Σ:$\square$] and [wh:$\square$]), the list of [Σ:$\square$] needs to conform to f-seq on the vP level (which it does: [Σ:$\boxed{\text{vV}}$]); and the list of [wh:$\square$] needs to conform to f-seq on the CP level (which it also does: [wh:$\boxed{\text{CTvV}}$]).

Against this background, the existence of superiority effects with long-distance wh-movement (see (34-a)) and the existence of weak crossover effects with long-distance wh-movement (see (34-b)) follow without further ado; see Frey (1993), Büring & Hartmann (1994), Fanselow (1996), Heck & Müller (2000b), and Pesetsky (2000). It is the presence of a criterial ([$\bullet\Sigma_{\text{v/V}}\bullet$]-based) configuration that helps to avoid superiority effects and weak crossover effects in German, and since such features cannot be checked by long-distance movement to Specv/V domains (because of the Williams Cycle), superiority effects and weak crossover effects cannot be circumvented.[16]

(34) a. *Wen$_2$ hat wer$_1$ geglaubt [$_{\text{CP}}$ dass der Fritz t$_2$ mag] ?
whom$_{acc}$ has who$_{nom}$ believed that the Fritz likes

b. *Wen$_1$ hat seine$_1$ Mutter gesagt [$_{\text{CP}}$ dass wir t$_1$ einladen
whom has his mother said that we invite
sollten] ?
should

3.7. Super-Raising and Other Local Movement Types

The ban on super-raising as in (35) (cf. (4-b)) can be derived in a similar way.

(35) *Mary$_1$ seems that t$_1$ likes John

[16] Assuming that restructuring infinitives are vPs or VPs in German, and other infinitives have CP status (see, e.g., Fanselow (1991), Haider (2010)), it is correctly predicted that the former constructions permit scrambling to the matrix domain in accordance with the Williams Cycle whereas the latter ones do not. The prediction would then also be that restructuring infinitives, unlike structures involving CP or TP embedding, make it possible to circumvent superiority effects and weak crossover effects. These predictions may ultimately prove tenable, but there are intervening factors of various kinds that blur the picture, both with superiority effects (where a non-identity requirement seems to hold in addition) and with weak crossover effects (where judgements are not uniform, and differences compared with superiority effects have been argued to arise); see Fanselow (1991), Kim & Sternefeld (1997) and Haider (2000b; 2004), among others.

By assumption, the relevant movement-related feature on *Mary* is [EPP]; matrix T bears the corresponding structure-building feature [•EPP•]. Successive-cyclic movement must take place via the embedded TP and CP domains, and via the matrix VP and vP domains. In the final matrix SpecT position where [•EPP•] is discharged by attracting the moved DP, [EPP] on DP has the value $\boxed{\text{TvVC}}$, which fatally violates f-seq (hence, the Williams Cycle) because the symbol C has not yet been removed.

The prohibition against a combination of super-raising to matrix SpecT followed by wh-movement to matrix SpecC is derived in the same way as the prohibition against long-distance scrambling feeding wh-movement; see (36-a) (= (14)) and (36-b): EPP-driven movement to matrix SpecT gives rise to a violation of the Williams Cycle (because $\boxed{\text{TvVC}}$ does not conform to f-seq) which cannot subsequently be made undone by matrix wh-movement.

(36) a. *Who$_1$ seems [$_{CP}$ C t$_1$ will leave] ?
b. *What$_1$ seems [$_{CP}$ that it was said t$_1$] ?

Other movement types that target positions in the TP, vP or VP areas (like Scandinavian object shift, German pronoun fronting, clitic climbing in Romance, and extraposition) also cannot apply long-distance via CP, and for the same reason: When the (criterial) target position is reached, there will at least be an f-seq-violating symbol C on the list of the movement-related feature on the moved item, and so a violation of the Williams Cycle will be unavoidable. Thus, the basic generalization correlating the height of the landing site of a movement type and its ability to apply long-distance highlighted at the beginning of the chapter is derived.

4. Legitimate Long-Distance Scrambling and Super-Raising

There is prima facie counterevidence to the approach to improper movement developed so far, in the form of well-formed cases of long-distance scrambling and super-raising from what look like fairly uncontroversial cases of embedded CPs (or at least from XPs that dominate the embedded vP and TP domains, which is all that is needed to create the problem, given that scrambling and raising target vP-internal and TP-internal positions, respectively).

Thus, long-distance scrambling from CP is an option in languages like Russian (see, e.g., Müller & Sternefeld (1993) and Bailyn (2001)) and Japanese (see Saito (1985) and Grewendorf & Sabel (1999), among many others; Korean and Persian also belong in this group), and the final landing site of the movement in these cases is clearly within the TP domain (or at least it can be; see Takahashi (1993) for a possible exception in Japanese that he accordingly reanalyzes as optional wh-movement), which is unexpected from a Williams Cycle perspective under present assumptions. The following example from (colloquial) Russian taken from Zemskaja (1973) illustrates long-distance scrambling.

(37) Vy [$_{\text{DP}}$ pocylku]$_1$ videli [$_{\text{CP}}$ kak zapakovali t$_1$] ?
you-2.PL parcel$_{acc}$ saw how (they-)wrapped

Similarly, super-raising from CP seems to be available in a number of languages, among them Greek (see Perlmutter & Soames (1979, ch. 43) and Alexiadou & Anagnostopoulou (2002), among others) and Kilega and other Bantu languages (see Obata & Epstein (2011) and references cited there). A Greek example is given in (38) (see Perlmutter & Soames (1979)):

(38) [$_{\text{DP}}$ I kopeles]$_1$ fenonde [$_{\text{CP}}$ na t$_1$ fevgun]
the girls$_{nom}$ seem-3.PL SUBJ leave-3.PL

I take these counterexamples to be real. However, this does not mean that the approach to improper movement developed above needs to be abandoned. Rather, it needs to be modified in such a way that it permits variation to some extent, so that a less fine-grained system of deletion in values of movement-related features can be employed in certain constructions and languages.[17]

I would like to suggest that a key to a solution of the problem posed by data such as those in (37) and (38) is that category features are not ontological primitives, but can be assumed to be composed of combinations of more elementary features (see Chomsky (1970)); their cross-classification yields the standard category labels, and underspecification with respect to these features makes it possible to refer to sets of categories as natural classes in syntactic operations. Thus, Stowell (1981, 21) (based on earlier work by Chomsky) suggests that the primitive features [±N] and [±V] yield the four syntactic categories V, N, A, and P (via cross-classification), as well as natural classes of these categories

[17] Also see Obata & Epstein (2011) for this general strategy; and also note that these counterexamples also raise problems for virtually all other existing analyses of improper movement.

(via underspecification): [+V] = V, A; [–V] = N, P; [+N] = N, A; [–N] = V, P.

Suppose now that the categories C, T, v, and V are composed of primitive features in such a way that C and v form a natural class, and T and V form a natural class. Following Chomsky's (2000) original motivation for phases, it can be postulated that the relevant feature is [$\pm\pi$], where π stands for propositionality (in an extended sense): C and v are characterized by [+π], and T and V are characterized by [–π].[18] The crucial assumption now is that deletion in the lists of movement-related features may not have to (but always can) apply under full identity in all languages; "categorial information" in the sense of (22-b) may refer only to a small (but fundamental) part of the category label, viz., information related to the [$\pm\pi$] status of the phase head. Given this assumption, there are four possibilities: First, the full feature set making up a category always needs to be considered in order to find out whether deletion in feature sets applies. This is the option assumed so far throughout the chapter: A category label values the movement-related feature, and deletion of category information takes place only under full identity (i.e., the symbol V deletes an earlier V, and so forth). Second, another option is that only [$\pm\pi$] needs to be shared for deletion in feature sets to apply. This has drastic consequences for improper movement. An edge feature with the categorial information T will now delete a V symbol in a buffer (and vice versa), and an edge feature with the categorial information C will delete a v symbol (and vice versa). The effects are illustrated in (39-abc), for long-distance movement to the VP, vP, and TP domains, respectively. A language that chooses this option can have both long-distance scrambling (to Specv) and super-raising (to SpecT); Russian might be a case in point.[19]

[18] Other features will then also have to be present to distinguish C from v, T from V, V from N, v from n, functional from lexical categories, and so on, but since these features will not play a role in the analysis that follows, I disregard them here. – Also note that the present reasoning does not imply that only C and v qualify as phase heads in the sense of the PIC; they are just the phase heads characterized by propositionality.

[19] Note that in addition to cases of long-distance scrambling such as (37), there are indeed constructions in Russian that look like instances of super-raising; however, see Stepanov (2007) for qualifications.

(39) [+π] deletes [+π], [–π] deletes [–π] (plus deletion under identity)

a. * $\boxed{\mathrm{V}_{[-\pi]}\mathrm{C}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ *f-seq

b. $\boxed{\mathrm{v}_{[+\pi]}\mathrm{V}_{[-\pi]}\cancel{\mathrm{C}}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ $\surd$f-seq

c. $\boxed{\mathrm{T}_{[-\pi]}\mathrm{v}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}\cancel{\mathrm{C}}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ $\surd$f-seq

The third possibility is that [+π] suffices for deletion to apply in cases of categories that are not fully identical; [–π], in contrast, does not. (As before, deletion under full identity is also still available.) This gives rise to a system of improper movement that is more liberal than the first option (requiring full categorial identity) but more restrictive than the second option. Now C and v delete one another, but T and V do not. The effects are shown in (40-abc) for long-distance movement to the VP, vP , and TP domains: Super-raising is possible, but long-distance scrambling is not. This might characterize the situation in Greek.[20]

(40) [+π] deletes [+π] (plus deletion under identity)

a. * $\boxed{\mathrm{V}_{[-\pi]}\mathrm{C}_{[+\pi]}\mathrm{T}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ *f-seq

b. * $\boxed{\mathrm{v}_{[+\pi]}\mathrm{V}_{[-\pi]}\cancel{\mathrm{C}}_{[+\pi]}\mathrm{T}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ *f-seq

c. $\boxed{\mathrm{T}_{[-\pi]}\mathrm{v}_{[+\pi]}\mathrm{V}_{[-\pi]}\cancel{\mathrm{C}}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ $\surd$f-seq

Finally, a fourth option might be that it is [–π] (rather than [+π]) that suffices for deletion to apply in cases of categories that are not fully identical. For (criterial) movement to the matrix VP, vP, and TP domains, this makes predictions that are extensionally equivalent to the first possibility (where only full identity leads to deletion in feature lists) under a C/T/v/V clause structure; cf. (41-abc).

[20] Such an analysis does not by itself correlate the availability of super-raising in a language with some other, independently established property. Obata & Epstein (2011) devise an analysis according to which Kilega and other Bantu languages permit super-raising ultimately because case of the moved item is checked in the embedded clause, and ϕ-features are checked in the matrix clause. However, in super-raising constructions in Greek, the opposite is the case (see Alexiadou & Anagnostopoulou (2002)): ϕ-features are checked in the embedded clause, and case is assigned in the matrix clause. This state of affairs would seem to suggest that an independent factor related to case or ϕ-features cannot easily be identified; and whereas Kilega super-raising is problematic from an Actitivity Condition point of view (as Obata & Epstein (2011, 139) note), Greek super-raising is problematic for the feature splitting approach.

(41) [–π] deletes [–π] (plus deletion under identity)

a. * $\boxed{\mathrm{V}_{[-\pi]}\mathrm{C}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\mathrm{v}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ *f-seq

b. * $\boxed{\mathrm{v}_{[+\pi]}\mathrm{V}_{[-\pi]}\mathrm{C}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ *f-seq

c. * $\boxed{\mathrm{T}_{[-\pi]}\mathrm{v}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}\mathrm{C}_{[+\pi]}\cancel{\mathrm{T}}_{[-\pi]}\cancel{\mathrm{v}}_{[+\pi]}\cancel{\mathrm{V}}_{[-\pi]}}$ *f-seq

As for option (41), it is not clear whether it is actually needed (given that reference to the full categorial information is also an option – arguably the default option –, as assumed throughout this chapter). A relevant case to look at in this context is ECM constructions. As observed by Abels (2008) and Bader (2011), a strict interpretation of the Williams Cycle is problematic if exceptional case marking (ECM) constructions are analyzed in terms of raising to object position (cf. Postal (1974)), rather than in terms of truly exceptional case assignment by a matrix verb to an embedded infinitival subject (as in Chomsky (1981)), and if ECM complements are TPs.

Given the present implementation, the reason is that the relevant movement-related feature on the raised object (whatever this ultimately turns out to be; see above on similar issues with EPP-driven movement to SpecT) would then end up having a value $\boxed{\mathrm{vVT}}$ (with the symbols v and V assigned in the infinitive deleted by movement through matrix SpecV to matrix Specv). Since, by assumption, Specv is a criterial position, such a syntactic buffer would violate the Williams Cycle.[21] However, it is unclear whether ECM constructions should be analyzed via raising to object; the literature contains arguments both for and against such a view.

Thus, Stowell (1991) notes that adverbs which uncontroversially belong to the matrix clause cannot intervene between the DP merged as an external argument of the embedded verb and the rest of the infinitive in English; see (42-abc) (where (42-b) is well formed only if the the adverb *repeatedly* is construed with the embedded clause, an option that does not arise with *sincerely* in (42-c)). This is an argument for exceptional case marking, and against raising to object.

(42) a. John promised repeatedly to leave
b. #John believed Mary repeatedly to have left (She left repeatedly.)

[21] Raising to object cannot possibly be assumed to reach the TP domain, as would be required to circumvent a Williams Cycle violation.

c. *John believed Mary sincerely to have left

On the other hand, Lasnik (1999) points out that sentences like (43) in English permit anaphoric binding of *each other* by *the defendants*. Unless further assumptions are made (e.g., about the role of linear precedence in the licensing of reflexives and reciprocals), this would seem to suggest that the latter DP has undergone movement from the infinitival clause into the matrix clause, crossing the adverbial expression *during each other's trials* and feeding Principle A satisfaction. This piece of evidence would thus seem to support a raising to object analysis, but not an approach to exceptional case marking.

(43) ?The DA proved the defendants$_1$ to be guilty during each other$_1$'s trials.

Taken together, it seems fair to conclude that there is no uncontroversial case for raising to object in English to be made yet. Still, for the sake of the argument, let us assume that ECM constructions in English do indeed involve raising to object. Then, as noted, under present assumptions, the feature list $\boxed{\text{vVT}}$ on the movement-related feature of a moved item in a criterial position that invariably arises if only full categorial identity can lead to symbol deletion in feature lists (assuming raising to object and a TP status of the infinitival complement) would induce a violation of the Williams Cycle. However, if option (41) is adopted, such raising to object will create a feature list $\boxed{\text{vV}}$ that is in accordance with the Williams Cycle: First, the original V is deleted by an incoming T, and secondly, the V information resulting from valuation in the matrix VP suffices to delete T in the feature list. (Finaly, v deletes the lower v symbol.) Still, movement that crosses C and ends up in a position below C in the matrix clause will continue to be rule out as ungrammatical.

To sum up this subsection, in light of languages that permit long-distance scrambling and super-raising from a CP, somewhat less restrictive versions of the Williams Cycle can be introduced alongside the original approach. I have suggested that languages can choose whether full identity of the categorial information is required for symbol deletion in feature lists on moved items, or whether identity of a major subfeature [π] (encoding propositionality of a phase head) of the full categorial information also suffices. Ultimately, the question to what extent individual languages make use of the resulting more liberal systems of improper movement can only be addressed by in-depth empirical studies of the relevant constructions; this is beyond the scope of the present chapter.

5. A Further Extension: DP-Internal PP Preposing

Finally, I will discuss a possible further extension of the present analysis. It concerns a movement type that affects a DP-internal position.[22] German has a movement operation that involves PPs and targets SpecD (see Lindauer (1995)); in what follows I will refer to this as "DP-internal PP preposing". The construction is usually considered slightly substandard, but it is fully productive. A relevant example is given in (44).

(44) $[_{\mathrm{DP}_2}$ $[_{\mathrm{PP}_1}$ Über die Liebe] $[_{\mathrm{D}'}$ das/ein Gerücht t_1]] kenne ich t_2
about the love the/a rumour know I

So far, I have been silent on whether f-seq should be assumed to comprise both the clausal and the nominal domain, or whether two separate f-seqs should be postulated. Suppose now that the former option is pursued, and, more specifically, that the comprehensive f-seq is CTvV<u>DNP</u>. This reflects the fact that C (rather than D) is the root node, and that nominals are typically parts of clauses.[23] Under this assumption, DP-internal PP preposing in local contexts, as in (44), is inherently unproblematic from the perspective of improper movement: Given that there are designated movement-inducing and movement-related features triggering DP-internal PP preposing (say, [$\bullet\omega_{\mathrm{D}}\bullet$] on D, and matching [$\omega$:☐] on the PP), the list on the movement-related feature on the moved PP in the criterial SpecD target position respects the Williams Cycle: [ω:$\boxed{\mathrm{DN}}$]. Similarly, simple cases of extraction of some item from DP into the embedding clause will be unproblematic from an improper movement perspective because the extended f-seq will be maintained.

---

[22] I hasten to add that what follows is tentative, and in some sense orthogonal to my main concern here, which has been to show that a local reformulation of one specific theory of improper movement – viz., the Williams Cycle-based approach – becomes possible if one adopts the concept of syntactic buffers.

[23] As a matter of fact, since V can also embed PP rather than DP, there also has to be a second option for f-seqs comprising verbal and non-verbal domains: The full f-seq is either CTvVDNP or CTvVPDNP, i.e., f-seqs must be branching. This does not affect the argument to be made below.

However, things should be different for Complex Noun Phrase Condition (CNPC) contexts, where a DP embeds a CP, and some PP item is extracted from within CP to end up in SpecD, as an instance of DP-internal PP preposing. Such constructions should always violate the Williams Cycle, in contrast to long-distance PP wh-movement that goes into the matrix clause. In addition (and somewhat less interestingly from the present perspective), long-distance PP scrambling is also predicted to violate the Williams Cycle, just like any other case of long-distance scrambling (see section 3.6.).

These predictions are borne out. Consider first (45), an instance of long-distance topicalization of a PP from an argument CP embedded in an object DP. The example has a degraded status, but it is generally not perceived as completely impossible, as is typical of CNPC violations with argument extraction.[24] However, (45) does not violate the Williams Cycle: In the matrix SpecC position targetted by topicalization, the PP *über die Liebe* ('about the love') has its [top] feature valued as $\boxed{\text{CTvVDN}}$, which is in accordance with the extended f-seq.

(45) ?*[$_{\text{PP}_1}$ Über die Liebe] kenne ich [$_{\text{DP}}$ das/ein Gerücht [$_{\text{CP}}$ dass sie ein
about the love know I the/a rumour that she a
Buch t_1 geschrieben hat]]
book written has
$\surd$[top: $\boxed{\text{CTvVDN}\cancel{\text{CTvVDN}}}$]

Consider next (46), which involves DP-internal PP preposing from within the CP to the SpecD position. This example is completely ungrammatical, much more so than one would expect if only a standard CNPC effect were involved; in particular, the contrast to (45) is striking.[25] This follows from the present version of the Williams Cycle: Movement originates in a CP; hence, given the extended f-seq, in must not end in a DP domain but needs to target the matrix CP domain again. PP_1 in (46) fails to do this; consequently, the movement-

---

[24] For present purposes, I leave open how CNPC island effects should be derived, but I assume that it is CP (not DP) that creates the island. I will return to this issue in the next chapter.

[25] Of course, the string as such is not excluded; if PP_1 is construed with *Gerücht* ('rumour'), as in *Gerücht über die Liebe*, the example becomes fully acceptable (if somewhat weird since it strongly suggests an allegorical interpretation of *Liebe* ('love')). Strong illformedness only results under the reading indicated here, where PP_1 is construed with *Buch* ('book'), as in *Buch über die Liebe*.

related feature, which is valued as $\boxed{\text{DNCTvV}}$ in the criterial position, violates the Williams Cycle, and the derivation crashes.

(46) *[$_{DP}$ [$_{PP_1}$ Über die Liebe] [$_{D'}$ das/ein Gerücht [$_{CP}$ dass sie ein Buch
about the love the/a rumour that she a book
t$_1$ geschrieben hat]]] kenne ich
written has know I *[ω: $\boxed{\text{DNCTvV}\cancel{\text{DN}}}$]

Finally, in (47), PP undergoes long-distance scrambling in a CNPC configuration. As with other cases of long-distance scrambling in German, the Williams Cycle is violated, and the construction is thus correctly predicted to be much more ill formed than one would expect if only a CNPC effect were involved.

(47) *Es kennt [$_{PP_1}$ über die Liebe] keiner [$_{D'}$ das/ein Gerücht [$_{CP}$
EXPL knows about the love no-one the/a rumour
dass sie ein Buch t$_1$ geschrieben hat]]]
that she a book written has *[Σ: $\boxed{\text{vVDNCT}\cancel{\text{vVDN}}}$]

6. Conclusion

The main result of this chapter is that it is possible to come up with a theory of improper movement in a local–derivational approach to syntax in which phrase structure is generated bottom-up, only small parts of syntactic structure are accessible at any given step of the derivation, and look-ahead and backtracking are not theoretical options. This goal can be achieved by (a) assuming syntactic buffers as values of movement-related features, and (b) reformulating the Williams Cycle, a constraint on improper movement that has been argued for in Williams (1974; 2003), Sternefeld (1992), Grewendorf (2003; 2004), Abels (2008), Neeleman & van de Koot (2010), Bader (2011), and Keine (2014). In the existing analyses where a version of the Williams Cycle is put to use, it is generally formulated in a non-local way, such that large amounts of syntactic structure must be scanned in order to decide whether a given interaction of movement steps counts as improper or not (though see footnote 6). In contrast, in the reformulation that I have suggested, all relevant pieces of information are locally available; in order to determine whether there is an improper movement configuration or not, no more structure needs to be considered than the moved item itself. In addition to the *locality problem*, the new formulation of the Williams Cycle also solves two other problems for existing approaches to improper movement, viz., the *generality problem* and the *promiscuity problem*. The generality problem does not arise because the Williams Cycle applies to all kinds of movement; and, perhaps most importantly, the promiscuity problem (which consists in the fact that massive use of intermediate landing sites is difficult to reconcile with the characterization of these same landing sites as specific for certain kinds of movement) is solved by assuming that the relevant (categorial) information of the domains that a moved item passes through is successively picked up and registered in a buffer on the moved item but can subsequently be deleted again if identical information is read in; only when a criterial position is reached does the Williams Cycle spring into action and determine whether movement has been improper or not, by checking the list of categorial information on the movement-related feature of the moved item and comparing it with the functional sequence (f-seq).

The theoretical machinery needed to implement this approach is, I think, innocuous, and to a significant extent independently motivated: Given that edge features are needed to bring about intermediate movement steps, it looks as though the simplest solution to the problem of how to generate them that is compatible with the Inclusiveness Condition is to copy the label of the phase head; and to make the resulting feature usable at all, it has to be stripped off its original content, which nonetheless is retained as an index on the newly generated edge feature. The central remaining assumption that is not (as far as I can tell) independently motivated then is that the movement-related features on items that need to undergo displacement have (initially) empty lists as values which are successively filled by the categorial information on edge features (as regular instances of feature valuation), subject to the requirement that a symbol on the list is deleted once an identical symbol is read in. Here, there is no disjunctive ordering, as it is typically assumed for other kinds of feature valuation.[26]

Against the background of the assumptions that I have made, there are various other possibilities for implementing a local version of the Williams Cycle, sometimes with slightly different empirical consequences, sometimes not. For instance, one might want to do away with the assumption that the value of

[26] That said, instances of case stacking and constructed plural in the world's languages arguably also require multiple incremental valuation of a single feature; see Nordlinger (1998), Richards (2012), and Adger (2003), Trommer (2006), respectively.

movement-related features is a (potentially multi-membered) list, and replace it with the hypothesis that it is a single category symbol. Let me briefly consider how this idea can be executed. Suppose first that every incoming symbol replaces an existing symbol as the value of the movement-related feature; thus, the deletion operation based on identity is dispensed with in favour of a more general deletion operation. However, other things being equal, this would create problems for ruling out long-distance scrambling to vP (where the highest value would be V, yielding vV, in line with f-seq) and super-raising to TP (where the highest value would be v, yielding Tv, again in line with f-seq).

Consider next the possibility that it might suffice to keep track of the topmost symbol in an f-seq, in the sense that only the highest label of an f-seq that a moved item has passed through is maintained. Such an approach would work for most cases, but there is a potential conceptual problem because f-seq is employed *twice* in the account of improper movement: (i) to determine whether a new symbol can replace an existing one as the value of a movement-related feature on a moved item (e.g., C replaces T, but matrix V does not replace C); and (ii) to check the Williams Cycle in criterial positions (e.g., so as to distinguish between criterial movement to matrix vP and criterial movement to matrix CP). In addition, given the approach to cross-linguistic variation in section 4, there is an empirical problem with this proposal: In (40-b), T (rather than the highest head of the f-seq in the embedded clause, viz., C) is the incriminating symbol on a feature list, which blocks long-distance scrambling while permitting super-raising.

Second, one might give up the assumption that every phrase is a phase (see section 3.1. and chapter 1), or, more specifically, that edge feature-driven intermediate movement steps leading to feature valuation with category information on the moved item occur in every phrase between the base position and the criterial position. Assuming, for instance, that only CP and vP trigger intermediate movement steps whereas TP and VP do not (or, in fact, that *only* SpecC is an intermediate landing site), a system would result in which many cases of improper movement could still be excluded by the Williams Cycle (as involving an illegitimate list $\boxed{\text{vC}}$ that violates f-seq). Empirically, such an approach would make predictions that are by and large identical to those of the present approach (assuming that criterial movement, unlike intermediate movement, may also target SpecT and SpecV positions, and thus be able to activate the Williams Cycle). However, it may be viewed as conceptually inferior since feature lists would then contain only a part of the information that the f-seqs contain against which they are measured.

Here is a third alternative: Assuming that the movement-related feature on the moved item is associated with a unique position on f-seq, one might adopt a version of the Williams Cycle stating that a movement-related feature on a moved item cannot license (intermediate) movement to any category that is higher on f-seq. There are several potential problems with such an approach, though: First, movement-related features are in fact not necessarily associated with a particular f-seq position (this holds, e.g., for [D] if this is the feature sought by an EPP T, and for [Σ], given that scrambling can target VP, vP and perhaps even TP in German). Second, movement-related features are not necessarily criterial (this distinguishes Attract-based from Greed-based approaches to movement, with the present account belonging to the former); but the presence of, say [wh] on a wh-phrase that does not undergo wh-movement in a multiple question does not license long-distance scrambling of this wh-phrase. A separate (but related) third problem is that if there is more than one movement-related feature on a moved XP, this approach would make a wrong prediction: A feature like [wh] on a long-distance scrambled XP cannot render long-distance scrambling (before wh-movement) well formed (see (34) above). Fourth, this account is not sufficiently flexible as it stands (cf. the previous section's remarks on licit long-distance scrambling and super-raising). To accomodate these problems, the Williams Cycle would minimally have to be reformulated in such a way that it states that there is no unchecked movement-related feature F such that the position P_1 in which F will ultimately be checked is lower on f-seq than the position P_2 just reached. However, even such a more complex formulation (involving look-ahead) would not evade what I take to be the most pressing conceptual problem with this proposal: The underlying logic of the present analysis would be stipulated, not derived. Thus, intermediate steps (and the history of movement) would be irrelevant, and whereas in the present approach it always suffices to look at the moved item alone in order to determine whether an improper movement configuration is present or not, on the alternative approach one has to compare the moved item's properties with the properties of its syntactic environment. In line with this, the idea of temporary improper movement configurations in the derivation that can be tolerated for a while but must eventually be gotten rid of would play no role – but see Heck & Müller (2013) for an independent argument that derivations recognize, and try to repair as quickly as possible, temporary improper movement configurations; also see chapters 3 and 4 for cases that crucially rely on the presence of more than just the value $+$ in the movement-related feature (which is also incompatible with the first two alternative approaches mentioned in the main text.)

In view of all these problems with alternative approaches, the analysis developed above strikes me as a fairly straightforward one because it is both fine-grained and potentially flexible, and (not least of all) because it structurally assimilates the central operation (viz., creating buffers) to other syntactic operations (in particular, to Agree). Needless to say, the proposed analysis of improper movement also gives rise to a number of further questions, and may suggest a number of extensions. However, for reasons of space and coherence, and since the main goal throughout is to highlight the role of syntactic buffers in a strictly derivational approach to syntax, I will not delve into improper movement as such any further at this point. However, I will pursue two issues that will turn out to be closely related to the account of improper movement given in the present chapter, both with respect to the problems they raise for a local–derivational approach (viz., backtracking issues), and with respect to the proposed solution (viz., syntactic buffers): In chapter 3, I consider remnant movement constructions; and in chapter 4, I turn to resumption.[27]

[27] As a matter of fact, Grewendorf (2003; 2004) and Abels (2008) suggest that a Williams Cycle-based approach to improper movement can be modified so as to not only cover cases of two operations applying to a single item α, but also cases of remnant movement or, more generally, two operations applying to two different items α, β that are initially in a dominance relation ($[_{\alpha}$ ... β ...]), with both items eventually targetting an α-external position. This presupposes an extended version of (10) according to which movement of either α or β in $[_{\alpha}$ ... β ...] is reflected on the remaining item. I address these proposals in the appendix of the next chapter.

Chapter 3

Remnant Movement

1. A Non-Local–Derivational Approach: Bleeding and Counter-Bleeding

1.1. Remnant Movement

Like the constructions discussed in subsection 2.3 of chapter 1 (*wanna*-contraction, movement and reflexivization, ergative vs. accusative movement), remnant movement constructions like the German example in (1) have been taken as an argument for a derivational approach to syntax since they exhibit opacity – more specifically, a counter-bleeding effect (see Müller (1998)).

(1) [$_{\text{VP}_2}$ t_1 Gelesen] hat das Buch$_1$ keiner t_2
read has the book no-one

(1) involves a combination of two movement operations. First, there is scrambling from VP (to a pre-subject position which is arguably a specifier of vP – it follows the position occupied by weak pronouns in German but can precede a subject DP that has not undergone optional movement to SpecT). Second, there is remnant VP topicalization to SpecC (or to the specifier of some other functional head in the left periphery). More generally, remnant movement constructions are characterized by a pre-movement α-over-β configuration, as in (2), in which both α and β undergo separate movement operations (and β targets an α-external position).

(2) ... [$_\alpha$ ... β ...] ...

Movement in α-over-β configurations of this type has a number of conspicuous properties. Assuming that all movement operations leave traces (or copies), remnant movement (i.e., movement of α in the presence of β-movement in (2)) creates an unbound trace (or copy). Accordingly, the properties of remnant movement constructions can be accounted for by postulating specific *restrictions* on unbound traces (see Thiersch (1985), den Besten & Webelhuth (1987; 1990), Müller (1993), Grewendorf & Sabel (1994), Saito (2003), Collins & Sabel (2007) for attempts along these lines). However, it is of course a priori preferable to account for the properties of multiple movement in α-over-β configurations without invoking designated constraints referring to unbound traces; such constraints will eventually qualify as construction-specific. In fact, from a current, third-factor-based minimalist perspective (see Chomsky (2007; 2008; 2013), it would seem to be impossible to postulate constraints referring to traces (like the Empty Category Principle (ECP) of Chomsky (1981)) – let alone constraints referring to *unbound* traces; and this conclusion holds independently of whether displacement in syntactic derivations is or is not assumed to leave a trace (copy) in the first place (see Epstein & Seely (2002) and chapter 1 above).

As argued in Müller (1998), it is possible to come up with a straightforward, reasonably simple analysis of three conspicuous properties of remnant movement constructions (which I will call *freezing*, *anti-freezing*, and *Müller-Takano generalization*) that does not have to resort to concepts like "unbound trace", provided that a *derivational* (rather than *declarative*) approach is adopted, which is also *non-local* in essential respects. I address the relevant generalizations and the derivational analysis in the next subsection.[1]

[1] Throughout this chapter, I presuppose that examples like (1) do indeed involve remnant movement from an α-over-β configuration, with scrambling preceding topicalization, and that the three properties discussed in the next subsection do indeed properly characterize α-over-β configurations. Both assumptions have been called into question. Thus, De Kuthy & Meurers (2001), Fanselow (2002), Hale & Legendre (2004) and Thoms & Walkden (2013), among others, argue against a remnant movement analysis of examples like (1). In contrast, Grewendorf (2003; 2004) and Abels (2008), while basically assuming a remnant movement analysis of these kinds of data, adopt some generalizations about α-over-β configurations that are not co-extensive with those I adopt in what follows (involving, e.g., systematic exceptions from freezing and anti-freezing). For reasons of space and coherence, and since the main focus of the present chapter is on a reconstruction of an existing theoretical analysis in a slightly different (i.e., more local) framework rather than on justifying this analysis, I will not discuss these alternative proposals at this point; but

1.2. *Three Properties*

1.2.1. *Freezing*

The first relevant generalization is that movement of α and β in α-over-β configurations as in (2) induces a *freezing* effect if movement of β ends up in a higher position than movement of α; i.e., if the trace of β (assuming there is one) is bound in the final representation: A trace in a moved item leads to illformedness when its antecedent is outside of the moved item and c-commands the trace. Relevant German data illustrating this interaction of movement operations are given in (3-a) (with VP topicalization accompanied by wh-movement from VP) and (3-b) (with VP scrambling accompanied by wh-movement from VP).[2]

(3) a. *Was$_1$ denkst du [$_{\mathrm{VP}_2}$ t$_1$ gelesen] hat keiner t$_2$?
    what think you read has no-one

  b. *Was$_1$ hat [$_{\mathrm{VP}_2}$ t$_1$ gelesen] keiner t$_2$?
    what has read no-one

Freezing effects as in (3) instantiate a case of transparent (rather than opaque) interaction of operations. More precisely, we are dealing with a *bleeding* effect here: XP$_2$ movement bleeds XP$_1$ movement. Consequently, deriving freezing effects in α-over-β configurations is unproblematic under both a declarative and a derivational approach (provided the approach is not strictly local; see below): Recall that it is only *opaque* interactions like counter-bleeding and counter-feeding that distinguish derivational from declarative approaches (see chapter 1). As shown by Browning (1991), in a declarative approach it suffices to look at the output representations in (3-ab) to correctly determine illformedness. Given a constraint like the Condition on Extraction Domain (CED) in (4)

---

see the appendix below (pp. 99-121). For present purposes, suffice it to say that I think there is neither strong empirical evidence against remnant movement in German, nor strong evidence for relativizing the generalizations that I will now turn to in the main text.

[2] On freezing, see Ross (1967) and Wexler & Culicover (1980), among others. Culicover & Winkler (2010) argue that all freezing effects can be traced back to processing difficulties, and that they can typically be improved “with context and prosody”. While this may be true for some of the cases that have been discussed in the literature under the label ‘freezing’ (and that may not involve genuine extraction), I do not concur with this assessment in the case of examples like those in (3).

(see Huang (1982), Chomsky (1986), Browning (1987), Cinque (1990)), VP_2 in (3) qualifies as a barrier between t_1 and its antecedent *was*$_1$ ('what') (since VP_2 shows up in a specifier position), and movement therefore illegitimately crosses a barrier.

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

Similarly, in a derivational approach, the freezing effect can be straightforwardly derived: Given the Strict Cycle Condition (SCC; see Chomsky (1973), Perlmutter & Soames (1979), among many others), a moderately updated version of which is presented in (5) (= (i) of footnote 2 in chapter 2, with variable names changed to avoid confusion with the variable names α, β used here for the two items involved in remnant movement configurations), movement of XP_2 (which targets a lower position) must precede movement of XP_1 (which targets a higher position) in (3). Consequently, movement of XP_1 takes place from XP_2 when the latter has already become a barrier, and a CED violation will be unavoidable (see (4)): Movement of α bleeds extraction of β from α.[3]

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

[3] There is a qualification, though, that foreshadows the problems with a *local*–derivational approach I will address below. As observed by Collins (1994), to derive illformedness in cases like those in (3), it must be ensured that a derivation in terms of 'chain interleaving' is blocked where β is moved from α to an intermediate (scrambling) position first, α is moved to its target position next, and β is finally moved from the intermediate position to its ultimate landing site. This derivation respects both the CED and the SCC. Collins (1994) argues that it can be excluded by (transderivational) economy considerations; in Müller (1998, ch. 4), I suggest that such a derivation via chain interleaving is ruled out as an instance of a constraint on improper movement (the Principle of Unambiguous Binding (PUB), see chapter 2), which I argue to be derivable from a combination of derivational and transderivational economy constraints. With transderivational constraints being widely considered dubious nowadays, the unwanted chain interleaving derivation remains problematic, irrespective of whether a phase-based approach is adopted or not.

1.2.2. Anti-Freezing

The second generalization about movement in α-over-β configurations as in (2) is that an *anti-freezing* effect arises if movement of β ends up in a lower position than movement of α; i.e., if the trace of β is not bound in the final representation because remnant movement has applied: A trace in a moved item does not have to lead to illformedness when its antecedent is outside of the moved item and does not c-command the trace. Relevant data from German are (6-a) (= (1), with scrambling from VP accompanied by VP topicalization in the same clause), (6-b) (with scrambling from VP accompanied by long-distance VP topicalization), and (6-c) (with wh-movement from VP accompanied by long-distance VP topicalization).[4]

(6) a. [$_{\text{VP}_2}$ t_1 Gelesen] hat das Buch$_1$ keiner t_2
read has the book no-one

b. [$_{\text{VP}_2}$ t_1 Zu lesen] glaubte sie [$_{\text{CP}}$ t'_2 habe [$_{\text{DP}_1}$ das Buch]
to read believed she has$_{subj}$ the book
keiner t_2 versucht]
no-one tried

c. ?[$_{\text{VP}_2}$ t_1 Zu lesen] weiß ich nicht [$_{\text{CP}}$ was$_1$ sie t_2 versucht hat]
to read know I not what she tried has

In contrast to freezing effects, anti-freezing effects in α-over-β configurations distinguish between declarative and derivational approaches because here, the interaction of operations is opaque. More specifically, there is a *counter-bleeding* effect that can receive a straightforward account in a derivational approach: Given the SCC, movement of XP_2 (which targets a higher position) must follow movement of XP_1 (which targets a lower position). Thus, extraction of XP_1 from XP_2 takes place when XP_2 is still in its base (complement) position, in accordance with the CED; and subsequent movement of XP_2, which turns XP_2 into a barrier, comes too late to prevent extraction. In other words: Movement of α counter-bleeds extraction of β from α. However, from a declarative perspective, anti-freezing poses a problem: By only looking

[4] The last case gives rise to a wh-island effect, but these effects are often not very strong in German, particularly if the item that moves from the wh-island does so by topicalization; see Fanselow (1987).

at the output representations, it is not clear why the sentences in (6) can be grammatical; we should expect a straightforward violation of the CED because t_1 is separated from its antecedent (*das Buch*$_1$ (‘the book’) or *was*$_1$ (‘what’)) by a barrier. The only possible way out in a declarative approach, it seems, would be to artificially enrich the CED in the same way that Barss (1984; 1986) modified Principle A of the binding theory in view of the counter-bleeding effect with moved items that are, or contain, anaphors (cf. the discussion of (15-a) in chapter 1), and this would imply, again, stipulating the effects of constraint interaction (here: the SCC and the CED) by integrating the interaction into the formulation of a single, much more complex, constraint.

1.2.3. Müller-Takano Generalization

The third generalization about movement in α-over-β configurations that I will focus on here is about cases where the two items undergo the same kind of movement. The generalization is due to Müller (1993) and Takano (1994), and has sometimes been referred to as the “Müller-Takano generalization” (see, e.g., Sauerland (1999), Pesetsky (2012)). It states that remnant XPs cannot undergo Y-movement if the antecedent of the unbound trace has also undergone Y-movement, where Y stands for a movement-related feature (like [wh] for wh-movement, [top] for topicalization, [Σ] for scrambling, etc.). The sentences in (7) fall under this generalization. In the German example (7-a), DP scrambling from VP to a position following the subject is accompanied by remnant VP scrambling; in the German example (7-b), DP scrambling from VP to a position preceding the subject is accompanied by remnant VP scrambling (cf. Fanselow (1991), Grewendorf & Sabel (1994), Frank et al. (1992), Stechow (1992), Haider (1993), and De Kuthy & Meurers (2001), among others, for these kinds of data); and in the Japanese example (7-c), long-distance scrambling of DP (which is an option as such in Japanese; cf. chapter 2) is accompanied by remnant CP scrambling (cf. Saito (1992)). In all these cases, ungrammaticality results; i.e., anti-freezing is somehow suppressed.

(7) a. *dass [$_{\text{VP}_2}$ t_1 zu lesen] keiner [$_{\text{DP}_1}$ das Buch] t_2 versucht hat
that to read no-one the book$_{acc}$ tried has

b. *dass [$_{\text{VP}_2}$ t_1 zu lesen] [$_{\text{DP}_1}$ das Buch] keiner t_2 versucht hat
that to read the book$_{acc}$ no-one tried has

c. *[$_{CP_2}$ Mary-ga t_1 yonda-to] [$_{DP_1}$ sono hon-o] John-ga t_2
Mary$_{nom}$ read-COMP that book$_{acc}$ John$_{nom}$
itta (koto)
said fact

Again, there is a very simple analysis in a derivational approach. As has been observed by Kitahara (1994; 1997), Fox (1995), Koizumi (1995), Müller (1998), and others, Müller-Takano generalization effects follow from the Minimal Link Condition. Here is why: Suppose that movement of of XP_2 and XP_1 is triggered by the same movement-related feature, and that XP_2 dominates XP_1. In this case, XP_2 is invariably closer to the attracting head, and must therefore move first; early movement of the lower XP_1 would give rise to a violation of the (generalized) Minimal Link Condition (MLC; the generalization consists of an extension of minimality from c-command to dominance contexts, thereby incorporating a *relativized* A-over-A Principle as it has been proposed in Chomsky (1973), Bresnan (1976b), Fitzpatrick (2002)). Therefore, a CED effect is unavoidable in these contexts; in addition, subsequent movement of XP_1 will also have to violate the SCC if it ends up in a lower position that is included in another phrase. Again, a declarative reconstruction of this analysis (that preserves its gist) does not seem to be forthcoming.

1.2.4. Dilemma

To sum up so far, in a *non-local–derivational* approach like standard Principles-and-Parameters theory, the freezing, anti-freezing and Müller-Takano generalizations for α-over-β configurations can straightforwardly be derived by the unstipulated interaction of three well-established constraints (CED, SCC, MLC) – assuming, that is, that there are no intermediate traces (except perhaps for those in SpecC postulated for non-clause bound movement). Unfortunately, this is not the case anymore if a *local–derivational* approach is adopted, as it is required under more recent minimalist assumptions.

Given the Phase Impenetrability Condition (PIC; Chomsky (2001)), XP movement from a phase can only take place via its specifier, and given that vP and CP are phases, at least some of the relevant movement types will have their eventual landing sites in an area that is beyond the minimal phase in which they originate. The PIC is repeated here from (1) of chapter 1.

(8) *Phase Impenetrability Condition* (PIC; Chomsky (2000; 2001)):
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.

If more XPs qualify as phases, more movement types will have their ultimate landing sites in a higher phase. In what follows, I will in fact assume that *all* phrases are phases (cf. chapters 1 and 2), which will then amplify (but not substantially change) the problem that arises with α-over-β configurations in a strictly local approach. And the problem is this:

In the legitimate cases (anti-freezing), extraction of XP_1 from XP_2 will have to take place immediately to an intermediate phase edge position, *before* XP_2 undergoes an intermediate movement step itself. Thus, suppose that Y is the first phase head above the base position of XP_2, such that XP_1 (β) and XP_2 (α) must both move to SpecY because of the PIC (given a non-recursive definition of edge, pace Richards (2011)). Then, an account of the anti-freezing effect in remnant movement constructions would seem to necessitate a derivation as in (9), with XP_1 moving first to SpecY, and XP_2 moving to another SpecY position after that, in line with the CED.[5]

(9) *Anti-freezing, first intermediate steps*:

a. $[_{Y'}$ Y $[_{XP_2{}^a}$ $XP_1{}^b$ $[_{X_2{}'}$ X_2 ...]]]]
b. $[_{Y'}$ $XP_1{}^b$ $[_{Y'}$ Y $[_{XP_2{}^a}$ t_1 $[_{X_2{}'}$ X_2 ...]]]]
c. $[_{YP}$ $[_{XP_2{}^a}$ t_1 $[_{X_2{}'}$ X_2 ...]] $[_{Y'}$ $XP_1{}^b$ $[_{Y'}$ Y t_2]]]

In contrast, in the illegitimate cases (freezing and Müller-Takano generalization), it looks as though extraction of XP_1 from XP_2 will have to *follow* the first intermediate movement step of XP_2, so as to produce a CED violation. This is shown in (10) and (11), respectively.

(10) *Freezing, first intermediate steps*:

a. $[_{Y'}$ Y $[_{XP_2{}^b}$ $XP_1{}^a$ $[_{X_2{}'}$ X_2 ...]]]]
b. $[_{Y'}$ $[_{XP_2{}^b}$ $XP_1{}^a$ $[_{X_2{}'}$ X_2 ...]] $[_{Y'}$ Y t_2]]
c. *$[_{YP}$ $XP_1{}^a$ $[_{Y'}$ $[_{XP_2{}^b}$ t_1 $[_{X_2{}'}$ X_2 ...]] $[_{Y'}$ Y t_2]]

[5] Here and in the following two subderivations, the letters "a" and "b" stand for different movement-related features that will eventually be checked in a higher position but are not checked in the intermediate positions. The sole purpose of these symbols is to enhance perspicuity; they play no role in the analysis.

(11) *Müller-Takano generalization, first intermediate steps*:

a. $[_{Y'}$ Y $[_{XP_2{}^{a}}$ $XP_1{}^{a}$ $[_{X_2{}'}$ X_2 ...]]]]
b. $[_{Y'}$ $[_{XP_2{}^{a}}$ $XP_1{}^{a}$ $[_{X_2{}'}$ X_2 ...]] $[_{Y'}$ Y t_2]]
c. *$[_{YP}$ $XP_1{}^{a}$ $[_{Y'}$ $[_{XP_2{}^{a}}$ t_1 $[_{X_2{}'}$ X_2 ...]] $[_{Y'}$ Y t_2]]

However, the problem is that it looks as though the decision must be made at a point in the derivation (viz., at the YP level) when the relevant information (concerning where XP_1 and XP_2 will eventually end up, and concerning the question of whether XP_1 and XP_2 will eventually check the same kind of movement-related feature or not) is not yet present: In all three cases, there is just PIC-driven intermediate movement of both XP_1 and XP_2. This problem could only be solved at this point if look-ahead were permitted in syntax (in violation of locality) – but even granted that, there does not seem to be an obvious possibility to evem technically formulate a look-ahead analysis. On the other hand, at a later point, the relevant distinction is lost; both XPs are in specifiers of phase edges from the first cycle onwards, and invoking the CED will be impossible from then on. This problem could only be solved if backtracking were permitted in syntax (in violation of both locality and strict cyclicity), such that information is made available specifying whether some item has moved out of XP_2 at an earlier stage of the derivation, and whether this item has already reached its criterial position.[6]

For these reasons, we end up with a severe dilemma: Core properties of movement in α-over-β configurations do in fact not follow anymore if a local–derivational (e.g., phase-based) approach is adopted. In view of this, I will develop a new analysis, one that does not account for the three generalizations in terms of the CED anymore (because this constraint's activity window in α-over-β configurations comes too early); thus, I do not classify the problem at hand as a look-ahead issue. I take it to be unavoidable (in a local–derivational

[6] One might think that reference to the movement-related feature on the intermediately moved item ("a" and "b" in the above subderivations) could help to avoid these problems. This is not the case. Even if one were to assume that edge features triggering intermediate movement steps to phase edges are sensitive to the nature of the criterial position that the movement will ultimately end up in (see references on page 8 above for such a proposal), this could at best help with the Müller-Takano generalization, not with freezing and anti-freezing: In these latter cases, the crucial factor is the *relative height* of the two criterial positions (which do not even need to be in the same clause), and this information cannot be encoded by a single feature.

approach) that the relevant stage of the derivation where it is decided whether movement in α-over-β configurations is legitimate or not does indeed (potentially) arise only late in the derivation, when a moved item has reached its ultimate, criterial landing site, and the problem at hand then emerges as a backtracking issue since information from earlier derivational stages must be accessed. Against this background, I would like to suggest that crucial aspects of (recent) movement steps in α-over-β configurations are recorded on a *buffer*, in the form of a list that acts as the value of the movement-related feature.

2. A Local–Derivational Approach: Counter-Feeding and Feeding

2.1. Defective Valuation of Movement-Related Features

The starting point of the new analysis is the hypothesis that remnant movement is not completely unproblematic from a theory-internal point of view.[7] On this view, languages ideally want to do without situations in syntax where some item β moves out of a category α that itself needs to undergo movement, because of the intermingling of dependencies and the potential ambiguities that ensue. However, this does not mean that remnant movement is excluded, and that α-over-β configurations are ruled out by stipulation as possible sources of well-formed derivations; it just means that there is a price to pay. For concreteness, I would like to suggest that if β moves out of α in an α-over-β configuration like (2), repeated here as (12), α is *contaminated*: β provides a defective value for α's movement-related feature (e.g., [wh], [top], [Σ]), which invariably brings about a crash of the derivation if it is not removed in time, before a criterial position is reached; thus, the movement-related feature acts as a buffer that stores a crucial aspect of an earlier part of the syntactic derivation.

(12) ... [$_{\alpha}$ ... β ...] ...

---

[7] This contrasts with Stabler (1999) and Koopman & Szabolcsi (2000), where it is assumed that remnant movement is completely innocuous; Stabler's argument centers around the size of derivations with and without remnant movement. However, Kobele (2010) (in fact elaborating on an earlier conjecture in Stabler (1999)) shows that remnant movement increases generative capacity, and I take this to be indicative of the problems with remnant movement presupposed in the main text.

This does not keep α (or β) from undergoing movement itself; a *temporary* contamination of a movement-related feature is unproblematic as long as a criterial position has not yet been reached. A moved item β can in principle *decontaminate* a category α again by removing the defective symbol; but this only happens when β itself reaches a criterial position, under c-command. Thus, the *timing* of movement steps of α and β will be crucial. Criterial remnant movement of α is legitimate if β has been able to remove the fatal value from α's feature list before the criterial movement step; otherwise criterial remnant movement of α is illegitimate.

At this point, two questions arise. First, what is this feature value that turns an XP α from which extraction of β has taken place into an illegitimate item? And second, why does such a feature value of a movement-related feature on α lead to illformedness unless it is deleted before α reaches a criterial position? The answer to the first question that I would like to give here is that the incriminating feature value that α gets from β when β moves out of α is β's *index*.[8] Consequently, contamination and decontamination of movement-related features in α-over-β configurations can be defined as in (13).

(13) a. *Contamination*:
Movement of β from a position within α to a position outside of α values a movement-related feature γ on α with β's index.

b. *Decontamination*:
Movement of β to a criterial position deletes β's index on all movement-related features of items that c-command it.

Removal of a defective value under c-command in (13-b) can be viewed as an instance of Agree, with the feature bearing the defective value on α acting as a probe. Crucially, this only becomes possible when β has reached a criterial

---

[8] This implies that I do not follow Chomsky (1995) in assuming that indices on syntactic categories do not exist. However, indices are standardly taken to be needed anyway for semantic interpretation (see Heim & Kratzer (1998)); and it is also worth pointing out that indices do not violate the Inclusiveness Condition or the No Tampering Condition, given that they are present before the syntactic derivation starts. That said, other options (that register prior extraction of β from α without using indices) would be readily available, as long as the value on α's movement-related feature that α gets from β uniquely identifies β. (Note incidentally that this role cannot be played by *category labels*; otherwise any category with the same label as β could decontaminate α, not just β.)

position; before that, β's index does not qualify as a proper goal.

Now that it has been clarified how α registers extraction of β in α-over-β configurations by acquiring β's index as a value of its movement-related feature, the second question can be addressed: How can such a feature value lead to illformedness? One possibility might be to simply assume that this is due to a specific constraint stating that such values are not tolerated in criterial positions; see the Index Filter in (14).

(14) *Index Filter*:
A movement-related feature (like [wh], [top], [Σ]) must not have an index as (part of) its value in a criterial position.

However, closer inspection reveals that such an extra constraint is actually not necessary: The Index Filter is straightforwardly derivable under the assumptions about syntactic buffers made in chapter 2. Recall that I have proposed that the value of movement-related features is not a simple symbol like "+" or "–", but rather a list – more precisely, a queue of category symbols (a first-in/first-out list) that is successively generated by movement steps, recording all category labels passed by the moved item on its way to the ultimate landing site and deleting symbols at the bottom of the list under identity with new incoming symbols. This way, proper and improper movement can be locally determined (in improper movement configurations, the symbol list that acts as the value of a movement-related feature on some moved item does not conform to the functional sequence (f-seq) of heads in a clause, and thus violates a local version of the Williams Cycle (cf. (24) of chapter 2, repeated here as (15)), which states that the information on a list of a movement-related feature must conform to f-seq when a moved item reaches a criterial position.

(15) *Williams Cycle*:
Information on a list of a movement-related feature β must conform to f-seq when β is checked by an inherent structure-building feature [$\bullet\beta_{\pi}\bullet$] of a phase head π (i.e., in criterial positions).

Thus, a typical feature value of an object wh-phrase that has undergone long-distance movement from an embedded clause to SpecC will look like [wh: $\boxed{\text{CTvV}}$], which conforms to f-seq. In contrast, a DP that has undergone illicit long-distance scrambling from a finite CP will look like [Σ: $\boxed{\text{vVCT}}$], which does not conform to f-seq.

Against the background of this analysis, it is clear why the index of β as (part of) a value of a movement-related feature on α will lead to illformedness in criterial positions: Such an index invariably brings about an f-seq violation because it is not a category label and *can therefore never show up in any well-formed f-seq* (which consists only of category labels by definition); thus a movement-related feature with a value like [wh: $\boxed{\text{CTvV1}}$] can never conform to f-seq. In sum, then, the Index Filter in (14) can be derived as a theorem of a local–derivational approach to improper movement based on the Williams Cycle: The Williams Cycle not only rules out improper movement, it also blocks a remnant XP in a criterial position that has an index of some other item as part of the value of its movement-related feature.

It remains to be shown, then, how the Williams Cycle is violated in freezing and Müller-Takano generalization environments (because the incriminating index on the remnant XP α's movement-related feature has not yet been removed when α reaches a criterial position), and respected in anti-freezing contexts (because the incriminating index has successfully been removed when α enters a criterial position). Thus, everything now boils down to the *timing* of syntactic operations. As I will argue in the next subsection, independently motivated constraints on the timing of syntactic operations correctly predict *feeding* in the good (anti-freezing) contexts, and *counter-feeding* in the bad ones (freezing, Müller-Takano).

2.2. On the Order of Movement Operations

Minimalist syntax envisages certain basic operations, most notably Merge ('external Merge'), Move ('internal Merge') and Agree. These operations may interact with one another in syntactic derivations, thereby potentially giving rise to feeding, bleeding, counter-feeding and counter-bleeding relations (see section 2.3. of chapter 1). The question then arises of whether there are restrictions on the order of elementary operations. This issue is highly relevant in derivational approaches to syntax since resolving the order in one way or the other may make radically different empirical predictions. A well-known example involves English expletive constructions; consider the sentence pair in (16).

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

As observed by Chomsky (2000), if Merge (of the expletive *there*$_1$) precedes Move (of *someone*$_2$) on the TP cycle, (16-a) is predicted to be the correct output; in contrast, if Move precedes Merge on the TP cycle, we expect (16-b) to be well formed. The distribution of grammaticality in (16) can therefore be taken to support a general principle *Merge over Move*. Closer scrutiny reveals a surprising number of general principles determining the order of grammatical operations that have been suggested over the last decades.[9]

In what follows, I will make the following assumptions about the order of operations in the case of multiple movement to phase edges.

(17) *Movement to phase edge*:
   a. If α c-commands β in the pre-movement structure, then α moves first, and β moves after that, to a *lower* specifier.
   b. If α does not c-command β in the pre-movement structure, the order is not fixed; the second item that moves ends up in a *higher* specifier.

(17-a) states that in cases of multiple movement of the same type, of items that are in a c-command relation, the derivation proceeds by tucking in; see Richards (2001) and Branigan (2013), among many others (and Fox & Pesetsky (2005), Stroik (2009), Unger (2010) and Assmann & Heck (2013) for related concepts). (17-b) will permit movement of β to apply first in α-over-β environments (where α, β both initially undergo intermediate movement), which is a precondition for CED satisfaction of any derivation in which this configuration occurs.[10] The assumptions in (17) can arguably be shown to

---

[9] Here are some proposals: (i) *extrinsic ordering* (Chomsky (1965), Perlmutter & Soames (1979)), possibly with free variation to capture cross-linguistic variation (Georgi (2014)); (ii) *obligatoriness vs. optionality* (Pullum (1979)); (iii) *specificity* (Sanders (1974), Pullum (1979), van Koppen (2005), Lahne (2012), Georgi (2013)); (iv) *anti-specificity* (Chomsky (2000; 2001) – note that ordering Merge before Move has originally been justified by assuming that Merge is the more general, i.e., less specific, operation); (v) *strict cyclicity* (McCawley (1984; 1998)); (vi) *strata/levels* (Chomsky (1981): D-structure, S-structure, LF/PF; Riemsdijk & Williams (1981): NP-structure); (vii) *rule vocabulary* (Arregi & Nevins (2012, ch. 6)); (viii) *minimal search* (Chomsky (2013)).

[10] Thus, the MLC must not force movement of α in $[_\alpha \ldots\beta\ldots]$ configurations, at least not if both items undergo intermediate movement steps driven by edge features. More generally, the MLC will play no role anymore in the account of the Müller-Takano generalization to be developed

follow from more elementary concepts, but for present purposes I will simply take them as given.[11] On this basis, let me now turn to the three generalizations about movement in α-over-β configurations. It will turn out that the interaction of the assumptions about the order of criterial and intermediate movement in (17) and the assumptions about contamination and decontamination of movement-related features in (13) correctly predicts the distribution of anti-freezing, freezing, and Müller-Takano effects in α-over-β configurations.

2.3. Feeding vs. Counter-Feeding: Generalizations Derived

2.3.1. Initial Steps

To begin with, recall that I assume that all phrases qualify as phases. Thus, the PIC forces intermediate movement through every intervening phrase edge domain on the way to the criterial landing site of a moved item. Furthermore, the CED and the SCC continue to hold (in contrast to the MLC).

In all cases, the decisive stage of the derivation starts when the higher XP_2 (α) has merged with a head Y, as a complement, where XP_1 (β) has earlier undergone movement to XP_2's specifier (for PIC reasons); cf. (18-a). In the first step, XP_1 must now move out of XP_2, to SpecY (if XP_2 moves first, the CED will be violated); cf. (18-b). Such movement is typically unproblematic because a fixed order of operations is not required here; cf. (17-b). However, movement of XP_1 contaminates XP_2 by valuing X_2's movement-related feature (γ) with XP_1's index, thereby creating a situation that must be remedied before XP_2 reaches a criterial position.[12] In the next step, XP_2 moves to an

---

below. See Müller (2011) for arguments against the MLC, and for a proposal of how to derive most of the intervention effects that it is supposed to cover.

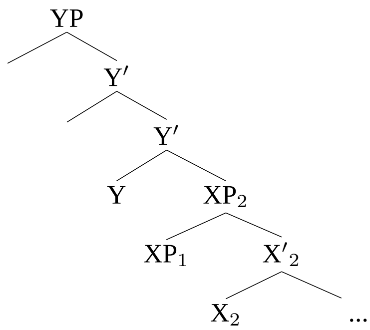
[11] See Müller (2013) for discussion. Basically, (17) can be understood in such a way that it brings about a *minimization of changes to existing structures*, as required under a (non-categorical) version of the No Tampering Condition (NTC, Chomsky (2007; 2008; 2013)) that incorporates Pullum's (1992) assumptions about the origins of cyclicity: Once established, linear order is preserved throughout the derivation as much as possible (i.e., as long as triggers for movement can be satisfied, etc.), even at the cost of minimally counter-cyclic tucking in operations (cf. (17-a)); but such violations are not permitted when maintaining linear order is not an issue (cf. (17-b)).

[12] Throughout, I gloss over the X/XP distinction here, and presuppose that an appropriate theory of pied piping ensures that XP is moved when X bears the movement-related feature; see Heck (2008) for comprehensive discussion.

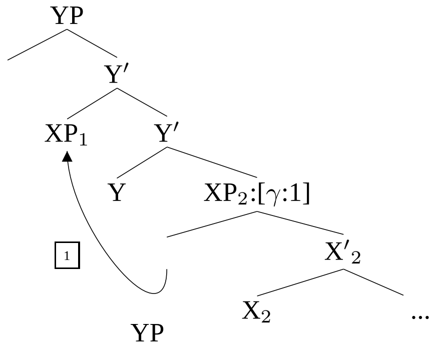
outer SpecY (because of the PIC and (17-b)); cf. (18-c), with the movement-related feature γ of XP_2 now bearing a contaminated value: [γ:Y1].[13]

(18) *Initial steps in α-over-β configurations*

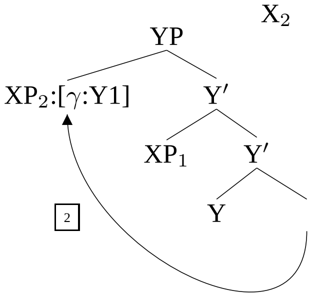
a.



b.



c.

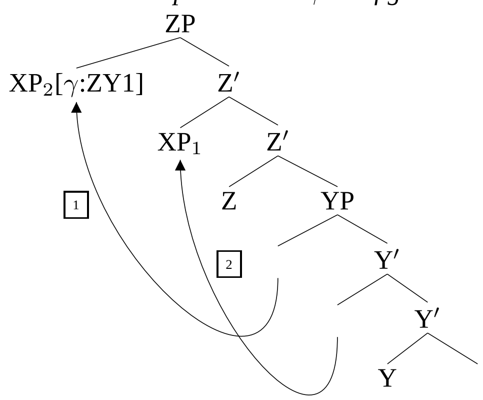


[13] Here and in the remainder of this chapter, I abstract away from category symbols on buffers of movement-related features of items from which no extraction has taken place, such as XP_1 in (18). This is purely for ease of exposition, and unproblematic since standard instances of improper movement (as they figure prominently in chapter 2) are not at issue in the present context.

2.3.2. *Intermediate Steps*

Consider next what happens on intermediate cycles, where both XP_2 (α) and XP_1 (β) undergo intermediate (edge feature-driven) movement to specifier domains, as required by the PIC. Given (17-a), the two intermediate movement steps must be order-preserving, with the first, higher item (the remnant category XP_2) moving first, and the second, lower item (XP_1) moving afterwards to a lower specifier, by tucking in; see (19). Of course, this pattern can be applied recursively, leading to order-preservation with multiple movement spanning arbitrarily long distances.

(19) *Intermediate steps in α-over-β configurations*

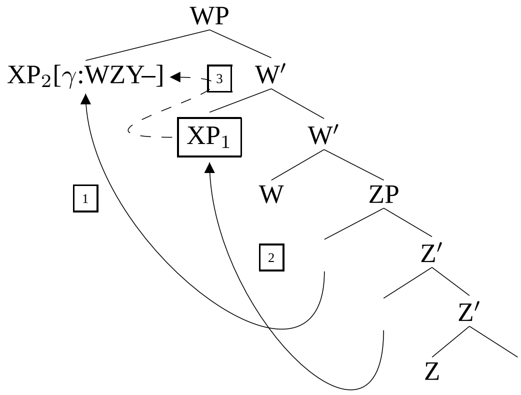


Finally, building on either (18-c) or (19), criterial movement steps of either XP_1 (β), XP_2 (α), or both XP_2 and XP_1 can take place, giving rise to anti-freezing, freezing, and Müller-Takano effects.

2.3.3. *Anti-Freezing*

In the case of anti-freezing as in typical remnant movement constructions like those in (6) (with a relevant example repeated in (21)), again XP_2 undergoes intermediate movement first. Subsequently, criterial movement of XP_1 takes place to an inner specifier, which then removes XP_1's index from XP_2. XP_2 is free to undergo criterial movement in accordance with the Williams Cycle from now on. This *feeding* effect in anti-freezing configurations (decontamination feeds criterial remnant movement) is shown in (20) ($\boxed{1}$ signals intermediate movement of XP_2, $\boxed{2}$ subsequent criterial movement of XP_1; and $\boxed{3}$ index removal (while XP_1 is c-commanded by XP_2); as in chapter 2, a box around a category indicates that the category has reached a criterial position).

(20) *Criterial steps in α-over-β configurations: XP$_1$*



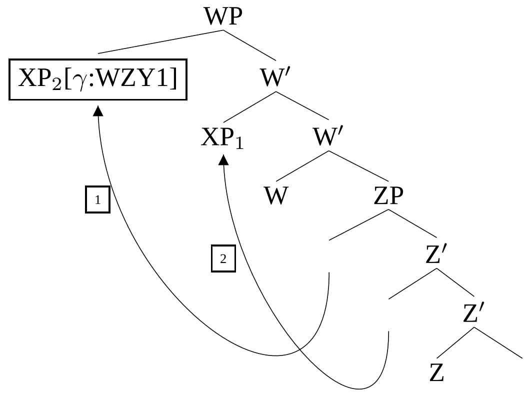
($\sqrt{}$f-seq)

(21) [$_{\text{VP}_2}$ t$_1$ Gelesen] hat das Buch$_1$ keiner t$_2$
read has the book no-one

2.3.4. Freezing

In contrast, freezing configurations as in (3) (also see (23)) involve *counter-feeding*: Criterial movement of XP$_1$ comes *far* too late to remove the fatal index from XP$_2$: XP$_2$ undergoes criterial movement when it still has a contaminated value on its movement-related feature and thereby violates the Williams Cycle; see (22). Here, XP$_2$ has a defective value 1 in a criterial position, and XP$_1$ is far from being able to remove the incriminating symbol 1 because it has undergone intermediate rather than criterial movement itself. In addition, note that movement of XP$_1$ would have come too late anyway (this will become relevant below).

(22) *Criterial steps in α-over-β configurations: XP_2*



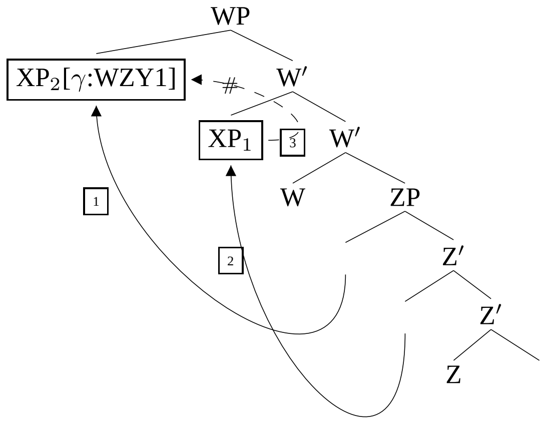
(*f-seq)

(23) *Was$_1$ denkst du [$_{VP_2}$ t$_1$ gelesen] hat keiner t$_2$?
what think you read has no-one

2.3.5. Müller-Takano Generalization

The third possible continuation of a derivation involving criterial movement is that both XP_2 and XP_1 undergo criterial movement to a given specifier domain. This is the situation underlying Müller-Takano effects as in (7) (a relevant example is repeated here as (25)), and the derivation also involves *counter-feeding* under present assumptions: This time, criterial movement of XP_1 comes *a bit* too late to be able to remove the fatal index from XP_2: Given (17-a), XP_2 undergoes criterial movement first and thereby violates the Williams Cycle; cf. (24). Subsequent criterial movement of XP_1 (via tucking in) creates a configuration in which the defective index on XP_2 could be removed (signalled by $\boxed{3}$), but at this point of the derivation, the damage has already, and irrevocably, been done.

(24) *Criterial steps in α-over-β configurations: XP_1 & XP_2*



(*f-seq)

(25) *dass [$_{VP_2}$ t_1 zu lesen] [$_{DP_1}$ das Buch] keiner t_2 versucht hat
that to read the book$_{acc}$ no-one tried has

2.3.6. Criterial Initial Steps

The discussion so far has presupposed that the initial operations in (18-a) are two intermediate movement steps, and this is arguably the prototypical situation. Still, it is worth investigating whether the conclusions concerning anti-freezing, freezing, and Müller-Takano effects can be maintained if XP_1, XP_2, or both XP_1 and XP_2, undergo extremely local criterial movement to SpecY positions in (18).[14]

[14] Such a situation cannot occur if extremely local criterial movement is blocked; see Bošković (1997), Abels (2003; 2012a), and Grohmann (2003)) for proposals to this effect. Throughout the present chapter, the analyses presuppose that extremely local intermediate movement is available (given that every phrase is a phase); but this does not necessarily imply that extremely local criterial movement is also an option. As a matter of fact, given that the structure-building features triggering the operation are provided in the course of the derivation in the former case, whereas they are intrinsically present on heads in the latter case, one might plausibly argue that whatever constraint blocks extremely local criterial movement (e.g., economy, as in Abels (2003)) does not also have to block extremely local intermediate movement. Still, these considerations notwithstanding, in view of the highly local nature of some of the criterial scrambling operations involved in relevant examples, I will assume in what follows that extremely local movement is an option, in both criterial and intermediate contexts; i.e., there is no constraint whatsoever demanding anti-locality.

There are three cases to be addressed. First, XP_2 undergoes criterial movement to SpecY in (18-a), and XP_1 undergoes intermediate movement. If XP_2 moves first, and XP_1 moves after that, this violates the CED (it does not produce a counter-feeding effect of the same type as seen in (22) because the value of XP_2's movement-related feature is not yet contaminated by extraction of XP_1). Alternatively, XP_1 is extracted first, thereby contaminating XP_2's movement-related feature with its index. Subsequent criterial movement of XP_2 (to an outer specifier of Y) will then violate the Williams Cycle.

The second possibility is that XP_1 undergoes criterial movement to SpecY in (18-a), and XP_2 undergoes intermediate movement. XP_1 extraction can (given (17-b)) and must (given the CED) apply first, which instantiates XP_1's index on the value of XP_2's movement-related feature. Next, XP_2 undergoes intermediate movement to a higher specifier position of Y, and the incriminating symbol is deleted again on XP_2, under c-command. So, this derivation is legitimate, as suggested by the empirical evidence. For instance, a sentence like (21) is also possible in German if the scrambled XP_1 follows rather than precedes the subject DP, which may in turn be assumed to be in situ, in Specv; cf. (26-a). An argument for the low (in situ) position of *keiner* ('no-one$_{nom}$') is that there is an option of a preceding unstressed pronoun in minimally different double object constructions; also see section 3 below); cf. (26-b).

(26) a. [$_{VP_2}$ t_1 Gelesen] hat keiner das Buch$_1$ t_2
read has no-one the book

b. [$_{VP_2}$ t_3 t_1 Geschenkt] hat ihm$_3$ keiner das Buch$_1$ t_2
given has him$_{dat}$ no-one$_{nom}$ the book$_{acc}$

The third and final possibility to be discussed here involves extremely local criterial movement of both XP_1 and XP_2 to SpecY in (18-a). Again, XP_1 needs to extract first (because of the CED), which contaminates XP_2. Since decontamination can only take place when XP_2 c-commands XP_1, XP_2 must move next, thereby immediately giving rise to a violation of the Williams Cycle. Thus, the order of operations is (i) criterial movement of XP_1, (ii) valuation of XP_2 with XP_1's index, (iii) criterial movement of XP_2 (giving rise to a violation of the Williams Cycle), and (iv) deletion of XP_1's index on XP_2 (which comes too late).

Consequently, for the three cases involving one (or two) criterial initial movement steps, the analysis makes correct predictions without further ado:

There is a freezing effect in the first case (only XP_2 undergoes local criterial movement) that is derived in terms of counter-feeding of decontamination by criterial movement of XP_1 (which comes much too late); there is an anti-freezing effect in the second case (only XP_1 undergoes local criterial movement) that is derived via feeding of decontamination of XP_2 by criterial movement of XP_1; and there is a Müller-Takano effect in the third case, derivable as counter-feeding of decontamination by criterial movement of XP_1 (which comes a bit too late): The timing of operations is crucial.

To sum up, assuming that remnant-creating movement of β in α-over-β configurations does not come entirely for free but involves the creation of a temporarily contaminated value on the movement-related feature of the remnant XP α which must eventually be removed by criterial movement of β, the freezing, anti-freezing, and Müller-Takano generalizations can all be shown to follow from independently established restrictions on multiple movement: On this view, examples subsumed under the freezing and Müller-Takano generalizations emerge as counter-feeding effects, with criterial movement of β applying too late in the derivation.

3. Consequences

The analysis makes a number of further predictions. I will discuss three of them, concerning (i) multiple remnant movement, (ii) the scope of Müller-Takano effects, and (iii) temporary defectivity.

3.1. Multiple Extraction, Multiple Remnant Movement

The analysis is compatible with multiple extraction from a remnant XP, as in (27). Here, there is both pronominal object shift of *es*$_1$ ('it') and scrambling of *dem Fritz* ('the Fritz'), and both items are extracted from VP_2 while the latter is still in situ, thereby contaminating it with their separate indices (recall that the value of a movement-related feature is a list). Both indices are subsequently removed from VP_2 when DP_1 and DP_2 reach their respective criterial landing sites, and when VP_2 reaches the topic position, it has the value of its movement-related feature restored to normalcy.

(27) [$_{VP_3}$ t$_2$ t$_1$ Gegeben] hat es$_1$ die Maria am Ende [$_{DP_2}$ dem Fritz] t$_3$
given has it$_{acc}$ the Maria$_{nom}$ in the end the Fritz$_{dat}$

The analysis is also compatible with multiple remnant movement, as in examples like (28-a) vs. (28-b) in German (see den Besten & Webelhuth (1990)).

(28) a. [$_{VP_3}$ t_2 Gerechnet] hat da_1 wie immer keiner t_3 [$_{PP_2}$ t_1 mit]
counted has there as always no-one with

b. *[$_{VP_3}$ t_2 Gerechnet] hat da_1 wie immer [$_{PP_2}$ t_1 mit] keiner t_3
counted has there as always with no-one

Under the analysis in Müller (1998), (28-a) involves a combination of scrambling of *da*$_1$ from PP_2, (globally string-vacuous) extraposition of the remnant PP_2 from VP_3, and topicalization of the remnant VP_3; suppose that this is indeed correct (also cf. the appendix to this chapter below). Given the present assumptions, this derivation is unproblematic (an instance of anti-freezing): The R-pronoun *da*$_1$ is moved from PP_2 while the latter is in situ, thereby creating a defective value on PP_2's extraposition feature that can be removed after *da*$_1$ has reached its criterial scrambling position, before PP_2 has reached the (higher) extraposition position. Furthermore, *da*$_1$ and PP_2 both undergo movement from VP_3 while the latter is still in situ, thereby contaminating it with their indices; but decontamination can take place for both defective values of VP_3's [top] feature before VP_3 actually reaches its criterial position (which is higher than that of PP_2). In contrast, in (28-b), PP_2 undergoes scrambling, and reaches this position before *da*$_1$ shows up in *its* criterial position and has a chance to remove the illicit feature value on PP_2. Thus, there is a freezing effect.

The underlying logic is shown schematically in the abstract derivation in (29), for three phrases (XP_1, XP_2, XP_3) that all need to ungergo movement. First (see (29-a)), XP_1 moves to SpecX_2, X_3 is merged with XP_2, and XP_1 moves on to SpecX_3, thereby contaminating XP_2 with its index (this movement must apply before movement of XP_2 because of the CED).

Second, (see (29-b)), XP_2 also moves to SpecX_3, a new head Y is merged (possibly also a new specifier, here indicated by WP), and both XP_2 and XP_1 move (order-preservingly) to SpecY, thereby contaminating XP_3 with their indices. Movement of XP_2 and XP_1 must (given the CED) and can (given (17-b)) take place before movement of XP_3. Movement of XP_1 can be intermediate or criterial at this point. If it is criterial, removal of XP_1's index on XP_2 (and, subsequently, XP_3) will take place within YP already. Although this would seem to be the case in (28-a) (where YP = vP, XP_3 = VP_3, XP_2 = PP_2, and

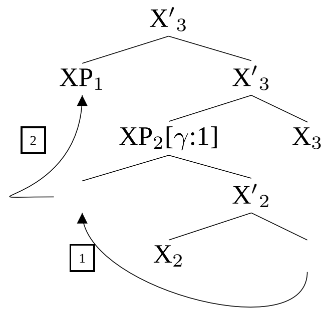
XP_1 = DP_1, and scrambling of DP_1 ends in Specv), let us consider the more complex derivation where XP_1 has not yet reached a criterial position on the YP cycle, for purely expository purposes.

Thus, third (see (29-c)), XP_3 moves to SpecY, a new head Z is merged (with possibly also a new specifier, here indicated by UP), and XP_3, XP_2, and XP_1 all undergo movement to SpecZ, in that order, and maintaining pre-movement c-command relations (given (17-a)). Suppose that XP_1 undergoes criterial movement to SpecZ, whereas XP_2 and XP_3 undergo intermediate movement. As a consequence, XP_1 removes its index from XP_2 and XP_3, which c-command it in ZP.

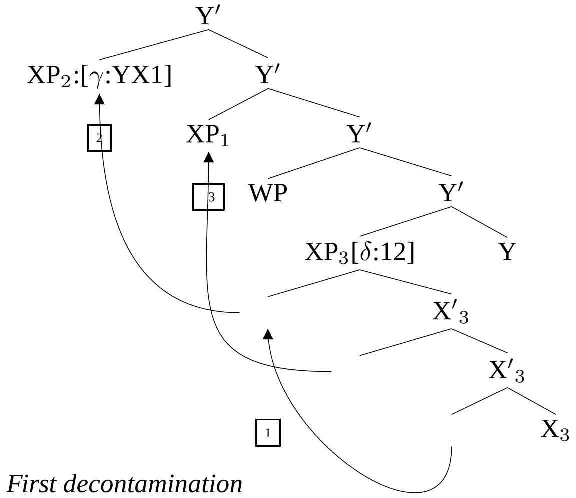
Fourth and finally (see (29-d)), a new head is merged (R). XP_2 undergoes criterial movement on the next cycle (here: to a right-peripheral position, as in (28-a), but this issue is orthogonal), whereas XP_3 still undergoes intermediate movement (before that). At this point, the derivation is virtually indistinguishable from the standard example of anti-freezing with remnant movement in (20): XP_2's criterial step decontaminates XP_3, and XP_3 is free to move on in what follows; there is no danger anymore of violating the Williams Cycle (unless, of course, a standard context of improper movement is created).

(29) *Multiple remnant movement: anti-freezing*

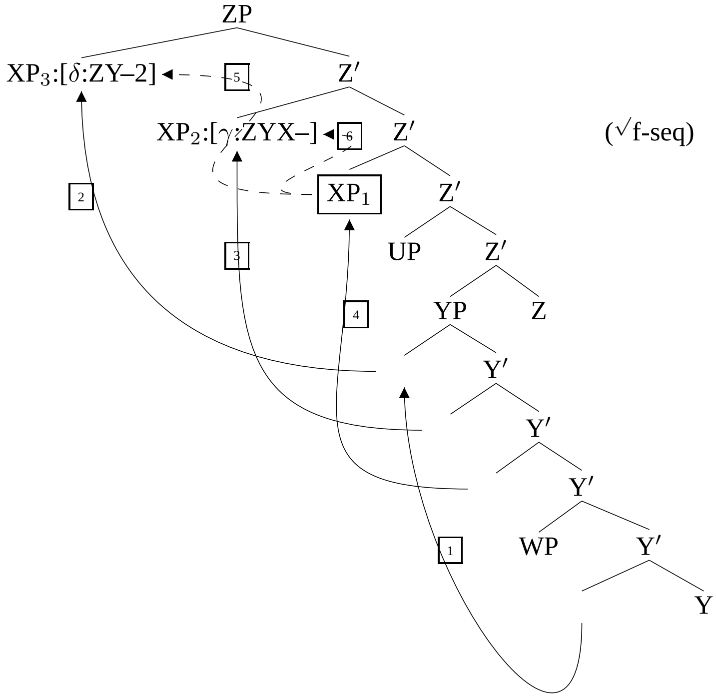
a. *Generation of the first remnant category*



b. *Generation of the second remnant category*

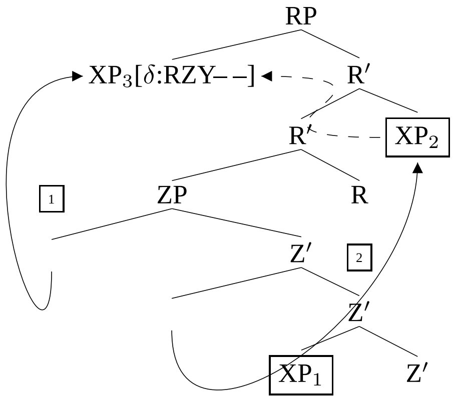


c. *First decontamination*



(√f-seq)

d. *Second decontamination*



($\surd$f-seq)

Clearly, if XP_1 fails to undergo criterial movement before XP_2 and XP_3 (as in (28-b)), or if XP_2 fails to undergo criterial movement before XP_3, the derivation will crash.

3.2. *The Scope of Müller-Takano Effects*

The present approach makes a prediction that is different from the accounts of Müller-Takano generalization effects highlighted in section 1.2.3 above. The prediction is that remnant movement should be possible, in violation of the Müller-Takano generalization as it is formulated above, if the remnant XP_2 has the same movement-related feature as XP_1, but checks this with some higher head in the clause. The reason is that ungrammaticality can only occur if the feature is checked in the same domain; if XP_2 does not check the feature in the same phase edge as XP_1, it can be decontaminated before it reaches its criterial position.[15] At present, I take it to be an open question whether this loophole for identical movement operations in α-over-β configurations is desirable or not. A straightforward way to close it would be to assume that there is a one-to-one correspondence between movement-attracting features and the functional heads on which they occur. Still, an approach recognizing this loophole might possibly receive empirical confirmation. For instance, Fanselow (2002) and Hale & Legendre (2004) argue that sentences like (25) (repeated here as (30))

---

[15] Note that this consequence is in fact one that also arises in Takano's (1994) original approach.

improve with a rise/fall (I-topicalization) intonational pattern that might indicate that a higher functional head is targetted by the VP fronting operation than by prior DP scrambling (see Frey (2000; 2004)).

(30)	*dass	[$_{VP_2}$ t$_1$	zu lesen]	[$_{DP_1}$ das Buch]	keiner	t$_2$	versucht	hat
	that		to read	the book$_{acc}$	no-one		tried	has

If different movement-triggering features are involved in the two positions targetted by DP$_1$ and VP$_2$, both analyses of Müller-Takano effects discussed in the present chapter would be able to accomodate a possible amelioration of (30) under I-topicalization. If, however, the same movement-inducing feature is involved (viz., a feature that uniformly triggers all kinds of scrambling in German, like [Σ]), then an improvement under I-topicalization would follow under the local–derivational approach in terms of contamination and decontamination of syntactic buffers, but not under the non-local–derivational approach based on the MLC.[16]

3.3. Temporary Defectivity

As a consequence of the buffer-based analysis of α-over-β constructions, a remnant XP$_2$ (α) is not uniformly the same kind of syntactic object throughout a derivation. At the very beginning, before extraction from it takes place, α is a legitimate object; and it also qualifies as a legitimate object as soon as the item that has undergone movement out of XP$_2$ (viz., β) has reached a criterial position, and index removal is effected. However, α is defective in the intermediate parts of the derivation, after extraction of β and before the criterial movement step of β. Therefore, it seems plausible to venture the hypothesis that the temporary defectivity of a remnant XP in the middle of its active life cycle should be reflected in a special behaviour with respect to other syntactic processes, and since the incriminating symbol is an index, we may expect problems to arise in index-sensitive domains like binding and scope.

Consider binding first. For concreteness, suppose that a temporarily defective remnant category α, as well as the items contained in it, cannot participate in the computation of binding principles. With this in mind, consider the sen-

---

[16] Similar conclusions hold for allegedly acceptable cases of left dislocation of remnant VPs; cf. Fanselow (2002) vs. Müller (1998).

tence in (31).

(31) ?[$_{\text{VP}_2}$ Dem Peter t$_3$ zu geben] weiß ich nicht [$_{\text{CP}_4}$ was$_3$
the Peter$_{dat}$ to give know I not what$_{acc}$
Maria denkt [$_{\text{CP}_5}$ dass Karl t$_2$ versuchen sollte]]
Maria$_{nom}$ thinks that Karl try should

In (31), there is long-distance wh-movement of DP$_3$ to the least deeply embedded SpecC position, plus long-distance topicalization of the remnant VP$_2$. The resulting sentence is only mildly degraded, like other cases of (non-adjunct) topicalization across a wh-island in German. The present analysis predicts that, as a result of contamination by DP$_3$ extraction, VP$_2$ has a defective value on its movement-related feature ([top]) in the subpart of the derivation that starts with the most deeply embedded vP and ends with CP$_4$. Given the above assumption about the consequences of temporary defectivity for binding theory, the prediction is that the R-expression *dem Peter*, as part of the remnant VP$_2$, cannot participate in the computation of binding principles at these stages of the derivation. More specifically, it is predicted to be immune to Principle C effects at these stages (but not before or after these stages). Interestingly, this prediction seem to be borne out. This is shown by the three examples in (32).

(32) a. *[$_{\text{VP}_2}$ Dem Peter$_1$ t$_3$ zu geben] weiß ich nicht [$_{\text{CP}_4}$ was$_3$
the Peter$_{dat}$ to give know I not what$_{acc}$
Maria denkt [$_{\text{CP}_5}$ dass er$_1$ (selbst) t$_2$ versuchen sollte]]
Maria$_{nom}$ thinks that he (self) try should

b. ?[$_{\text{VP}_2}$ Dem Peter$_1$ t$_3$ zu geben] weiß ich nicht [$_{\text{CP}_4}$ was$_3$
the Peter$_{dat}$ to give know I not what$_{acc}$
er$_1$ (selbst) denkt [$_{\text{CP}_5}$ dass Maria t$_2$ versuchen sollte]]
he (self) thinks that Maria$_{nom}$ try should

c. *[$_{\text{VP}_2}$ Dem Peter$_1$ t$_3$ zu geben] weiß er$_1$ (selbst) nicht [$_{\text{CP}_4}$
the Peter$_{dat}$ to give knows he (self) not
was$_3$ Maria denkt [$_{\text{CP}_5}$ dass ich t$_2$ versuchen sollte]]
what$_{acc}$ Maria$_{nom}$ thinks that I try should

In (32-a), there is a clear Principle C effect, in addition to the mild wh-island effect. Here, the incriminating co-indexed subject pronoun *er* ('he') shows up in the most deeply embedded CP$_5$; it is merged in Specv *before* DP$_3$ and VP$_2$ move to (outer) specifiers of vP (given Merge over Move). This means that *dem*

Peter$_1$ violates Principle C before DP$_1$ is extracted from VP$_2$, contaminating it with its index and rendering it defective.[17]

In (32-c), there is also a clear Principle C effect. Here, the incriminating co-indexed pronoun *er* ('he') shows up in the matrix vP; it successfully binds *dem Peter* in VP$_2$ *after* DP$_3$ (*was*, 'what') has reached its criterial position in CP$_4$, when VP$_2$ is in SpecV of the root clause.

However, there is a subtle but clearly discernible improvement in (32-b). Here, the subject pronoun *er* ('he') shows up in the intermediate clause, CP$_4$; and throughout the derivational stages where the pronoun might effect c-command of the co-indexed DP *dem Peter*, the latter is sheltered by the defective remnant VP$_2$. Thus, we get a reconstruction effect in (32-ac) and an anti-reconstruction effect in (32-b) that follows straightforwardly under present assumptions but must, it seems, remain a complete mystery under virtually all other theories of movement.

Next consider scope. According to *Barss's generalization* (see Barss (1986), Lechner (1998), Sauerland & Elbourne (2002), Bhatt & Dayal (2007), Neeleman & van de Koot (2010), Heck & Assmann (2012)), a quantified item in a moved remnant XP α cannot take scope, via reconstruction, over an item β that has undergone movement from α. Thus, (33-a) is ambiguous, but (33-b), with remnant VP topicalization, is not: The dative DP *jedem Studenten* ('every student') cannot take scope over the accusative DP *ein Buch* ('a book') (also see Thoms & Walkden (2013)).

(33) a. Jedes Buch hat sie einem Studenten gegeben
every book$_{acc}$ has she$_{nom}$ a student$_{dat}$ given
(∀ > ∃, ∃ > ∀)

b. Jedem Studenten gegeben hat sie ein Buch
every student$_{dat}$ given has she$_{nom}$ a book$_{acc}$
(*∀ > ∃, ∃ > ∀)

Suppose now that a temporarily defective XP$_2$ in intermediate stages of a derivation (i.e., after contamination, and before decontamination) is not vis-

[17] The example is pragmatically complex but by no means implausible: I don't know for which thing (e.g. a price, an assignment, a job) it is the case that Maria believes that Peter should try to give it to himself.

ible for the purposes of Barss's generalization. Then we expect that it can in fact take scope over β in this syntactic domain.[18] Needless to say, relevant examples are not completely trivial to construct, and quite hard to classify. Here is an attempt. First, (34) shows that a remnant VP with a quantified dative object can undergo long-distance topicalization in German (again, with a weak wh-island effect arising).

(34) ?[$_{\mathrm{VP}_2}$ Jedem Studenten t$_1$ gegeben] weiß ich nicht [$_{\mathrm{CP}}$ was$_1$ sie t$_2$
every student$_{dat}$ given know I not what$_{acc}$ she
hat]
has

Second, as observed by Beck (1997), sentences like (35) are ambiguous.

(35) Wieviele Bücher$_1$ hat sie jedem Studenten$_2$ t$_1$ gegeben ?
how many books$_{acc}$ has she every student$_{dat}$ given
($\exists > \forall$, $\forall > \exists$)

(35) can mean: For which number n: There are n books that she gave to every student ($\exists > \forall$). Alternatively, it can mean: For which number n: For every student there are n books that she have to him ($\forall > \exists$).

Third, by combining the two contexts, we expect a Barss's generalization effect. As shown in (36), this does indeed seem to be the case; (36) can hardly have a reading where *jedem Studenten* ('every student') takes scope over the existential quantifier.[19]

(36) ??[$_{\mathrm{VP}_2}$ Jedem Studenten$_3$ t$_1$ gegeben] weiß ich nicht [$_{\mathrm{CP}}$ wieviele
every student$_{dat}$ given know I not how many
Bücher$_1$ sie t$_2$ hat]
books$_{acc}$ she has
(*$\forall > \exists$)

[18] In the approach of Heck & Assmann (2012), this could be so because a temporarily defective XP$_2$ might exceptionally make acyclic reconstruction possible, where reconstruction normally has to be cyclic.

[19] For reasons unclear to me, the wh-island effect seems to be stronger with *wieviele Bücher* ('how many books') than with *was* ('what'), which is an orthogonal factor that nevertheless seems to complicate judgements somewhat.

Against this background, consider the example in (37). Given the present assumptions, the prediction is that a Barss's generalization effect could be absent here, i.e., that the sentence can have a reading with the universal quantifier outscoping the existential quantifier since the remnant VP_2 is defective at the relevant stages of the derivation (i.e., in the intermediate CP).

(37) ??[$_{\text{VP}_2}$ Jedem Studenten$_3$ t$_1$ gegeben] weiß ich nicht [$_{\text{CP}}$ wieviele
every student$_{dat}$ given know I not how many
Bücher$_1$ Fritz denkt [$_{\text{CP}}$ dass sie t$_2$ hat]
books$_{acc}$ Fritz thinks that she has

A similar, parallel example pair is given in (38). Again, the prediction would be that (38-b) can have a wide-scope reading for the universal quantifier more easily than (38-a).

(38) a. ??[$_{\text{VP}_2}$ Jedem Besucher$_3$ t$_1$ zu zeigen] weiß ich nicht [$_{\text{CP}}$
every visitor$_{dat}$ to show know I not
wieviele Bilder$_1$ sie t$_2$ versuchen will]
how many pictures$_{acc}$ she try wants

b. ??[$_{\text{VP}_2}$ Jedem Besucher$_3$ t$_1$ zu zeigen] weiß ich nicht [$_{\text{CP}}$
every visitor$_{dat}$ to show know I not
wieviele Bilder$_1$ Maria sagt [$_{\text{CP}}$ dass sie t$_2$ versuchen
how many pictures$_{acc}$ Maria says that she try
will]
wants

However, since judgements are extremely subtle in this case, and the exact theoretical implementation of Barss's generalization is far from clear (and certainly not innocuous), I will leave it at that, and draw a conclusion.

4. Conclusion

The gist of the analysis I have developed in this chapter is that movement of some category β from another category α that needs to undergo movement itself triggers a contamination of the movement-related feature of α; thus, there is a price to be paid for remnant movement in syntax. A contamination of α is temporarily acceptable, but decontamination must take place before α moves to a criterial position; and the required index removal can only apply when β has moved to a criterial position (and is c-commanded by α). On this basis, independently motivated assumptions about the order of movement operations ensure that freezing and Müller-Takano generalization configurations violate the Williams Cycle, a local constraint against improper movement demanding adherence to f-seq; and that anti-freezing configurations do not. Interestingly, there is opacity (and hence a potential argument for a derivational approach to syntax) in both the standard approach to movement in α-over-β configurations sketched in section 1, and the new local approach developed in section 2. However, whereas it is the anti-freezing effect that creates opacity as an instance of counter-bleeding in the former case, it is the freezing and Müller-Takano effects that create opacity as an instance of counter-feeding in the latter case.[20]

[20] To be sure, there is still counter-bleeding with remnant movement in the new approach since the CED would be violated with extraction of β after α moves; but this does not distinguish legitimate remnant movement from illegitimate kinds of movement in α-over-β constructions anymore.

Appendix

As mentioned in footnote 1, the basic assumption that constructions such as (1) in German (repeated here as (39)) involve remnant movement, i.e., a combination of DP scrambling and subsequent VP topicalization, has sometimes been called into question; cf. De Kuthy & Meurers (2001), Fanselow (2002), Hale & Legendre (2004) and Thoms & Walkden (2013), who all argue that remnant movement does not exist – either not at all, or at least not in contexts like the one exemplified by (39).

(39)	[$_{\text{VP}_2}$ t_1	Gelesen]	hat	das	Buch$_1$	keiner	t_2
		read	has	the	book	no-one	

Furthermore, Grewendorf (2003; 2004) and Abels (2008), while basically embracing a remnant movement approach to these kinds of constructions, assume slightly different generalizations from the ones that I have presupposed throughout (viz., freezing, anti-freezing, and Müller-Takano generalization). It is the purpose of this appendix to critically examine core arguments that have been brought forward to substantiate these claims. The appendix consists of two parts, A.1. and A.2. In A.1., I address arguments against a remnant movement analysis of (39); after that, in A.2., I turn to alternative generalizations about remnant movement.

A.1. Arguments Against Remnant Movement

As far as I can see, virtually all arguments against a remnant movement approach to (39) belong to one of three types: In constructions involving movement of what looks like an incomplete category α which lacks a category β that would at first sight seem to have been base-generated within α, there is in fact no X-movement of β out of α, feeding remnant Y-movement of α, because

- β can be an item that cannot normally undergo X-movement;
- β does not necessarily show island effects that are normally indicative of X-moved items;
- X-movement does not exist in the first place.

I address these three types of arguments against remnant movement in turn, in the following three subsections.

A.1.1. Categorial Selectivity

There are several kinds of items that are usually taken to tend to resist scrambling in German (to various degrees), such as wh-indefinites (see (40-a); Haider (1993)), negative DPs like *niemand* ('no-one'; see (40-b)), predicative APs (see (40-c); Stechow & Sternefeld (1988), Fanselow (1992b), Haider (1993)), adverbs (see (40-d); Stechow & Sternefeld (1988), Fanselow (1992b; 2001), Haider (1993), Haider & Rosengren (1998)), and expressions like *Gesindel* ('riff-raff') (see (40-e); Fanselow (1992b; 1995)).

(40) a. (i) dass der Karl der Maria was$_1$ gegeben hat
that the Karl$_{nom}$ the Maria$_{dat}$ something$_{acc}$ given has

(ii) ??dass der Karl was$_1$ der Maria t$_1$ gegeben hat
that the Karl$_{nom}$ something$_{acc}$ the Maria$_{dat}$ given has

(iii) ?*dass was$_1$ der Karl der Maria t$_1$ gegeben hat
that something$_{acc}$ the Karl$_{nom}$ the Maria$_{dat}$ given has

b. (i) dass der Fritz niemanden$_1$ küsst
that the Fritz$_{nom}$ no-one$_{acc}$ kisses

(ii) ?*dass niemanden$_1$ der Fritz t$_1$ küsst
that no-one$_{acc}$ the Fritz$_{nom}$ kisses

c. (i) dass der Karl das Fleisch roh$_1$ gegessen hat
that the Karl$_{nom}$ the meat$_{acc}$ raw eaten has

(ii) ??dass der Karl roh$_1$ das Fleisch t$_1$ gegessen hat
that the Karl$_{nom}$ raw the meat$_{acc}$ eaten has

(iii) ?*dass roh$_1$ der Karl das Fleisch t$_1$ gegessen hat
that raw the Karl$_{nom}$ the meat$_{acc}$ eaten has

d. (i) dass sie ihr Bier schnell getrunken hat
that she$_{nom}$ her beer$_{acc}$ quickly drunk has

(ii) *dass schnell sie ihr Bier t$_1$ getrunken hat
that quickly she$_{nom}$ her beer$_{acc}$ drunk has

e. (i) dass keiner hier Gesindel$_1$ begrüßen will
that no-one$_{nom}$ here riff-raff$_{acc}$ greet wants to

(ii) ?*dass Gesindel$_1$ keiner hier t$_1$ begrüßen will
that riff-raff$_{acc}$ no-one$_{nom}$ here greet wants to

However, as noted by Fanselow (2002, 99-104) and De Kuthy & Meurers (2001, 156), the items bearing index 1 in (40) can all show up as βs in remnant movement constructions; see the examples in (41).

(41) a. $[_{\mathrm{VP}_2}$ t_1 Gegeben] hat er ihr was$_1$ t_2
given has he$_{nom}$ her$_{dat}$ something$_{acc}$

b. $[_{\mathrm{VP}_2}$ t_1 Geküsst] hat der Fritz niemanden$_1$ t_2
kissed has the Fritz$_{nom}$ no-one$_{acc}$

c. $[_{\mathrm{VP}_2}$ t_3 t_1 Gegessen] hat der Karl das Fleisch$_3$ roh$_1$ t_2
eaten has the Karl$_{nom}$ the meat raw

d. $[_{\mathrm{VP}_2}$ t_3 t_1 Getrunken] hat sie ihr Bier$_3$ schnell$_1$ t_2
drunk has she$_{nom}$ her beer$_{acc}$ quickly

e. ?$[_{\mathrm{VP}_2}$ t_1 Begrüßen] will er Gesindel$_1$ t_2
greet wants to he$_{nom}$ riff-raff$_{acc}$

Both Fanselow and de Kuthy and Meurers take this as an argument that β_1 in the sentences in (41) does not leave the VP by scrambling prior to VP topicalization, and thus, that there is no remnant movement involved here. Since, then, incomplete category fronting of the type in (39) must in principle be possible without a first scrambling operation, they conclude that there is no reason to postulate remnant movement for other cases either (where β is an item that can undergo scrambling in other contexts).

This argument against remnant movement has already been addressed in Müller (1998, 204-210). The first thing to note is that the amelioration effect in (41) exclusively involves contexts in which, under the remnant movement analysis, β_1 undergoes extremely short, *locally string-vacuous* movement. Thus, instances of remnant movement where the preceding scrambling operation postulated under the remnant movement approach is not string-vacuous yield exactly the same kinds of deviance as the parallel examples in (40), and not well-formedness as in the cases in (41). This is shown by the examples in (42).

(42) a. (i) ??$[_{\mathrm{VP}_2}$ t_1 Gegeben] hat der Karl was$_1$ der Maria t_2
given has the Karl$_{nom}$ something$_{acc}$ the Marie$_{dat}$

(ii) ?*[$_{\mathrm{VP}_2}$ t$_1$ Gegeben] hat was$_1$ der Karl der Maria t$_2$
given has something$_{acc}$ the Karl$_{nom}$ the Marie$_{dat}$

b. ?*[$_{\mathrm{VP}_2}$ t$_1$ Geküsst] hat niemanden$_1$ der Fritz t$_2$
kissed has no-one$_{acc}$ the Fritz$_{nom}$

c. (i) ??[$_{\mathrm{VP}_2}$ t$_3$ t$_1$ Gegessen] hat der Karl roh$_1$ das Fleisch$_3$ t$_2$
eaten has the Karl$_{nom}$ raw the meat

(ii) ?*[$_{\mathrm{VP}_2}$ t$_3$ t$_1$ Gegessen] hat roh$_1$ der Karl das Fleisch$_3$ t$_2$
eaten has raw the Karl$_{nom}$ the meat

d. *[$_{\mathrm{VP}_2}$ t$_3$ t$_1$ Getrunken] hat schnell$_1$ sie ihr Bier$_3$ t$_2$
drunk has quickly she$_{nom}$ her beer$_{acc}$

e. ?*[$_{\mathrm{VP}_2}$ t$_1$ Begrüßen] will Gesindel$_1$ hier keiner t$_2$
greet wants to riff-raff$_{acc}$ here no-one$_{nom}$

Thus, it can be concluded that there is a problem with this argument against remnant movement because the contexts in (40) and (41) do not establish a minimal pair: In one case, scrambling leads to word order permutation; in the other case, movement is string-vacuous.[21] From this point of view, crucial examples that would show that items like indefinite *was*, *niemand*, *roh*, *schnell*, and *Gesindel* indeed cannot undergo scrambling at all would have to involve cases of string-vacuous scrambling as in (41), but without subsequent remnant movement; however, such structures will be string-identical to in situ variants (i.e., they are not only locally string-vacuous, but in fact globally string-vacuous), and will therefore be hard to justify.[22]

All this said, the difference between (41) on the one hand and (40) and (42) on the other hand of course needs to be accounted for in some way. In Müller (1998, 205), I suggest that the relevant movement operation that extracts β_1 from VP$_2$ prior to remnant topicalization in the examples in (41) might not be

[21] In addition, I have changed non-pronominal DPs as they show up in (40) into pronominal DPs in (41-a) and (41-e), in further violation of the minimal pair directive, so as to improve the acceptability of the sentences. The direct analogues involving non-pronominal DPs are actually still perceived as quite marked.

[22] Also note that some theories of scrambling explicity exclude globally vacuous application of the operation; see Ross (1967) and Frey & Tappe (1991), among others.

scrambling but, in fact, extraposition. However, as noted both in Müller (1998) and Fanselow (2002), there may also be a problem for this approach insofar as items to which extraposition must then be able to apply in these contexts seem to sometimes be precluded from undergoing other, well-established kinds of overtly visible extraposition. Furthermore, as Fanselow observes, if (41) involves extraposition, it is not quite clear why a language like English, which lacks scrambling but has extraposition, cannot have constructions of the type in (41) after all. Finally, Fanselow (2002, 103) argues that well-formed examples like (43-a) call the extraposition analysis into question since *wen* (‘whom’) occupies a position preceding the auxiliary *haben* (‘have’) in (43-a) that it cannot occupy in other contexts without VP topicalization; see (43-b).[23]

(43) a. [$_{\text{VP}_2}$ t_1 Geküsst] dürfte er schon öfter (t_2)
kissed might he$_{nom}$ already now and then
wen$_1$ (t_2) haben
someone$_{acc}$ have

b. *Er dürfte schon öfter t_1 geküsst wen$_1$
he$_{nom}$ might already now and then kissed someone$_{acc}$
haben
have

In view of all this, in what follows I will pursue two approaches to the data in (41), First, I outline an analysis in which the original assumption is maintained that the examples in (41) do indeed involve scrambling of β_1 prior to VP$_1$ topicalization; and after that I briefly reevaluate the extraposition approach.

A.1.1.1 Deriving (41) via Scrambling

Thus, suppose first that the data in (41) involve scrambling of the item with index 1. Obviously, then, there can be no *general* ban on scrambling of these items. At this point, it is worth investigating how a statement that certain items cannot undergo scrambling (conceived of as a movement operation) could be implemented theoretically in the first place. There are two obvious possibilities. The first one is that these items bear a property that characterizes them as not being targetable by the scrambling operation; this property could either

[23] The bracketed VP traces (t_2) in (43) are supposed to indicate neutrality with respect to leftward movement (scrambling) or rightward movement (extraposition) of the wh-phrase *wen* (‘whom’).

be encoded in terms of a feature that states an incompatibility with scrambling, or via the absence of a designated feature that is needed to qualify as a target for scrambling (such as the Σ feature I have assumed throughout this monograph). The second possibility is that ceteris paribus (i.e., if no other constraints unrelated to scrambling per se intervene that distinguish between classes of items), all items can in principle undergo scrambling; and restrictions on certain items in certain contexts are derived from output constraints that evaluate the well-formedness of scrambling structures by taking into account a variety of different interacting factors (which can be information-structural, semantic, prosodic, etc.) Clearly, the Fanselow/de Kuthy/Meurers argument against remnant movement presupposes the first approach. However, for a variety of independent reasons, I believe that the latter approach is correct (see Müller (1999)): The formal means that brings about scrambling (viz., Σ-driven movement) is dissociated from the formal means that evaluates the felicity of the resulting representation (viz., interacting information-structural, semantic, prosodic factors). Thus, the fact that an example like (44-a) (with scrambling of an indefinite inanimate object DP_1 to a position preceding a definite animate subject DP) is hardly possible (at least with normal intonation) cannot be taken as evidence that DP_1 (i.e., the linguistic expression *einen alten Baum* ('an old tree')) cannot undergo scrambling as an inherent property; as shown in (44-bcd), scrambling is vastly improved if the context is slightly different (with an indefinite subject instead of a definite subject, more material remaining in the rest of the clause, and less marked verb forms), and entirely inconspicuous if the landing site follows rather than precedes the subject (provided that information-structural, semantic, and prosodic requirements are optimally met).[24]

[24] Haider (2000a) assumes that the base order of dative and accusative arguments in German is variable: Some double object verbs have a base order of dative preceding and c-commanding accusative, some have a base order of accusative preceding and c-commanding dative, and some permit both orders. I have argued against such a variable approach, and for an invariant base order in Müller (1999). Still, it is worth noting that a ditransitive verb like *überlassen* ('leave something to someone') is unequivocally classified as a dative-before-accusative verb by Haider, and that (44-d) must therefore involve scrambling (resulting in a completely unmarked structure) even under the assumption that base order of object arguments can vary depending on the properties of the verb.

(44) a. ?*dass [$_{\text{DP}_1}$ einen alten Baum] Fritz t$_1$ fällte
that an old tree$_{acc}$ Fritz$_{nom}$ cut

b. dass [$_{\text{DP}_1}$ einen alten Baum] niemand t$_1$ zu fällen versuchen sollte
that an old tree$_{acc}$ no-one$_{nom}$ to cut try should

c. dass Fritz [$_{\text{DP}_1}$ einen alten Baum] einem jungen Baum t$_1$ vorzieht
that Fritz$_{nom}$ an old tree$_{acc}$ a young tree$_{dat}$ prefers to

d. dass Fritz [$_{\text{DP}_1}$ einen alten Baum] üblicherweise nur Experten t$_1$ überlässt
that Fritz$_{nom}$ an old tree$_{acc}$ typically only experts$_{dat}$ leaves

On this view, the only reason why scrambling of the β_1 items in (40) gives rises to various degrees of deviance is that the resulting output structures are negatively evaluated by interacting information-structural, semantic, and prosodic constraints. I will not try to speculate as to what individual properties of the β_1 items might be responsible in each case (in some cases – e.g., with prosodically light or reduced forms – the factors might seem to be fairly obvious, in others arguably less so). However, against this background, it is to be expected that the β_1 items that cannot undergo scrambling easily in contexts like those in (40) can in fact undergo scrambling in other contexts, where fewer (information-structural, semantic, prosodic) restrictions are imposed.[25] Of particular relevance here are contexts in which scrambling of β_1 does not have to cross a DP argument. Thus, in Müller (2000b), I argue that sentences with an unmarked order of a DP preceding a locative adverbial phrase in German (cf. Lenerz (1977)) must be derived by DP scrambling; as shown in (45-a), an indefinite pronoun *was* ('something') can easily undergo movement in this context. Furthermore, as noted in Heck & Müller (2000a), indefinite wh-pronouns can be scrambled in front of an adverbial CP in order to license a parasitic gap; see (45-b).

[25] In line with this, I would like to contend that most examples in (40) can be improved by modulating prosodic and/or information-structural factors.

(45) a. dass der Karl der Maria$_3$ was$_1$ [$_{\text{PP}}$ in der Kneipe]
that the Karl$_{nom}$ the Maria$_{dat}$ something$_{acc}$ in the pub
t$_3$ t$_1$ gegeben hat
given has

b. dass die Maria was$_1$ [$_{\text{CP}}$ ohne e$_1$ zu lesen] dem
that the Maria$_{nom}$ something$_{acc}$ without to read the
Fritz t$_1$ zurückgegeben hat
Fritz$_{dat}$ given back has

Interestingly, parasitic gap licensing by wh-indefinite scrambling can feed remnant VP topicalization in German; see (46). Given that there is strong evidence that (45-b) does indeed involve true parasitic gap licensing (see Assmann & Heck (2011; 2013), Assmann (2014) vs. Fanselow (2001)), this provides a direct argument that *was* can have undergone scrambling in (41-a).

(46) [$_{\text{VP}_2}$ t$'_1$ [$_{\text{CP}}$ Ohne e$_1$ zu lesen] dem Fritz t$_1$ zurückgegeben] hat
without to read the Fritz$_{dat}$ given back has
sie was$_1$
she$_{nom}$ something$_{acc}$

More generally, then, it seems to me that there is every reason to assume that the (relative) illformedness of examples such as those in (40) is not due to the β_1 items involved here not being accessible by scrambling; and that scrambling can therefore be assumed to underlie the examples in (41).

A.1.1.2 Deriving (41) via Extraposition

Given the reasoning of the last subsection, it may also be possible to critically re-evaluate at least some of the arguments suggesting that the examples in (41) do not involve extraposition. Recall that the first argument was that some items would then have to undergo extraposition which cannot normally undergo extraposition well in German. This is shown for wh-indefinites in (47-a). With other β_1 items in (41), the result of applying extraposition is more variable. For instance, extraposition of an adjective like *roh* ('raw') does not necessarily lead to complete inacceptability; see (47-b) (cf. Müller (1998, 209)).

(47) a. *dass er ihr t$_1$ gegeben hat was$_1$
that he$_{nom}$ her$_{dat}$ given has something$_{acc}$

b. ??dass er das Fleisch t$_1$ gegessen hat roh$_1$
that he$_{nom}$ the meat$_{acc}$ eaten has raw

However, extraposition, like scrambling, is a movement rule whose output is clearly evaluated by information-structural, semantic, and prosodic constraints. Thus, it is not at all clear that problems with extraposition in (47-a) and, to some extent, (47-b) should be formulated as constraints on a syntactic operation of extraposition as such. In line with this, factors like heaviness or focus can improve many examples, and on this view the main problem with (47-a) is that wh-indefinites are reduced forms (e.g., *irgend-was* $\rightarrow$ *was*, *irgend-wer* $\rightarrow$ *wer*) that can neither be phonologically heavy nor be focussed; as with scrambling, in locally string-vacuous contexts like those in (41), such factors can plausibly be assumed not to play a role. In principle, PPs (see (48-c)), APs (see (48-b)), and even DPs (see (48-a)) can undergo extraposition in German (see Müller (1996)). I would like to conclude from this that extraposition is not categorially selective at all in German.

(48) a. dass er ihr t$_1$ gegeben hat [$_{DP_1}$ Bücher, die sein
that he$_{nom}$ her$_{dat}$ given has books$_{acc}$ that$_{nom}$ his
Herz berührt hatten]
heart$_{acc}$ moved had

b. dass er das Fleisch t$_1$ gegessen hat [$_{AP_1}$ noch ganz roh und
that he$_{nom}$ the meat$_{acc}$ eaten has yet fully raw and
ohne Beilagen]
without side dishes

c. dass er schon t$_1$ gerechnet hatte [$_{PP_1}$ damit]
that he$_{nom}$ already reckoned had there-with

The second argument was that (49-a) cannot involve extraposition because extraposition would have to target a position in front of the auxiliary, which (49-b) suggests is impossible (both examples are repeated here from (43)).

(49) a. [$_{VP_2}$ t$_1$ Geküsst] dürfte er schon öfter (t$_2$)
kissed might he$_{nom}$ already now and then
wen$_1$ haben
someone$_{acc}$ have

b. *Er dürfte schon öfter t$_1$ geküsst wen$_1$
he$_{nom}$ might already now and then kissed someone$_{acc}$
haben
have

One might take this to indicate that extraposition can only go to the highest verbal projection in a clause in German. However, as noted by Haider (1993; 2010), under the extraposition analysis it must be the case that extraposition to a non-highest VP is an option after all, given the availability of topicalization in (50-a); if VP topicalization does not apply, illformedness results here just as it does in (49-b).

(50) a. [$_{\mathrm{VP}_2}$ t$_1$ Gerechnet [$_{\mathrm{PP}_1}$ damit]] wird keiner t$_2$ haben
reckoned there-with will no-one$_{nom}$ have

b. *Es wird keiner [$_{\mathrm{VP}_2}$ t$_1$ gerechnet [$_{\mathrm{PP}_1}$ damit]] haben
it will no-one$_{nom}$ reckoned there-with have

Thus, a straightforward way to account for these data is to postulate that extraposition can affect any verbal projection as a landing site in German; in addition, there is a filter that prohibits contexts where some item linearly intervenes between two members of a verbal cluster. This restriction is automatically lifted as soon as at least one of the three items involved in a V_a-XP-V_b configuration moves away; this is achieved by displacing VP_2 in (50-b) (which removes V_a and PP, leaving only V_b), and by displacing (the lower segment of) VP_2 in (49-a) (which removes V_a, leaving DP and V_b). Of course, the question arises as to what this V_a-XP-V_b filter follows from; but this issue is orthogonal to my concerns here (see Büring & Hartmann (1997) for discussion).

Assuming that a satisfactory solution can be found for the third problem as well (based on the question of why it is that English cannot have remnant movement constructions of the type in (41) if it has extraposition), there might therefore be reason to conclude that the data in (41) could in fact be structurally ambiguous between scrambling and extraposition of the items with index 1.[26]

[26] An obvious place to look for a solution to the third problem is the set of restrictions on case assignment in English, especially in view of the fact that Fanselow (2002) takes examples like (i-a) (see Phillips (2003)) to have a derivation with PP extraposition followed by remnant VP fronting (so as to have a stage of the derivation where *every girl* can c-command and thus license the co-indexed variable pronoun *her*).

(i) a. [$_{\mathrm{VP}_2}$ Given every girl$_3$ a book t$_1$] he certainly has t$_2$ [$_{\mathrm{PP}_1}$ on her$_3$ eleventh birthday]

b. *[$_{\mathrm{VP}_2}$ Given every girl$_3$] he certainly has [$_{\mathrm{DP}_2}$ a book]

A.1.2. Unexpected Anti-Freezing

In Müller (1998, 203-210), I discuss an unexpected anti-freezing effect with multiple remnant movement in German. An example like (51-a) involves PP_1 scrambling from VP_2 preceding VP_2 topicalization; (51-b) shows that PP_1 blocks further extraction of an R-pronoun from it, as a regular, expected instance of freezing. (51-c) is a variant of (51-a) that has locally string-vacuous PP_1 movement: PP_1 precedes the subject in (51-a) but follows the subject in (51-c). And, as shown in (51-d), this context gives rise to an a priori unexpected effect – PP_1 is transparent for extraction of the R-pronoun *da* ('there') even though it has undergone movement.

(51) a. [$_{VP_2}$ t_1 Gerechnet] hat wie immer [$_{PP_1}$ damit] keiner t_2
counted has as always therewith no-one

b. *[$_{VP_2}$ t_1 Gerechnet] hat da_3 wie immer [$_{PP_1}$ t_3 mit] keiner t_2
counted has there as always with no-one

c. [$_{VP_2}$ t_1 Gerechnet] hat wie immer keiner [$_{PP_1}$ damit] t_2
counted has as always no-one therewith

d. [$_{VP_2}$ t_1 Gerechnet] hat da_3 wie immer keiner [$_{PP_1}$ t_3 mit] t_2
counted has there as always no-one with

Thus, the pattern is similar to that discussed in the previous subsection: Extremly local string-vacuous movement feeding remnant fronting behaves differently from what one might normally expect, and this may be taken to cast doubt on the hypothesis that scrambling is involved here. The solutions to this problem considered in Müller (1998) are the same as the ones developed for categorial selectivity, viz., to either maintain that scrambling is involved after all, or reanalyze local scrambling as local extraposition. Let me consider the two approaches in turn, this time beginning with the extraposition analysis (since the scrambling analysis is somewhat more involved).

A.1.2.1. Deriving (51-d) via Extraposition

By assuming extraposition of PP_1 to end up in a higher (right-adjoined) position than a preceding scrambling operation moving da_3, both the Strict Cycle

On this view, the problem with (i-b) might be that *a book* cannot be assigned case by the verb (essentially this amounts to an adjacency effect of the type discussed in Stowell (1981) – a requirement that can be satisfied with leftward movement of DP to a higher position, under reconstruction).

Condition (SCC) and the Condition on Extraction Domain (CED) can be maintained in the derivation underlying (51-d); as a matter of fact, such a derivation has been presupposed in section 3.1 above, where I discuss multiple remnant movement (recall the discussion of (28-b) (= (51-b)) and (28-a) (= (51-d))). Evidently, PP_1 extraposition is not an analytic option in (51-b).

So far, so good. However, Fanselow (2002, 106-114) raises two problems for this analysis. First, a DP can play the role of the PP in a derivation that is structurally identical to that in (51-d) even though DPs "typically resist extraposition". This problem is acknowledged in Müller (1998) but not taken to be severe there; given the reasonings above, it can arguably be put aside (as we have seen, DPs *can* in principle undergo extraposition in German; also see Fanselow (1992a) for examples involving train announcements). Another problem raised by Fanselow that is somewhat more pressing at first sight is this: In Müller (1998, 207), I noted that an example like (52-a), which differs from (51-b) in that extraposition of PP_1 is not an analytic option, is "significantly degraded". Still, it seems that there is a contrast to (51-b), which is much less acceptable; and Fanselow also gives examples like (52-b) and (52-c), which do indeed seem to be fully well formed even though they are structurally very similar.

(52) a. ??[$_{VP_2}$ t_1 Rechnen] hat da_3 keiner [$_{PP_1}$ t_3 mit] [$_{VP_4}$ t_2 können]
count has there no-one$_{nom}$ with could

b. [$_{VP_2}$ t_1 Widerlegt] dürfte er sie da_3 ja wohl kaum [$_{PP_1}$ t_3 mit] haben
refuted might he her there yes well barely with have

c. [$_{DP_2}$ Welche Ansichten t_1] hat er da_3 denn [$_{PP_1}$ t_3 zu] geäußert ?
which opinions has he there then to uttered

As before, the amelioration effect only occurs with extremely local string-vacuous movement. (53-bc) illustrate that if PP_1 shows up further to the left, (52-bc) become ungrammatical (and recall that (51-b) is fully ungrammatical, (52-a) much less so).

(53) a. *[$_{\text{VP}_2}$ t$_1$ Widerlegt] dürfte er sie da$_3$ ja [$_{\text{PP}_1}$ t$_3$ mit] wohl
refuted might he her there yes with well
kaum haben
barely have

b. *[$_{\text{DP}_2}$ Welche Ansichten t$_1$] hat er da$_3$ denn [$_{\text{PP}_1}$ t$_3$ zu]
which opinions has he there then to
vermutlich geäußert ?
probably uttered

However, this problem can also be addressed if one assumes (as I have tentatively done above) that not only the highest verbal projection, but lower verbal projections too can qualify as targets for extraposition. On this view, the examples in (52) can have derivations where scrambling of *da*$_3$ is followed by remnant extraposition of PP$_1$ to a higher position, which is finally followed by remnant VP$_2$ topicalization. Where extraposition is clearly not an option (as in the examples in (53), and in (51-b)), illformedness results, as expected. There may be some intricate further questions raised by the analysis concerning the exact structure of verbal projections, and the role of vP vs. VP, but in general this approach would seem to work satisfactorily. Nevertheless, let me also address the question of what can be said if extraposition is not an option in (52), and if one assumes scrambling of PP$_1$ to be the only option here.

A.1.2.2. Deriving (51-d) via Scrambling

As a matter of fact, den Besten & Webelhuth (1990, 87-91) already addressed this issue, suggesting that movement (in this case: scrambling) does not create a barrier for extraction if it is highly local, and such an approach is also noted as a possible alternative (to the extraposition approach) in Müller (1998, 205). In Müller (2002, 224-225), a version of the CED is proposed (on the basis of other phenomena that are not related to scrambling in German) that incorporates den Besten and Webelhuth's insight; see (54).

(54) *Condition on Extraction Domain* (CED; revised):
a. Movement must not cross a barrier.
b. An XP γ is a barrier unless there is a head σ such that
a. σ c-selects γ.
b. σ and γ are in the same minimal domain.

(54) is only slightly less restrictive than the version of the CED in (4), with the requirement of complement status for non-barriers replaced by the combined requirements of c-selection (subcategorization) and co-occurrence in the same phrase. This way, all standard CED effects can still be derived (including in particular all freezing effects), whereas extremely local string-vacuous movement does not create an island; however, as soon as movement is slightly less local, targetting a higher phrase, extraction from the moved item will violate the CED. In sum, instances of "unexpected anti-freezing" turn out not to be unexpected at all.

Or do they? Recall that the difference between freezing and anti-freezing in α-over-β constructions cannot be fully derived by the conspiracy of the SCC and the CED in a local–derivational approach with intermediate, PIC-driven movement to all phrase edges (see section 1.2.4. above). I have suggested that the problem with freezing configurations is that an incriminating index on a remnant category α has not yet been removed via criterial movement of β when α itself reaches its criterial position. So the question now is this: Does the assumption that extremely local criterial movement of some item (as I have assumed it to be possible throughout this chapter; cf. footnote 14) keeps it transparent for extraction ensure that the Williams Cycle can be respected in the derivations underlying the examples in (52) but not in all other freezing contexts where criterial movement is less local? It turns out that one more assumption is needed in the local–derivational system, one that distinguishes between criterial and intermediate movement: The CED must ensure that extraction from a locally moved item is possible only when this item has reached its criterial position. This is accomplished by the formulation in (55).

(55) *Condition on Extraction Domain* (CED; extended):
a. Movement must not cross a barrier.
b. An XP γ is a barrier unless there is a head σ such that
   a. σ c-selects γ.
   b. σ and γ are in the same minimal domain.
   c. γ is not required in its position by an edge feature.

The new requirement in (55-c) distinguishes between intermediate and criterial movement; and it treats criterial positions and base positions as a natural class of contexts in which barrier status can be removed from an XP (provided that the phrase of the c-selecting head has not yet been left). Thus, an example like (52-b), repeated below as (56), can have a derivation where PP_1 *first* undergoes local criterial movement to a specifier of VP_2 (in accordance with (17-b)), thereby contaminating VP_2's movement-related (i.e., topic) feature with its index and immediately decontaminating it since it has reached a criterial position (see the definitions in (13-a), (13-b)). Only *then* does *da*$_3$ extract from PP_1, to an outer specifier (as required by (13-b)); this option of *da*$_3$ movement following PP_1 movement exists because of the version of the CED in (55). (It would not be available under the CED in (4).) Movement of *da*$_3$ leaves an index on VP_2's movement-related feature (which must eventually be removed before VP_2 undergoes a criterial movement step), and it may also leave an index on PP_1's movement-related (i.e., Σ) feature, depending on whether or not this feature is still accessible at this point (an issue about which I have been silent since it has not played a role in the analysis so far) – but even if it does, this contamination comes too late to create any problems for the Williams Cycle because PP_1 has already undergone its *criterial* movement step. After this, the derivation proceeds more or less exactly as discussed before: *da*$_3$ is extracted from VP_2 to its criterial position (while VP is still in situ), thereby decontaminating VP_2, and then (a lower segment of) VP_2 is also extracted to a higher specifier of the same domain, driven by edge features. From this point onwards, only VP_2 moves, until it reaches its final landing site.

(56) [$_{VP_2}$ t_1 Widerlegt] dürfte er sie da_3 ja wohl kaum [$_{PP_1}$ t_3 mit] haben
refuted might he her there yes well barely with have

To conclude, by postulating that extraction from an XP that has undergone extremely local criterial movement can satisfy the CED, the anti-freezing effect with multiple movement in examples of the type in (51-d), (52) follows both under the standard approach to remnant movement that solely relies on the CED and the SCC *and* under the new buffer-based approach introduced in the present chapter that envisages highly local intermediate movement steps and employs symbol lists on movement-related features as buffers. Thus, the remnant movement approach to constructions of the type in (1) is not called into question, even if examples like those in (51-d), (52) involve locally string-vacuous scrambling rather than locally string-vacuous extraposition.

A.1.3. Does Scrambling Exist?

The most far-reaching, radical objection to remnant movement analyses of constructions like the one in (1) is to claim that scrambling does not exist as a genuine movement operation. On this view, the relevant phenomena must involve variable base-generation, accompanied by either a host of different head movement operations at LF (as in Fanselow (2001)) or by unconstrained reanalysis operations (as in De Kuthy & Meurers (2001)). Seriously discussing the issue whether scrambling exists as a syntactic movement operation in German would lead me too far astray at this point. For present purposes, it may suffice to state that in order to show that scrambling does not exist, the many cases of parallel behaviour of well-established movement constructions (wh-movement, topicalization, relativization) and free word order constructions in German with respect to locality and other constraints on movement that have been pointed out in work like Grewendorf (1989), Webelhuth (1992), and Müller (1995) need to be addressed, and accounted for in some other, non-conspiratorial way; and I would like to contend that this has not yet been accomplished.[27]

A.2. Alternative Generalizations about Remnant Movement

The generalizations about α-over-β configurations that I have adopted in the present chapter (freezing, anti-freezing, Müller-Takano generalization) are not co-extensive with those underlying either Grewendorf's (2003) or Abels's (2008) approaches to α-over-β configurations (which also differ from one another in various respects). Interestingly, both Abels and Grewendorf assume that a version of the Williams Cycle restricts the interaction of movement operations in α-over-β derivations, and in their analyses the Williams Cycle exerts a much more direct influence than it does in the analysis developed in the present chapter. Thus, Grewendorf (2003; 2004) and Abels (2008) suggest that a Williams Cycle-type approach to improper movement can be modified so as to not only cover cases of two operations applying to a single item α, but also

[27] Furthermore, several pieces of empirical evidence brought forward by Fanselow (2001) and De Kuthy & Meurers (2001) strike me as inconclusive at best. For instance, De Kuthy & Meurers (2001, 149) adduce putative exceptions to the Specificity constraint on extraction from DP that involve *von*- ('by') phrases; but these are known to often involve external generation of an optional argument instead of extraction (see Koster (1987, 196f.), Cinque (1990, 47), Sternefeld (1991, 121), Müller (1995, 397f.), Barbiers (2002, 54), and Gallego (2007, 349), among others).

cases of two operations applying to two different items α, β that are initially in a dominance relation ([$_\alpha$ ... β ...]), with both items eventually targetting an α-external position.[28] This presupposes an extended version of the Williams Cycle according to which movement of either α or β in [$_\alpha$ ... β ...] has consequences for movement of the other item.

A.2.1. Abels' Analysis

In Abels' (2008) approach, the more abstract notion of being *affected* by movement (rather than simply having *undergone* movement) is employed in his version of the Williams Cycle; (57) is his version of Williams' Generalized Ban on Improper Movement (GBOIM) (cf. (10) of chapter 2), which he dubs Generalized Prohibition against Improper Movement (GENPIM):

(57) *Generalized Prohibition against Improper Movement* (GENPIM):
No constituent may undergo movement of type τ if it has been affected by movement of type σ, where $\tau < \sigma$ according to the hierarchy A-movement < scrambling < wh < topicalization.

Basically, affectedness is then defined in such a way that if β moves out of α in [$_\alpha$ ... β ...], all the nodes on its movement path, including α (and β) are affected by β-movement; and if α moves somewhere with β still in it, β is affected by α-movement (as is α, and everything else dominated by α). More formally, Abels' (2008) concept of affectedness is defined in (58).[29]

(58) *Affectedness of constituents*:
A constituent γ is affected by a movement operation iff
a. γ is reflexively contained in the constituent created by movement, and
b. γ is in a (reflexive) domination relation with the moved con-

---

[28] That said, Grewendorf (2003; 2004) does not actually make reference to Williams' work; also, as we will see, Grewendorf's analysis is silent on standard cases of improper movement that do not involve α-over-β configurations.

[29] As a side remark, note that this extension of the Williams Cycle to α-over-β configurations of multiple movement is intrinsically non-local – huge amounts of structure are affected (in this technical sense) by any given movement operation: "Every movement has an effect on a well-defined, *potentially very big*, part of the structure" (see Abels (2008, sect. 2.1.); my emphasis).

stituent.

The restrictions that can be derived along these lines are orthogonal to the remnant/non-remnant (or bound/unbound trace) split that I have presupposed so far: First, it is predicted that freezing effects (with bound traces) should not arise if β targets a higher type of position than α; e.g., if α undergoes scrambling, and β (which has thus been affected by scrambling since it is contained in α) is extracted from α to SpecC by wh-movement. On the other hand, a freezing effect is predicted if α moves to a higher type of position (say, SpecC, by wh-movement) than β that is extracted from α (say, to a matrix Specv position, by scrambling). Second, it is predicted that anti-freezing effects (with unbound traces) should also not uniformly arise: An anti-freezing effect is created when β moves to a lower kind of position than the position type occupied by (the remnant category) α in a subsequent movement step; e.g., when β undergoes scrambling and α undergoes topicalization, as in standard examples of the type in (1). In contrast, there should be no anti-freezing effect with remnant movement if movement of β ends up in a higher kind of position than subsequent movement of α; e.g., when β undergoes topicalization, and α scrambling into the matrix clause.

For reasons discussed in some detail in Müller (2011), I do not think that relativizing freezing effects with bound traces is empirically warranted in German; I think if the few putative counter-examples are properly analyzed (see footnote 27 for one typical case), the assumption that all instances of movement (perhaps with the exception of extremely local movement, as just discussed in the first part of the appendix) create islands can (and, therefore, should) be maintained.

In the same way, I take it that the restrictions on unbound traces, i.e., on remnant movement, that are derivable under Abels' approach – and in particular the predicted exceptions from anti-freezing – are artefacts of the analysis. This holds, e.g., for illicit structures where β in [$_\alpha$ ...β...] undergoes topicalization and α undergoes wh-movement to a higher position, which are excluded in Abels's (2008) analysis by the GENPIM (but not in Grewendorf's (2003) analysis; see below); arguably this instantiates a more general topic island effect that is independent of α-over-β configurations (see Müller & Sternefeld (1993), among others). Thus, remnant movement is indeed impossible in (59-a), where embedded topicalization moves PP_1 out of DP_2 to the embedded topic position, and wh-movement subsequently transports the remnant DP_2 to the matrix SpecC position of an interrogative C. However, (59-b) is ungrammatical in just the same way; here, there is embedded topicalization of DP_1 followed by matrix topicalization of the remnant VP_2 (this would fall under the Müller-Takano generalization, but these effects are in fact not excluded in Abels' approach; see below). Furthermore, exactly the same effect shows up in constructions where the two moved items are not in a dominance relation at any point of the derivation, and the first movement operation therefore cannot lead to affectedness of the item that participates in the second movement operation; see (59-c). On the other hand, whatever accounts for the strict island nature of embedded topicalization in (59-c) will automatically cover (59-a) (and, perhaps redundantly if Müller-Takano effects are to be derived by a designated constraint, (59-b)) as well, irrespective of assumptions about remnant movement (or, more generally, α-over-β configurations).[30]

(59) a. *[$_{DP_2}$ Welches Buch t$_1$] glaubst du [$_{CP}$ [$_{PP_1}$ über dieses
which book$_{acc}$ believe you about this
Thema] hat der Karl t$_2$ gelesen ?
topic has the Karl$_{nom}$ read

b. *[$_{VP_2}$ t$_1$ Zu lesen] glaube ich [$_{CP}$ [$_{DP_1}$ dieses Buch] hat
to read believe I this book$_{acc}$ has
keiner t$_2$ versucht
no-one$_{acc}$ tried

c. *[$_{DP_3}$ Dem Karl] glaube ich [$_{CP}$ [$_{DP_1}$ dieses Buch] hat
the Karl$_{dat}$ believe I this book$_{acc}$ has
Maria t$_3$ t$_1$ zu geben versucht
Maria$_{nom}$ to give tried

In some other cases, the data underlying Abels' analysis are controversial. This holds, e.g., for [$_\alpha$ ...β...] structures where one item undergoes A-movement, and the other item undergoes scrambling. In Abels's (2008) analysis, a combination is permitted where β undergoes A-movement and α undergoes remnant scrambling; and a combination is excluded where β undergoes scrambling and α undergoes remnant A-movement. Examples like those in (60-ab) are as-

[30] Also note that this does not necessarily hold for *bound* traces, as in (3-a), where there is movement *out of* a topicalized item, not *across* a topicalized item, as in the case presently under consideration.

sumed to substantiate the former point (with judgements as provided by Abels (2008) and Grewendorf (2003), respectively).

(60) a. ?dass [$_{\text{VP}_2}$ t_1 von einem Studenten angefasst] kein einziges Reagenzglas$_1$ t_2 werden durfte
that by a student touched no single test tube become may.PAST

b. dass [$_{\text{VP}_2}$ t_1 zu küssen] der Student$_1$ von Maria t_2 versucht wurde
that to kiss the student by Maria tried was

These structures may in principle be legitimate (even though they do not really strike me as particularly felicitous).[31] However, ruling out the reverse combination of scrambling followed by remnant A-movement does not seem to be warranted empirically; German examples like (61-a) (from Takano (2000)) and (61-b) seem to be mildly degraded at worst, and certainly not less acceptable than those in (60).

(61) a. dass [$_{\text{DP}_2}$ ein Buch t_1] niemandem [$_{\text{PP}_1}$ über die Liebe] t_2 gegeben worden ist
that a book no-one about the love given been is

b. dass [$_{\text{DP}_2}$ ein Artikel t_1] gestern [$_{\text{PP}_1}$ darüber] t_2 veröffentlicht wurde
that an article$_{nom}$ yesterday about it published was

Next, there are inherent difficulties with the extremely liberal notion of affectedness of constituents in (58). The problem is that if β moves first in an α-over-β configuration such as (62), (58) does in fact not discriminate between movement to a position properly outside of α, as in (62-a) (which is the configuration that Abels is concerned with), and movement to a specifier of α, as in (62-b).

---

[31] In the approach developed in the present chapter, wellformedness is predicted for the examples in (60) if the two moved items end up in different domains, with the movement triggered by two distinct features – a scrambling feature and an EPP feature.

(62) ... $[_\alpha$ ... β]

a. ... β ... $[_\alpha$ ... t_β]

b. ... $[_\alpha$ β $[_\alpha$ ... t_β]

However, as noted by Philipp Weisser (p.c.) and Grewendorf (2013), the GEN-PIM makes predictions for the movement of CPs that are not empirically confirmed. For instance, a CP α in which wh-movement of some category β has applied should not be able to undergo either scrambling or A-movement. The reason is that after wh-movement of some XP β to the specifier of C′ (= α), CP (= γ in the sense of (58)) is affected by this movement operation because (i) CP is reflexively contained in the constituent created by movement (which is CP itself), and (ii) CP is in a domination relation with the moved constituent β. This prediction is not borne out; see (63-ab) (with scrambling and A-movement, respectively, of a CP in which wh-movement has taken place).

(63) a. dass $[_{CP_2}$ wen$_1$ sie t_1 dort getroffen hat] keiner der Anwesenden t_2 sagen konnte
that whom$_{acc}$ she there met has no-one$_{nom}$ of the attendees say could

b. dass $[_{CP_2}$ wohin$_1$ sie t_1 fährt] dem Karl nicht t_2 mitgeteilt wurde
that where she$_{nom}$ goes the Karl$_{dat}$ not told was

Finally, it is worth pointing out that there is an inherent tension between a Williams Cycle approach to Müller-Takano effects, and ‘proper’ cases of successive-cyclic movement as they figured prominently in chapter 2 above, in Abels’ analysis. Thus, Abels (2008) assumes that the second movement operation in remnant movement constructions must always go to a position that is *not* of a *lower* type than the first movement operation; see the definition of (57) above. Under this assumption, standard cases of proper and improper movement as they were discussed in the preceding chapter are unproblematic, but it also means that there can be no Williams Cycle-based account of Müller-Takano effects ((59-b), e.g., cannot be ruled out in this way); and this is indeed conceded by Abels.

A.2.2. Grewendorf's Analysis

Consider next the analysis developed in Grewendorf (2003; 2004). The analysis is first and foremost designed to capture properties of remnant movement, which is assumed to be regulated by the constraint in (64).

(64) *Improper Remnant Movement Constraint*:
Remnant movement is prohibited unless it is of a higher type than remnant-creating movement.

The hierarchy of positions that capture movement types postulated here is A-movement < scrambling < topicalization < wh-movement. Thus, as far as remnant movement is concerned, there are two major differences between the (otherwise similar) analyses of Abels (2008) and Grewendorf (2003; 2004). First, whereas Abels assumes that the second movement operation must not be of a lower type than the type of the first movement operation, according to (64) the second movement operation must always go to a position of a *higher* type. Accordingly, in Grewendorf's (2003) approach, the Müller-Takano generalization can be derived (because in the relevant cases, movement is to a position of the *same* type, not to a position of a *higher* type), but, unlike Abels's account, this account cannot be generalized to standard cases of proper and improper movement applying to one and the same item (e.g., movement from SpecC to SpecC would be blocked) – thus, we end up with a version of the Williams Cycle that regulates the proper and improper combinations of movement operations applying to two items in remnant movement constructions but not the proper and improper combinations of movement operations applying to one and the same item.

In addition to this, there is a second substantial difference to the approach developed by Abels: The order of topicalization and wh-movement is reversed.

Thus, unlike Abels' more far-reaching analysis, Grewendorf's approach derives the Müller-Takano generalization as a special case, and per se it does not say anything about freezing contexts, i.e., cases where α ends up in a lower position than β (since (64) is explicitly confined to remnant movement contexts). Still, as for anti-freezing contexts, Grewendorf's analysis predicts the existence of further restrictions, just like Abels' analysis does. To wit, scrambling from DP followed by remnant A-movement of DP is predicted to be ungrammatical, which does not strike me as an attractive consequence, at least not for German (see the examples in (61) above). Furthermore, by adopting a reverse hierarchy of topicalization and wh-movement, the spurious effects vis-à-vis topic island examples (recall (59-a) vs. (59-b), (59-c)) that arise in Abels' approach can be avoided; however, now the prediction arises that wh-movement from XP should be unable to feed topicalization of XP. As a matter of fact, examples of this type are quite acceptable in German; they are certainly not worse than other cases of topicalization across a wh-island; see the example in (31) and the surrounding discussion in section 3.3. above. As a matter of fact, Grewendorf (2003) acknowledges this empirical fact, presenting the example in (65), which he assumes to pose "a serious problem for generalization [(64)]."

(65)	[$_{\text{VP}_2}$ t$_1$	Zu	überreden]	weiß	ich	nicht [$_{\text{CP}}$	wen$_1$	sie	t$_2$	versucht	hat]
		to	persuade	know	I	not	whom$_{acc}$	she$_{nom}$		tried	has

Grewendorf's solution to this problem is to suggest that "in this particular case" of topicalization out of wh-islands, "topicalization is in fact focus movement", so that (64) can be respected after all (if focus movement outranks wh-movement on the hierarchy of movement types); still, whatever the merits of such a reanalysis may be, it seems clear that the simplest approach will be one where this problem does not arise in the first place. More generally, then, I would like to conclude from the discussion of Abels' and Grewendorf's Williams Cycle-based approaches that there is no reason to impose further restrictions on remnant movement constructions; and that the three generalizations that the present chapter set out to derive in terms of syntactic buffers are empirically well motivated for German.

Chapter 4

Resumptive Movement

1. Introduction

In this final chapter, I will be concerned with evidence for postulating buffers in the domain of resumption. Empirically, I will focus on resumptive strategies in long-distance dependencies in German and some Slavic languages. Based on the assumption that the only way to model resumption in a local–derivational (e.g., phase-based) approach to syntax is to postulate a movement dependency (at least as long cyclic Agree is excluded as an option, cf. footnote 4 from chapter 1), instances of resumptive movement that can cross what otherwise exists as an island for non-resumptive movement give rise to a backtracking problem: At the stage where the island is crossed by a moved item, it must be known whether there is a resumptive pronoun in the base position or not, but this position is typically too far away to be accessible at this point. Moreover, there are cases of resumptive movement where the moved item not only *can* cross an island but actually *has to* cross an island (and the German construction I will focus on below is one of them), and this also necessitates reference to information from earlier derivational stages that does not seem to be present anymore. As in the previous two chapters, I contend that a simple solution for these backtracking problems is available if moved items are equipped with buffers (in the form of lists that act as values of movement-related features) which store, in a minimal way, crucial aspects of the derivational history – in the case at hand, the fact that a resumptive pronoun has been split off the moved item in the base position.

All that said, a qualification is in order here: It is not my goal to attempt to come up with a comprehensive account of resumption in German (or Slavic), let alone to develop an approach that covers all known instances of cross-linguistic variation (which is substantial; see McCloskey (2006) for an overview, and Boeckx (2003) for the attempt to develop a unified analysis). At the end of the chapter, I will sketch possible ways in which the analysis to be developed below might be extended to some other cases of resumption that exhibit somewhat different properties, but it should be kept in mind that throughout, the focus is on solving island-related backtracking problems as they arise under a local–derivational approach (and, incidentally, also under a local–declarative approach; see page 141 below).

Let me begin by highlighting the properties of a resumptive strategy with relativization in German.

2. Resumption in German

Varieties of German exhibit resumptive relativization constructions of a type similar to those known from Swiss German and Southern German dialects (see Riemsdijk (1989), Salzmann (2006; 2012)), but with somewhat different properties. The first thing to note is that in cases of clause-bound dependencies that are completely transparent for standard movement, this resumption strategy is not available. The examples in (1) illustrate that movement of an empty relative operator to the specifier of a relative clause complementizer *wo* ('where') is possible for accusative objects (cf. (1-a)) and nominative subjects (cf. (1-b)), stricly blocking the resumptive strategy (here involving a pronoun *es* ('it')) in these contexts. In the case of dative object relativization (cf. (1-c)), neither strategy is available in German (in contrast to varieties of Swiss German, where both strategies can be legitimate, and optionality arises; see Salzmann (2012).)

(1) a. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich t$_1$/*es$_1$ gelesen habe]
this is a book where I it$_{acc}$ read have

b. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] t$_1$/*es$_1$ mir gefallen hat]
this is a book where it$_{nom}$ me$_{acc}$ pleased has

c. Das ist ein Mann [$_{CP}$ Op$_1$ [$_C$ wo] ich *t$_1$/*ihm$_1$ gedankt habe]
this is a man where I him$_{dat}$ thanked have

Before proceeding, it should be pointed out that the non-resumptive strategies in (1-a) and (1-b) are confined to highly colloquial, substandard varieties of German, and are generally stigmatized.[1] In what follows, I will not take this to be particularly significant from a theoretical point of view: There is an alternative relativization strategy involving an overt relative pronoun, which is clearly preferred by normative grammar (and, accordingly, the only strategy that can be heard or read in the media); cf. (2-abc).[2]

(2) a. Das ist ein Buch [$_{\text{CP}}$ das$_1$ [$_{\text{C}}$ Ø] ich t$_1$/*es$_1$ gelesen habe]
this is a book I it$_{acc}$ read have

b. Das ist ein Buch [$_{\text{CP}}$ das$_1$ [$_{\text{C}}$ Ø] t$_1$/*es$_1$ mir gefallen hat]
this is a book it$_{nom}$ me$_{acc}$ pleased has

c. Das ist ein Mann [$_{\text{CP}}$ dem$_1$ [$_{\text{C}}$ Ø] ich t$_1$/*ihm$_1$ gedankt habe]
this is a man I him$_{dat}$ thanked have

Another clause-bound context that is transparent for movement involves postpositions (i.e., postposition stranding). Suppose, following standard reasoning, that a precondition for extraction from PP in German is that a left-peripheral

---

[1] This also accounts for the fact that the sentence in (i), a real-life utterance of a German football player, is supposed to be funny, insinuating that the player should not criticize the French-speaking colleagues on his (German) team for not knowing German when he himself apparently does not know it sufficiently well, as indicated by the use of the empty operator plus *wo* relativization strategy.

(i) Ich lerne nicht extra Französisch für die Spieler [$_{\text{CP}}$ Op$_1$ wo t$_1$ dieser
I learn not deliberatley French for the players where this
Sprache nicht mächtig sind]
language$_{gen}$ not capable are

[2] Note that C can in principle also be realized by *wo* in (2-a), (2-b), and (2-c), which then again relegates the sentences to a stigmatized substandard variety of German. One may speculate that normative prohibition against using *wo* in relative clauses (no matter what the relative operator looks like, i.e., whether it is an overt pronoun like *dem* or an empty operator) also plays a role in accounting for the perceived illformedness of (1-c): On this view, two factors conspire in cases like (1-c): On the one hand, it has been noted that there is a general, independently verifiable recoverability problem with dative arguments in the absence of any morphological cues (see Bayer et al. (2001)), which presumably also underlies the fact that the resumptive strategy is an option in this transparent context in varieties of Swiss German (as argued by Salzmann (2012)); and on the other hand, there is the general prescriptive ban on using *wo*, which does not show up in this form in Swiss German and regional varieties of German.

specifier position of P can be occupied by the moved item (see Riemsdijk (1978), Koster (1987), Grewendorf (1989), Abels (2012a), among many others). On this basis, it looks as though the empty operator Op can be merged to the left of a P element like *für* ('for') (see (3-a)), just like R-pronouns like *da* ('there') (see (3-b)), and in contrast to regular pronouns like *ihn* ('him') or *den* ('which') (the latter would also qualify as the regular relative pronoun here; see (3-c)). Resumption with either the R-pronoun or the standard personal pronoun is impossible in this context (see (3-de)).

(3) a. Das ist ein Vorschlag [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht [$_{PP}$ t$_1$ für]
this is a proposal where I not for
gestimmt habe]
voted have

b. Da$_1$ habe ich nicht [$_{PP}$ t$_1$ für] gestimmt
there have I not for voted

c. *Das ist ein Vorschlag [$_{CP}$ den$_1$ [$_C$ Ø] ich nicht [$_{PP}$ für t$_1$]
this is a proposal which I not for
gestimmt habe]
voted have

d. *Das ist ein Vorschlag [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht [$_{PP}$ da-für]
this is a proposal where I not there for
gestimmt habe]
voted have

e. *Das ist ein Vorschlag [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht [$_{PP}$ für ihn$_1$]
this is a proposal where I not for him$_{\mathrm{inan}}$
gestimmt habe]
voted have

Interestingly, in this context where the strategy preferred by normative grammar is impossible (cf. (3-c)), it seems that the empty operator movement strategy (cf. (3-a)) is not only possible; it is in fact much less perceived as belonging to substandard (or 'dialectal') varieties than the examples in (1-a) and (1-b), where there is an alternative with an overt relative pronoun, and without a complementizer *wo* (cf. (2-a), (2-b)) – at least, this holds for those speakers of German who permit postposition stranding in the first place, i.e., for whom

(3-b) is unproblematic.[3]

Another context that is transparent for extraction in German involves postposition stranding within object DPs (see Koster (1987) and Grewendorf (1989), among others). Here, the strategy in terms of regular empty operator movement and a complementizer *wo* is available, and resumption is blocked (both with an R-pronoun and a normal personal pronoun); see (4).

(4) a. Das ist ein Plan [$_{CP}$ Op$_1$ [$_C$ wo] er [$_{DP}$ ein Buch [$_{PP}$ t$_1$
this is a plan where he$_{nom}$ a book$_{acc}$
über]] geschrieben hat]
about written has

b. ?*Das ist ein Plan [$_{CP}$ Op$_1$ [$_C$ wo] er [$_{DP}$ ein Buch
this is a plan where he$_{nom}$ a book$_{acc}$
[$_{PP}$ da$_1$-r-über]] geschrieben hat]
there EPENTH about written has

---

[3] One might think that the item *wo* in (3-a) is in fact not a complementizer, but a moved wh-pronoun of the same type as the wh-marked R-pronoun in (i-a), which would then be used as a relative pronoun in (3-a) in roughly the same way as the wh-pronoun in (i-b) is used as a relative pronoun.

(i) a. Wo$_1$ hat sie [$_{PP}$ t$_1$ für] gestimmt ?
where has she for voted

b. Das ist ein Buch [$_{CP}$ was$_1$ keiner t$_1$ kaufen wollte]
this is a book what$_{acc}$ no-one$_{nom}$ buy wanted

Such a reanalysis, however, is unlikely to be correct. Like, e.g., the marked wh-relative pronoun *was*, the R-pronoun *wo* cannot bear a plural feature (cf. (ii-a)) and, accordingly, leads to illformedness under a plural interpretation in the absence of an explicit distributor like *alles* ('all'); and it cannot be interpreted as human either in most varieties of German (cf. (ii-c); see Müller (2000a) for systematic exceptions in Northern varieties, where such a sentence is indeed well formed). However, in a context like (3-a), these restrictions are lifted (cf. (ii-b) and (ii-d), respectively), which supports the analysis in terms of an empty operator (that is not subject to special number and animacy requirements) and a complementizer (rather than pronoun) status of *wo*.

(ii) a. Hier sind einige Vorschläge. Wo$_1$ hat sie ?*(alles) [$_{PP}$ t$_1$ für] gestimmt ?
here are some proposals where$_{plur}$ has she all for voted

b. Hier sind einige Vorschläge dabei [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht [$_{PP}$ t$_1$ für]
here are some proposals included where I not for
stimmen werde
vote will

c. *Das ist ein Plan [$_{CP}$ Op$_1$ [$_C$ wo] er [$_{DP}$ ein Buch
this is a plan where he$_{nom}$ a book$_{acc}$
[$_{PP}$ über ihn$_1$]] geschrieben hat]
about him$_{\text{inan}}$ written has

In contrast, resumption with an R-pronoun improves signficantly in subject contexts, where standard movement is excluded, whereas resumption with a normal personal pronoun continues to be blocked; see (5)).

(5) a. *Das ist ein Plan [$_{CP}$ Op$_1$ [$_C$ wo] [$_{DP}$ ein Buch [$_{PP}$ t$_1$ über]]
this is a plan where a book$_{nom}$ about
Maria beeindruckt hat]
Maria$_{acc}$ impressed has

b. ?Das ist ein Plan [$_{CP}$ Op$_1$ [$_C$ wo] [$_{DP}$ ein Buch
this is a plan where a book$_{nom}$
[$_{PP}$ da$_1$-r-über]] Maria beeindruckt hat]
there EPENTH about Maria$_{acc}$ impressed has

c. *Das ist ein Plan [$_{CP}$ Op$_1$ [$_C$ wo] [$_{DP}$ ein Buch [$_{PP}$ über
this is a plan where a book$_{nom}$ about
ihn$_1$]] Maria beeindruckt hat]
him$_{\text{inan}}$ Maria$_{acc}$ impressed has

Consider next cases of dependencies that are not (strictly) clause-bound. Here, the data are not always crystal-clear, and there is some variation among speakers. Resumption would seem to be completely impossible with restructuring verbs like *versuchen* ('try') as in (6-a), which on many analyses do in fact not involve a biclausal structure (as is indicated here). The resumption strategy improves somewhat with non-restructuring verbs like *ablehnen* ('reject'), especially if the infinitival complement is extraposed; see (6-b). In a dependency crossing a finite CP embedded under a bridge verb, a resumptive pronoun be-

---

c. *Wo$_1$ hast du gerade [$_{PP}$ t$_1$ mit] geredet ?
where have you just with spoken

d. Das ist jemand [$_{CP}$ Op$_1$ [$_C$ wo] ich gerade [$_{PP}$ t$_1$ mit] geredet habe]
this is someone where I just with spoken have

comes tolerable, see (6-c).[4] Adding negation in the matrix clause (see (6-d)) further improves resumption, and resumption becomes perfect with non-bridge predicates like *know*, as in (6-e).

(6) a. *Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{VP}$ es$_1$ zu kaufen]
this is a book where I it to buy
versucht habe]
tried have

b. ?*Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich abgelehnt habe [$_{CP}$ es$_1$ zu
this is a book where I refused have it to
kaufen]]
buy

c. ??Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich gedacht habe [$_{CP}$ dass sie
this is a book where I thought have that she
es$_1$ kaufen würde]]
it buy would

d. ?Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht gedacht hätte
this is a book where I not thought had
[$_{CP}$ dass sie es$_1$ kaufen würde]]
that she it buy would

e. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich gewusst habe [$_{CP}$ dass sie
this is a book where I known have that she
es$_1$ kaufen würde]]
it buy would

The increasing degree of wellformedness of resumption from top to bottom in (6) correlates with a decrease of acceptability of the standard movement option. This is shown for movement of an empty operator in (7).

(7) a. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{VP}$ t$_1$ zu kaufen] versucht
this is a book where I to buy tried
habe]
have

[4] Also see Brandner & Bucheli's (2014) empirical study of Swiss German varieties, where constructions like (6-c) emerge as a widespread strategy of dependency formation.

b. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich abgelehnt habe [$_{CP}$ t$_1$ zu kaufen]]
this is a book where I refused have to buy

c. ??Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich gedacht habe [$_{CP}$ dass sie t$_1$ kaufen würde]]
this is a book where I thought have that she buy would

d. ?*Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht gedacht hätte [$_{CP}$ dass sie t$_1$ kaufen würde]]
this is a book where I not thought had that she buy would

e. *Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich gewusst habe [$_{CP}$ dass sie t$_1$ kaufen würde]]
this is a book where I known have that she buy would

As before, the well-formed examples in (7) belong to substandard or regional varieties and are stigmatized by proponents of a normative approach to grammar. Versions of (7-ab) involving an overt relative pronoun like *das* ('that') and an empty complementizer are fully well formed (cf. (8-ab)). In contrast, extraction of an overt relative pronoun from a finite clause is not completely unproblematic. As noted by Bayer & Salzmann (2009), many speakers of German do not permit long-distance relativization here (in contrast to wh-movement or topicalization); see, e.g., (8-c).[5] Movement of *das* becomes even worse with matrix negation and under non-bridge verbs; see (8-de).[6]

(8) a. Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich [$_{VP}$ t$_1$ zu kaufen] versucht habe]
this is a book that I to buy tried have

[5] Also see Plank (1983, 11) and Grewendorf (1988, 92) for some preliminary remarks in this direction.

[6] Note that the illformedness of (8-c) is not related to homophony of the relative pronoun and the embedded complementizer; Bayer & Salzmann (2009) give examples where the morphological forms of the two items are distinct.

b. Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich abgelehnt habe [$_{CP}$ t$_1$ zu
this is a book that I refused have to
kaufen]]
buy

c. ?*Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich gedacht habe [$_{CP}$ dass sie t$_1$
this is a book that I thought have that she
kaufen würde]]
buy would

d. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich nicht gedacht hätte [$_{CP}$ dass
this is a book that I not thought had that
sie t$_1$ kaufen würde]]
she buy would

e. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich gewusst habe [$_{CP}$ dass sie t$_1$
this is a book that I known have that she
kaufen würde]]
buy would

Turning finally to island contexts in which standard movement is always blocked, resumption becomes the only available strategy to express a long-distance dependency. The acceptability of resumption is shown for Complex Noun Phrase Condition (CNPC) islands in (9-a), and for adjunct islands in (9-b); both sentences are completely unmarked.[7]

(9) a. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{DP}$ einen Mann t$_{CP}$]
this is a book where I a man$_{acc}$
getroffen habe [$_{CP}$ der es$_1$ gelesen hat]]
met have who it read has

b. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich eingeschlafen bin
this is a book where I fallen asleep have
[$_{CP}$ nachdem ich es$_1$ gelesen habe]]
after I it read have

---

[7] As is well known, CP extraposition is not an option to circumvent CNPC effects; see, e.g., (i):

(i) *Wem$_1$ hast du [$_{DP}$ ein Buch t$_{CP}$] gelesen [$_{CP}$ das$_2$ Maria t$_2$ geschenkt hat] ?
whom$_{dat}$ have you a book given that Maria$_{nom}$ given has

In contrast, movement of an empty operator without resumption is impossible in these island contexts; see (10-ab).

(10) a. *Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{DP}$ einen Mann t_{CP}]
this is a book where I a man$_{acc}$
getroffen habe [$_{CP}$ der t_1 gelesen hat]]
met have who read has

b. *Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich eingeschlafen bin [$_{CP}$
this is a book where I fallen asleep have
nachdem ich t_1 gelesen habe]]
after I read have

The same goes for movement of an overt relative pronoun; see (11-ab).

(11) a. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich [$_{DP}$ einen Mann t_{CP}]
this is a book that I a man$_{acc}$
getroffen habe [$_{CP}$ der t_1 gelesen hat]]
met have who read has

b. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich eingeschlafen bin [$_{CP}$
this is a book that I fallen asleep have
nachdem ich t_1 gelesen habe]]
after I read have

As for dependencies that reach into PPs, recall that if PP is a complement and there are no other islands, resumption is blocked (see (3-a) vs. (3-d), (3-e)), independently of whether the resumptive pronoun is an R-pronoun *da* ('there') or a regular pronoun like *ihn* ('him'). In contrast, if PP is embedded in an island, e.g., a CNPC island, resumption becomes obligatory (see (12-a) vs. (12-b)). Interestingly, it is only the empty operator strategy that is completely unproblematic in this context; the sentence with the regular pronoun is much degraded in comparison (see (12-b) vs. (12-c)). This latter fact can arguably be viewed as an empirical argument that resumption does indeed involve movement, an assumption that is forced under a local–derivational approach to syntax: However the ban on preposition stranding in German is ultimately derived, it seems clear that we are dealing with a constraint on *movement*.[8]

(12) a. *Das ist ein Vorschlag [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{DP}$ jemanden
this is a proposal where I someone$_{acc}$
t$_{CP}$] kenne [$_{CP}$ der nicht [$_{PP}$ t$_1$ für] gestimmt hat]]
know who not for voted has

b. Das ist ein Vorschlag [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{DP}$ jemanden
this is a proposal where I someone$_{acc}$
t$_{CP}$] kenne [$_{CP}$ der nicht [$_{PP}$ da$_1$-für] gestimmt hat]]
know who not there for voted has

c. ?*Das ist ein Vorschlag [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{DP}$ jemanden
this is a proposal where I someone$_{acc}$
t$_{CP}$] kenne [$_{CP}$ der nicht [$_{PP}$ für ihn$_1$] gestimmt hat]]
know who not for him$_{\mathrm{inan}}$ voted has

The realization of a long-distance dependency by resumption is confined to relativization in German; as shown in (13-ab), the construction is not available with wh-movement or topicalization, even though movement without resumption is also not possible in the island contexts present here (a non-bridge verb context for wh-movement, a CNPC context for topicalization).[9]

(13) a. *[$_{CP}$ Was$_1$ hast du gewusst [$_{CP}$ dass sie t$_1$/es$_1$ kaufen würde]
what$_{acc}$ have you known that she it buy would

b. *[$_{CP}$ Solche Bücher$_1$ habe ich [$_{DP}$ einen Mann t$_{CP}$] getroffen
such books$_{acc}$ have I a man$_{acc}$ met
[$_{CP}$ der t$_1$/sie$_1$ gelesen hat]]
who them read has

And not only that: Resumption in relativization contexts is confined to an empty operator and a complementizer *wo*; in particular, regular overt relative pronouns can never co-occur with resumption; see (14-ab).

[8] This presupposes that *pre*positional phrases are islands that cannot be circumvented by resumptive movement, unlike the other islands discussed so far.

[9] In line with this, the experimental (magnitude estimation-based) study carried out by Alexopoulou & Keller (2003), which only considers wh-questions in German, shows that resumption is never preferred to a resumption-less strategy in German wh-clauses; i.e., resumption does not help to avoid islands in this context.

(14) a. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich gewusst habe [$_{CP}$ dass sie
this is a book that I known have that she
es$_1$ kaufen würde]
it buy would

b. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ Ø] ich [$_{DP}$ einen Mann t$_{CP}$]
this is a book that I a man$_{acc}$
getroffen habe [$_{CP}$ der es$_1$ gelesen hat]]
met have who it read has

In view of the evidence from (13) and (14), one might be tempted to speculate that *wo* in resumptive constructions is not actually a complementizer accompanied by an empty operator, but used here in its independently available function as a locative relative pronoun (see footnote 3 above for arguments against a non-complementizer status of *wo* in contexts without resumption). However, this cannot be right: Ungrammaticality results if there is no argument slot corresponding to the head noun in all the relevant constructions in (5), (6), (9), and (12) where *wo* co-occurs with a resumptive pronoun. Thus, compare the legitimate locative relative pronoun use of *wo* modifying a head noun like *Ort* ('place') in (15-a) with the ill-formed case in (15-b), where a locative interpretation is excluded and there is no argument variable (pronoun or other) that the empty relative operator Op$_1$ could bind ((15-b) = (9-a) without the embedded CP providing the required argument slot). Thus, it can be concluded that *wo* in (5), (6), (9), and (12) is a true complementizer of relative clauses accompanied by an empty operator.

(15) a. Das ist ein Ort [$_{CP}$ wo ich einen Mann getroffen habe]
this is a place where I a man met have

b. *Das ist ein Buch [$_{CP}$ wo ich einen Mann getroffen habe]
this is a book where I a man met have

To sum up so far: There is strong evidence that resumption in German relative constructions with an empty operator and a complementizer *wo* is a *last resort* operation. With some (minimal) idealization of the empirical evidence, and assuming that all finite clauses are barriers for resumption-less relative movement in German (as opposed to other movement types like wh-movement and topicalization), we can state that resumption not only *can* circumvent island effects (except prepositional phrase islands; see footnote 8), as in most other languages where the phenomenon shows up, but actually *must* cross an island to be legitimate in this language.

At this point, a remark on the status of the resumption phenomenon in (5), (6), (9), and (12) is in order. It has become customary to distinguish between two types of resumption: those where the strategy is fully *grammaticalized* (and typically able to circumvent islands) on the one hand, and those where the strategy is *intrusive* (and purely a last resort operation to save constructions in contexts where there is no legitimate way out) on the other; see Sells (1984), Boeckx (2003), and McCloskey (2006). In cases of intrusive resumption, the operation does not seem to belong to the grammar as such, but qualifies as what is essentially a metagrammatical device. A standard case of intrusive resumption shows up under the (optimality-theoretic) analysis that Pesetsky (1997; 1998) develops for the sentence pair in (16) in English: Assuming a high-ranked (non-local) constraint according to which two members of a movement chain must not be separated by an island (such as the wh-island in (16)), the only way to realize the input in this context is by partial spell-out of the trace (which is assumed to have the status of a copy).

(16) a. *[$_{NP}$ Which picture of John]$_1$ were you wondering [$_{CP}$ whether t$_1$ was going to win a prize at the exposition] ?
b. #[$_{NP}$ Which picture of John]$_1$ were you wondering [$_{CP}$ whether it$_1$ was going to win a prize at the exposition] ?

As indicated by #, the use of a resumptive pronoun in (16-b) does not really represent a grammaticalized way of realizing the long-distance dependency. It is worth pointing out that the German relativization by resumption construction in (5), (6), (9), and (12), although confined to last resort contexts, is decidedly not intrusive but rather fully grammaticalized. There are various pieces of evidence to support such a conclusion. First, the examples are generally perceived as completely natural and unmarked, in all varieties of German. In particular, they neither convey the impression of substandard language use in the way that resumption-less examples with *wo* complementizers like (1-a), (1-b), and (i) of footnote 1 (but, as noted, not (3-a)) do, nor do they look like the result of a meta-grammatical perfomance-based mechanism designed to say what one wants to say in the absence of the grammatical means to do so. Second, as shown above, there is a clear difference between the well-formed resumption construction in contexts with an empty operator and a complementizer *wo* and the constructions involving wh-movement, topicalization, and overt relative operators in (13) and (14), which would be completely unexpected if they all involved the same phenomenon (viz., intrusive resumption).[10] Finally, Sells (1984) develops some tests to distinguish grammaticalized from intrusive resumption (also see McCloskey (2006)). A crucial difference arises in contexts with quantificational antecedents: A grammaticalized resumptive pronoun can have all kinds of quantificational antecedents (including *every* and *most*), but an intrusive resumptive pronoun cannot. As shown by the examples in (17), resumptive pronouns in German relativization constructions with an empty operator can take quantificational antecedents without problems.

(17) a. Jedes Buch [$_{CP}$ OP$_1$ [$_C$ wo] man einschläft [$_{CP}$ nachdem
every books where one falls asleep after
man es$_1$ gelesen hat]] ist nicht gut
one it read has is not good

b. Die meisten Bücher [$_{CP}$ OP$_1$ [$_C$ wo] man niemanden finden
the most books where one no-one find
kann [$_{CP}$ der sie$_1$ gelesen hat]] sind auch nicht gut
can who they read has are also not good

Thus, we end up with the conclusion that German has an island-based last resort operation of resumption that is fully grammaticalized.[11]

3. Resumption in Slavic

The goal of this section is to broaden the perspective slightly by adding empirical evidence from two Slavic languages (Czech and Polish). In particular, this evidence will show two things. First, there are languages in which resumption

---

[10] In fact, even though there can be little doubt about their status as ungrammatical expressions, constructions like those in (13) and (14) can sometimes be heard in actual discourses, and may therefore be assumed to be instantiations of truly intrusive resumption of the type that Sells (1984) has in mind for English-type languages.

[11] This, as such, is not unusual given, e.g., Shlonsky's (1992) analysis of the Highest Subject Restriction (i.e., the ban on resumptive pronouns in subject positions that are close to the eventual landing site) in Hebrew and other languages (where resumption is fully grammaticalized) as an instance of last resort. Here and in what follows, I will remain agnostic as to how the Highest Subject Restriction can be derived; see Klein (2013) (and footnote 16 below) for a recent proposal in terms of orders of elementary operations.

may violate syntactic islands but does not actually *have* to do so – i.e., where resumption is not a last resort operation but independently available. Second, the voiding of islands effects by resumption may take place with moved items that are neither necessarily relative operators nor necessarily phonologically empty.

As noted by Toman (1998), resumption can void adjunct islands (see (18-a)) and CNPC islands (see (18-b)) in colloquial Czech (the resumptive is realized as a clitic pronoun here and shows up in a typical clitic position).

(18) a. To je ten chlap [$_{\text{CP}}$ Op$_1$ [$_{\text{C}}$ co] ti říkám že [$_{\text{CP}}$ když
this is the guy what to you I say that if
mu$_1$ nedáme dva lístky] tak budem mít potíže
to him we not give two tickets so we will have troubles

b. To je ta ženská [$_{\text{CP}}$ Op$_1$ [$_{\text{C}}$ co] sem ti dal ten
this is the woman what I AUX to you gave the
časopis [$_{\text{CP}}$ Op$_2$ [co] v něm$_2$ byla její$_1$ fotka]]
magazine what in it was her photograph

However, Toman's (1998) examples given in (19) show that resumption may also take place in local contexts in colloquial Czech (see (19-a), again with a clitic realization of the resumptive pronoun) where movement is another option (see (19-b)).

(19) a. chlap [$_{\text{CP}}$ Op$_1$ [$_{\text{C}}$ co] mu$_1$ nikdo nevěří]
man what to him nobody believes

b. chlap [$_{\text{CP}}$ kerýmu$_1$ [$_{\text{C}}$ Ø] nikdo nevěří t$_1$]
man to whom nobody believes

(19-b) involves an overt relative pronoun. There is also a resumptive-less movement strategy with a null operator, as in German. But again, unlike what generally seems to be the case in German, the strategies are not in (near-) complementary distribution, as they are in German. See (20-b) (resumption) vs. (20-a) (movement of an empty operator).

(20) a. To je ten nůž [$_{\text{CP}}$ Op$_1$ [$_{\text{C}}$ co] Petr našel t$_1$ na stole]
this is the knife what Petr found on table

b. To je ta socha [$_{\text{CP}}$ Op$_1$ [$_{\text{C}}$ co] se jí$_1$ dotk]
this is the statue what REFL her$_{\text{gen}}$ touched

The situation that resumption can circumvent islands but does not always have to do so is familiar from a variety of languages, including Hebrew (Shlonsky (1992)), Arabic (Aoun et al. (2001)), and Irish (McCloskey (2002)), among others (see Boeckx (2003), McCloskey (2006), Klein (2013) for overviews).

Next, recall that resumption in German is possible only with relativization, and only if the relative operator is phonologically non-overt. As shown by Szczegielniak (2005) for Polish, none of these properties is crucial. In Polish, an overt relative pronoun that bears a full set of ϕ-features (person, number, gender) as well as a case feature can license resumption; see (21).[12]

(21) a. Marysia zna chłopców [$_{\text{CP}}$ których$_1$ [$_{\text{C}}$ Ø] wiem że ich$_1$
Mary knows boys whom I.know that them
Ania lubi]
Anne likes

b. Marysia zna chłopców [$_{\text{CP}}$ których$_1$ [$_{\text{C}}$ Ø] wiem że Ania
Mary knows boys whom I.know that Anne
lubi t$_1$]
likes

This resumption strategy with an overt relative pronoun is also available in Polish in island contexts which block regular movement; see (22-a) (resumption) versus (22-b) (Joanna Zaleska, p.c.).

(22) a. To jest książ-k-a [$_{\text{CP}}$ któr-ą$_1$ znam [$_{\text{DP}}$ chłopak-a [$_{\text{CP}}$
this is book-DIM$_{nom}$ which$_{acc}$ I.know boy$_{nom}$

[12] As Szczegielniak (2005) notes, this is subject to an anti-clause mate requirement. One might speculate that this could be related to the question of island violation as a last resort, given that wh-extraction from complement clauses is much more restricted in Polish than in, say, English; cf. Witkoś (1995). However, in the case at hand, there is optionality: Szczegielniak (2005) does not attribute a different grammaticality status to the two sentences in (21-a) (resumption) and (21-b) (no resumption).
Also, it should be noted that an empty operator plus *co* (‘that’) construction, as in the Czech examples just mentioned, is possible in Polish as well; see, e.g., (i) (Joanna Zaleska, p.c.).

(i) To jest takie coś [$_{\text{CP}}$ OP$_1$ co większość ludzi tego nie produkuje]
this is such something that majority$_{nom}$ people$_{gen}$ it$_{gen}$ not produce

Cf. Pesetsky (1997; 1998) and Broihier (1995) for extensive discussion.

któr-y ją$_1$ przeczatał]]]
who$_{nom}$ it$_{acc}$ read

b. *To jest książ-k-a [$_{CP}$ któr-ą$_1$ znam [$_{DP}$ chłopak-a [$_{CP}$
this is book-DIM$_{nom}$ which$_{acc}$ I.know boy$_{nom}$
któr-y t$_1$ przeczatał]]]
who$_{nom}$ read

Wh-movement and topicalization can also involve resumption in Polish; see (23) for an example of the former construction.

(23) [$_{DP}$ Który komputer]$_1$ Marek podejrzewał że Maria wie że
which computer$_{acc}$ Marek suspected that Maria knows that
Jan chce go$_1$ kupić ?
Jan wants it buy

Finally, resumption can circumvent island effects with wh-movement in Polish (cf. (24-a)), in contrast to a movement strategy without resumption (cf. (24-b)).

(24) a. ?[$_{DP}$ Jakiego obrazu]$_1$ zadzwoniłem do Marii [$_{CP}$ po jego$_1$
which painting I.called to Maria after it
namalowaniu] ?
painting

b. *[$_{DP}$ Jakiego obrazu]$_1$ zadzwoniłem do Marii [$_{CP}$ po
which painting I.called to Maria after
namalowaniu t$_1$] ?
painting

In what follows, I will present a buffer-based account of the possibility of island violations with resumption (in German, Czech, Polish, and many other languages), and I will show how it can be that resumption is only permitted if it violates islands in German whereas it is also permitted in other contexts in Czech and Polish.[13]

[13] I will have nothing insightful to say about the inability of resumption in languages like Swedish (see Engdahl (1985)) and Vata (see Koopman (1984)) to avoid island effects. The logic of the approach to be developed below might lead one to postulate that the resumptive pronoun in these kinds of languages is not generated by a designated copy operation, but comes into being in some other way; see also footnote 28 below for some pertinent remarks.

4. A Local–Derivational Approach to Resumptive Movement

4.1. Resumption as Movement

First, I would like to contend that in a local–derivational (phase-based) approach, it is not possible to adopt a base-generation approach to resumption, as it is otherwise standardly assumed.[14] There is simply no way how any syntactic relation could be posited between a base-generated resumptive pronoun and a base-generated displaced item that can be separated from it by an arbitrarily large number of phases.[15] Consequently, if cyclic Agree is excluded, resumption must be derived by movement, and the differences between "standard" movement and resumptive movement with respect to locality constraints (as well as possibly other factors, like weak crossover) must be explained in some different way (see Boeckx (2003) and Klein (2013), and to some extent Koopman (1984), Engdahl (1985), and Aoun et al. (2001)).[16]

It is worth emphasizing that this consequence is independent of the ac-

---

[14] Note that this is so independently of whether there are strong empirical arguments against a base-generation approach; in this context, see, e.g., the arguments for movement based on reconstruction advanced by Salzmann (2006) for (Swiss) German.

[15] In the same way, from a phase-based perspective it is not possible to envisage an "A-bar bound *pro*" strategy for modelling long-distance dependencies, as it has been suggested by Cinque (1990) for cases of displacement that seem to selectively violate certain constraints on movement, and as it is still envisaged as a general possibility in recent work like Erlewine (2014, sect. 3).

[16] Boeckx (2003) assumes that resumption arises as a result of stranding: The resumptive pronoun is a D category that stays in situ, and the operator that has been merged as a complement of D then undergoes movement; also see Grewendorf (2002) for such an approach to resumptive pronouns occurring with left dislocation in German. This implies that movement dependencies with and without resumption have a different source. In contrast, Klein (2013) proposes that there is a single source for both derivations, viz., a ϕP embedding a DP throughout. On this view, whether resumptive movement or standard movement takes place depends on the order of elementary operations: If the next higher phase head (e.g., v) carries out Agree with the ϕP first, the latter becomes transparent for extraction (as suggested by Rackowski & Richards (2005) as a general means of rendering phases transparent for extraction), and DP undergoes (intermediate) movement to the edge of the phase, stranding ϕ, which is realized as a resumptive pronoun. If, on the other hand, the next higher phase head (e.g., v) triggers Move (internal Merge) first, ϕP still intervenes, and so DP cannot be attracted to an intermediate position alone but rather has to pied-pipe the ϕp; this instantiates the strategy of movement without a resumptive pronoun. As Klein (2013) shows, this approach in terms of the order of elementary operations makes it possible to straightforwardly derive the Highest Subject Restriction.

tual local–derivational approach adopted in this monograph (i.e., a phase-based minimalist syntax with step-by-step bottom-up derivations). As a matter of fact, exactly the same consequence holds for other local approaches to syntax, even if they are declarative rather than derivational in nature. In what follows, I will briefly consider HPSG and GPSG analyses, where movement is modelled by SLASH feature percolation (or sharing). Thus, Vaillette (2002) develops an HPSG analysis in which a feature [RESUMP] is assumed for resumptive dependencies that co-exists with the standard feature [SLASH] for regular movement dependencies; since two different features are involved, the differences between resumptive movement and standard movement (e.g., with respect to reflexives of successive-cyclic movement in Irish, i.e., complementizer choice) can be accounted for. In contrast, in the HPSG analysis in Assmann, Heck, Hein, Keine & Müller (2010) that sets out to derive hybrid resumption/movement dependencies in Irish as they are reported in McCloskey (2002), it is suggested that there is in fact only one kind of feature for modelling movement dependencies (whether by resumption or not), viz., [SLASH], and the differences arise with respect to whether there is sharing of the INDEX or LOCAL values, for resumptive pronouns and traces, respectively. These (and other) differences notwithstanding, both HPSG analyses thus basically presuppose that the moved items are of a different type in the two constructions; in this respect, there is convergence with the analyses in Boeckx (2003) and Klein (2013) (see above).

Going back a bit further in time, movement-type analyses of resumption have also been suggested in GPSG. Maling & Zaenen (1982) already distinguish between two types of [SLASH] features: the standard [SLASH] feature on the one hand, and one which is accompanied by a special diacritic ($\underline{\text{[SLASH]}}$) on the other; the two versions are referred to as [GAP-SLASH] and [PRO-SLASH] by Sells (1984). Locality constraints (conceived of as feature co-occurrence restrictions that block the passing on of [SLASH] across certain specified categories) can then be made sensitive to the difference, such that [GAP-SLASH] cannot be instaniated on certain categories (i.e., islands) – and hence fails to be passed on throughout the tree, connecting the base position and the moved item – where [PRO-SLASH] can be instantiated.

Also based on GPSG, Sells (1984) contemplates both an approach in which there is only one [SLASH] feature for both dependencies, and an approach that envisages two different kinds of [SLASH] features that are both passed on locally through syntactic trees. After rejecting the first approach because it cannot accomodate the different kinds of reflexes of standard movement and resumption in a language like Irish, and because it is not able to distinguish between possible and impossible violations of islands, he then goes on to ultimately reject the alternative as well: A first argument is that given an approach of the type pursued by Maling & Zaenen (1982), it is completely unclear why there could be no languages where resumption is systematically more island-sensitive than regular movement; i.e., where [PRO-SLASH] cannot be instaniated on certain categories where [GAP-SLASH] can be. Another argument has to do with morphological realization: If resumptives are introduced by a special type of [SLASH] feature that is different from ordinary [SLASH] features, one would expect them to be realized in some special way morphologically (as with other items that introduce foot features that are passed on in trees), and not like ordinary pronouns.

Sells's (1984) conclusion from all this is that resumption does not in fact involve a purely syntactic dependency after all: The moved item ('filler') introduces a [SLASH] dependency, but this does not go down the tree, ultimately reaching a resumptive pronoun; rather, it 'dries up' somewhere in the middle, licensed by a special rule of [SLASH] feature termination (p. 330), and the creation of the dependency linking the moved item and the resumptive pronoun in its base position will eventually have to be brought about by the semantic component of grammar, which arguably amounts to acknowledging defeat from a syntactic perspective.

I take the problems that Sells raises for movement (or gap feature percolation) analyses to be real. However, it will turn out that the approach to be developed in what follows is not subject to the counterarguments raised by Sells against movement-based approaches to resumption.

4.2. Buffers for Resumption

Given this state of affairs, and given the generalizations about island violation in the preceding sections, there are three questions that need to be addressed: First, how can resumptive movement in German (and many other languages) circumvent islands? Second, why does resumptive movement sometimes *have to* cross an island (as an instance of a *last resort* operation), as in the German data in section 2. above (see Shlonsky (1992), Pesetsky (1998)), and why does resumptive movement sometimes *not* have to cross an island (as an instance of *optional* resumption, as in the Slavic data in section 3.)? And third, how can the locality (i.e., backtracking) problem be solved that arises under a local–derivational approach? This latter problem consists in the observation that, at the point where it encounters an island, a moved item (e.g., an empty operator Op_1 in some of the above sentences) must "know" whether there is a resumptive pronoun in the base position or not.[17]

I would like to suggest that a buffer-based approach makes a unified account of all three problems possible. More specifically, solving the problem of passing on information from the bottom of the dependency by postulating an appropriate symbol on the moved item's buffer in cases of resumptive movement will be shown to simultaneously address the other two problems (why an island can be voided, and why it sometimes has to be voided).

To begin with, suppose that resumption involves a copy mechanism applying to DP. (Recall from chapter 1 that I am assuming throughout that standard, non-resumptive instances of movement leave neither a copy nor a trace). In a language like German, where only null operators participate in resumptive movement, it can be postulated that this first operation of generating a copy is simply excluded for wh-phrases, topics, overt relative operators, etc.; as we have seen, the situation is somewhat different in Slavic. Next, I assume, following Pesetsky (1998), Toman (1998), McCloskey (2006), and many others, that independent principles ensure that the copy is spelled out as a pronominal element, i.e., as the minimal well-formed realization of a DP. Third, and

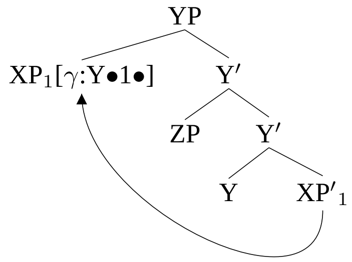
[17] See Lavine (2003) for an early formulation of this kind of problem in phase-based syntax, based on the task to correctly determine the different shapes of displaced items in the target positions with resumption and standard movement in view of the fact that the base position containing the relevant information is normally separated from the displaced item by phase boundaries. It is worth pointing out that Lavine's own solution for this problem cannot be adopted in the present analysis. On the one hand, his approach is not supposed to derive the different behaviour of resumption and resumption-less movement with respect to islands in a local way – rather, he follows Pesetsky (1998) in assuming that whether or not islands are respected can be decided on a strictly non-local basis, by simultaneously checking the properties of the displaced item and the base position, and that the presence of a resumptive pronoun (classified as a reduced copy) simply suffices to circumvent island effects. On the other hand, his solution is eventually based on postulating that there is no movement in resumption constructions at all (a position that I do not think is tenable in a strictly local approach, as noted above), so that different kinds of displaced items in target positions in Slavic resumption and regular movement constructions can be chosen simply on the basis of whether or not a moved item is present in the specifier of the last phase.

crucially, like remnant movement, resumption *does not come for free*. The creation of a copy in the base position (as part of the movement operation) is registered on the moved item, more specifically, on the list that acts as the value of a movement-related feature. For concreteness, I would like to suggest that if a copy of a category γ with index *n* has been made, this information is registered on the buffer that is the value of γ's movement-related feature as $\bullet n \bullet$; so if a copy has been split off from a category γ bearing index 1, both items bear index 1 as a consequence (as is standardly assumed), and $\bullet 1 \bullet$ shows up on γ in addition. Why "$\bullet 1 \bullet$", rather than, say, "1", or something yet different? This reflects the underlying assumption that copying is not innocuous: $\bullet n \bullet$ indicates that the copy should ideally not be present as a separate unit but should be amalgamated – merged – with the moved item again (which, of course, can never happen in the derivation), and this is something that the $\bullet$-free representation would not indicate; but the obvious diacritic to encode this is the symbol $\bullet$ used to bring about Merge operations.[18]

The generation of a copy in resumptive movement constructions is shown in (25); the copy (XP'_1) remains in the base position, and the original item (XP_1) undergoes an intermediate movement step to the phase edge, because of the PIC.

[18] This technically addresses the "deep mystery" raised by the existence of resumptive movement that is mentioned in McCloskey (2006, 113): If resumption is available (and arguably preferable from a functional point of view because it can show up in many contexts where pure movement is blocked *and* because would seem to simplify parsing efforts), why is non-resumptive movement possible in the first place? In the present approach, the answer is twofold. First, as we have just seen, resumption involves an additional, hence costly, operation, viz., the generation of a copy. And second, the information that a copy has been produced is registered on the buffer of the moved item; as will be discussed momentarily in the main text, this leads to temporal defectivity that must be removed in the derivation before a criterial position is reached. Thus, resumption emerges as a marked operation, just like remnant movement (see page 76 above).

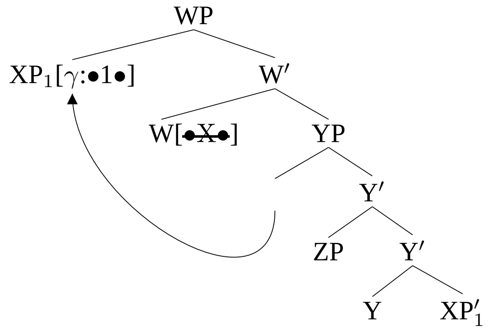
(25) *Initial steps of resumptive movement*



Application of the copy operation (to items where it is permitted, like empty operators in German) is optional throughout (recall, e.g., that an empty operator is available in transparent non-resumptive contexts in German; cf. examples like (1-a), (1-b) (both restricted to substandard varieties), (3-a), (4-a), and (7-a), (7-b) (the latter two again restricted to substandard varieties). However, when it applies, as in (25), its application is registered on the value of the movement-related feature of the moved item. This means that after the first movement step, resumptive movement and standard, non-resumptive movement can be distinguished in a local way in the derivation: Moved items that are accompanied by a resumptive pronoun in the base position are singled out by a •*n*• symbol on their buffer, where *n* is the index shared by the resumptive pronoun and the moved item. Resumptive movement of an item with a symbol •*n*• on its buffer must be unproblematic as such. However, a symbol •*n*• on the value of a movement-related feature of some moved item implies a *temporary defectivity* that a derivation can live with for a while, but that must be remedied before the moved item reaches a criterial position (i.e., a position in which an intrinsic structure-building feature of some lexical head it satisfied, rather than an all-purpose edge feature). The reason is that, just as we have seen in the preceding chapter for a symbol generated by remnant movement, this symbol recording an instance of resumption will invariably trigger a violation of the Williams Cycle in a criterial position.

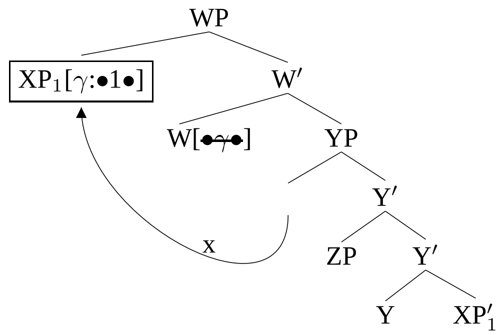
Consequently, a symbol indicating the early generation of a copy (i.e., resumption) must be removed from an item before it reaches a criterial position (a specifier of a $C_{[\bullet\text{rel}\bullet]}$, in the case at hand). (26-a) shows a legitimate case of intermediate resumptive movement where the Williams Cycle is satisfied vacuously because the moved item XP$_1$ is not in a criterial position yet; (26-b) shows how criterial resumptive movement leads to illformedness. Here the moved item XP$_1$ is attracted by a head intrinsically requiring XP$_1$ to become its specifier; the head is looking for XP_1's movement-related feature. ($\cancel{[\bullet X\bullet]}$, $\cancel{[\bullet\gamma\bullet]}$ stand for structure-building features – edge features and inherent, movement-inducing – i.e., criterial – features of a head, respectively – that have been discharged and deleted.)

(26) a. *Intermediate steps of resumptive movement*



($\sqrt{}$f-seq)

b. *Criterial steps of resumptive movement*



(*f-seq)

Thus, the locality (backtracking) problem with resumption is solved: The information that a resumptive pronoun has been split off earlier in the derivation is accessible at later stages because it has been placed on the buffer associated with the moved item.

4.3. Circumventing Islands by Resumption

Next, the task is to show how the presence of such a symbol can make it possible to circumvent what is otherwise an island for movement. Here, the worst case scenario would be that one has to stipulate that a moved item with a symbol •*n*• on the list that acts as the value of its movement-related feature can cross an island whereas a moved item without such a symbol cannot.[19] Still, depending on the properties of the theory of islands that is assumed as background, simpler approaches may be available.

I would like to suggest that given the present assumptions about resumption, island circumvention follows without further ado under the theory of locality constraints on movement developed in Müller (2010; 2011). I briefly outline this approach in the next subsection, and return to resumption after that.

4.3.1. An Approach to Islands

In Müller (2010; 2011), it is argued that island effects can be derived from the PIC: Last-merged specifiers and adjuncts (and, in some cases, complements) are islands because their entering the phase (= phrase, as assumed throughout this monograph) is the *final* operation taking place in a phase that is triggered by the (structure-building or probe) features of the phase head. After a phase head has discharged its final (structure-building or probe) feature, it becomes inactive. This has a potentially fatal consequence given that edge features required to effect intermediate movement steps to phase edges cannot be assigned anymore (more precisely, cannot be generated by defective copying anymore) if the phase head is inactive (given the Edge Feature Condition): It follows that no edge feature can be provided for a moved item in the last-merged XP of a phase head, and subsequent extraction will have to violate the PIC.

More specifically, the approach works as follows. (Most of the background assumptions it requires are already in place; they have been introduced in chapters 1 and 2.) All phrases are phases; all operations are driven by features (structure-building or probe features); and intermediate movement steps require edge features which are generated on phase heads in accordance with the Edge Feature Condition (see (18) of chapter 2, repeated here as (27)).

---

[19] However, as such, such a step would arguably not be radically different in nature from what is standardly assumed, viz., that resumptive pronouns (and, possibly, *pro*s in some cases, see footnote 15) can find an antecedent outside an opaque domain whereas traces (or copies that are not phonologically realized) cannot.

(27) *Edge Feature Condition*:
An edge feature [$\bullet$X$_\gamma\bullet$] can be generated by defective copying of the categorial feature of a head π of a phase only if (a) π is active and (b) this has an effect on outcome.

Activity in the sense of (27) is defined as in (28).

(28) *Activity of a phase head*:
A phase head is active iff it has not yet discharged all its structure-building or probe features.

Furthermore, the PIC (see (1) in chapter 1, which is repeated here in (29)) presupposes a non-recursive definition of edge (such that the specifier of a specifier of a phase head is not accessible from outside the phase; see chapter 3).

(29) *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.

Finally, one additional assumption that is required is that the structure-building features that a head is inherently equipped with (i.e., subcategorization features and features triggering movement operations) are *ordered*; this brings about linking (i.e., correlating the lexically determined argument structure with the hierarchical order of arguments in syntax). Thus, inherent structure-building features come in stacks (first-in/last-out lists, which are not to be confused with the first-in/first-out queues introduced for buffers in chapter 2). Edge Features assigned in accordance with (27) always end up on top of an existing stack, and are discharged before the inherent structure-building feature below is. Consequently, the lowest structure-building feature on a given stack will introduce an XP in the syntax for which an edge feature (normally) cannot be provided anymore.

Let us look at the consequences of this set of assumptions. Three cases need to be taken into account: (i) last-merged specifiers (including, by assumption, adjuncts, and hence also relative CPs); (ii) non-last-merged specifiers and complements; and (iii) last-merged complements (where complements are defined as sisters of lexical items and specifiers are defined as sisters of complex items).

Consider last-merged specifiers first. The relevant changes on the stack of structure-building features of the phase head are shown in (30). Here, [•β•] is the last structure-building feature associated with the phase head π. After discharging this feature and merging with the XP (β) that becomes its specifier, π does not have any structure-building feature left (on the question of probe features, see below). Therefore, π is inactive at this point, and an edge feature [•X$_\pi$•] cannot be generated, given the Edge Feature Condition. However, if an edge feature cannot be inserted on a phase head π, an intermediate movement step of some category α in the last-merged specifier XP (β) to Specπ is blocked, and a PIC violation will arise once the derivation moves beyond the phase headed by π and tries to extract α (given a non-recursive concept of phase edge).

(30) *Last-merged specifiers as islands*:

	π: [•β•]
$\rightarrow$	π: —
$\rightarrow$	π: [•X$_\pi$•]

$\rightsquigarrow$ violates (27)

This derives the illformedness of extraction from subject DPs (and other last-merged specifiers), as shown for wh-movement from in-situ subjects in German in (31).

(31) a. *Was$_1$ haben denn [$_{\text{DP}_3}$ t$_1$ für Bücher] [$_{\text{DP}_2}$ den Fritz]
what have PRT for books$_{nom}$ the Fritz$_{acc}$
beeindruckt ?
impressed

b. *[$_{\text{PP}_1}$ Über wen] hat wohl [$_{\text{DP}_3}$ ein Buch t$_1$] [$_{\text{DP}_2}$ den Fritz]
about whom has PRT a book$_{nom}$ the Fritz$_{acc}$
beeindruckt ?
impressed

Turning to non-last-merged specifiers and complements next, the situation looks as in (32). Here, the phase head still has two subcategorization features on its stack of structure-building features. The feature at the top, viz., [•δ•], is discharged first. (If the XP merged in virtue of [•δ•] is the first item merged with the phase head, it qualifies as a complement, and if there has been another Merge operation triggered by a previous subcategorization feature, XP is a specifier; but the analysis does not distinguish these two cases.) Since, after discharge of [•δ•], there is another structure-building feature left on the phase head, the phase head is still active at this point, and an edge feature can be generated that attracts an item out of the non-last-merged complement or specifier, thereby satisfying the PIC on the next cycle.

(32) *Non-last-merged complements as non-islands*:

$$\begin{array}{r|l|} \cline{2-2} & \pi\text{: } [\bullet\delta\bullet] \succ [\bullet\beta\bullet] \\ \rightarrow & \pi\text{: } [\bullet\beta\bullet] \\ \rightarrow & \pi\text{: } [\bullet \mathrm{X}_{\pi}\bullet] \succ [\bullet\beta\bullet] \\ \rightarrow & \pi\text{: } [\bullet\beta\bullet] \\ \rightarrow & \pi\text{: } — \\ \cline{2-2} \end{array}$$

⇝ violates nothing

An interesting consequence is that this approach actually predicts a transparency for extraction for those subjects where the phase head (i.e., v) has yet another structure-building feature left after merging the subject. This situation obtains with cases of scrambling to an outer specifier of v in languages like German or Czech; and indeed, subjects turn out to lose island status if there is extremely local scrambling to a position in front of it. This *melting effect* induced by local movement to an outer specifier is illustrated by the German examples in (33-a) vs. (33-b).

(33) a. *Was$_1$ haben [$_{\text{DP}}$ t$_1$ für Bücher] [$_{\text{DP}_2}$ den Fritz] beeindruckt ?
what have for books$_{nom}$ the Fritz$_{acc}$ impressed

b. Was$_1$ haben [$_{\text{DP}_2}$ den Fritz] [$_{\text{DP}}$ t$_1$ für Bücher] t$_2$
what have the Fritz$_{acc}$ for books$_{nom}$
beeindruckt ?
impressed

Finally, as for last-merged complements, one might at first sight expect them to be islands in the same way that last-merged specifiers are: In both cases, it looks as though the phase head has become inactive after the Merge operation. This is shown in (34).

(34) *Last-merged complements as islands*?

$$\begin{array}{r|l|} \cline{2-2} & \pi\text{: } [\bullet\beta\bullet] \\ \rightarrow & \pi\text{: } — \\ \rightarrow & \pi\text{: } [\bullet \mathrm{X}_{\pi}\bullet] \\ \cline{2-2} \end{array}$$

⇝ violates (27)

However, it is argued in Müller (2010; 2011) that the island status of a last-merged complement can be voided by a probe feature on the phase head (that shows up there on a separate stack) in a way that the island status of a last-merged specifier can never be. (35) shows how a probe feature ([$*$F$*$]) can keep a phase head that has discharged all its structure-building features active, and thereby permit extraction from a last-merged complement.

(35) *Last-merged complements as non-islands*:

$$\begin{array}{r|ll|}\hline & \pi\text{:} & [\bullet\beta\bullet] \\ & & [*\text{F}*] \\ \rightarrow & \pi\text{:} & [*\text{F}*] \\ \rightarrow & \pi\text{:} & [\bullet\text{X}_{\pi}\bullet] \\ & & [*\text{F}*] \\ \hline \end{array} \qquad \rightsquigarrow \text{ violates nothing}$$

Such a way out is available for complements but not for specifiers because of the interaction of two requirements: First, unlike discharge of a structure-building feature, discharge of a probe feature via Agree requires c-command (so such a feature cannot help a last-merged specifier directly); and second, strict cyclicity precludes carrying out an Agree operation with a complement (or an item contained in a complement) after a specifier has been merged.[20] As a consequence, extraction from a last-merged specifier is still blocked throughout, and extraction from a last-merged complement can only take place when there is an Agree relation between the phase head and the complement. Evidence for this latter prediction comes from the observation that extraction from a complement CP is typically only possible with bridge verbs, and that extraction from an object DP also depends on the choice of embedding verb.

This approach to islands is compatible with all the analyses given above in chapters 1–3.[21] Let me now show how the assumption that resumption leaves

[20] Note that this requires a minimal – and arguably independently motivated – strengthening of the Strict Cycle Condition adopted so far (see (5) from chapter 3), with projections replacing phrases as cyclic domains; see Müller (2011, 183). Incidentally, the revised version of the Strict Cycle Condition must not exclude tucking in, which I have argued to be necessary for multiple movement (see (17-a) in chapter 2). Since the two configurations are sufficiently different (Merge plus Agree in one case, two Merge operations obeying order preservation in the other case), making the relevant distinction would seem to be unproblematic.

[21] Two qualifications are in order, though. First, something extra would ultimately have to be

a symbol •*n*• on the movement-related buffer of the moved item accounts for the absence of island effects against the background of this approach.

4.3.2. Münchhausen Movement

At this point, the analysis is straightforward. With resumptive movement, there is simply no need for an edge feature when an island (i.e., simplifying a bit, a last-merged item) is encountered: A moved XP_1 bears a symbol •1•, and thus *brings its own designated edge feature* that may transport it (but no other category) to the specifier of an otherwise inert, non-active phase, thereby making

---

said to accomodate the analysis of the ban on extraction mentioned in chapter 1 for the purposes of illustrating opacity effects (because that analysis is based on the assumption that Agree with a specifier is possible). And second, the stack-based approach to structure-building features just sketched actually implies that intermediate movement steps precede final movement steps (see Müller (2011) for detailed clarifications as to why this is so). Given the assumptions about movement to phase edges in (17) of chapter 3, this will have no discernible effect on the structures generated standardly in anti-freezing contexts (see (20) of chapter 3), freezing contexts (see (22) of chapter 3), and Müller-Takano generalization contexts (see (24) of chapter 3); and it will not in any way affect the gist of the results given there. However, it will imply that intermediate movement of XP_1 in (22) of chapter 3 takes place before criterial movement of XP_2 (i.e., (17) is then only relevant if other principles do not determine the order of operations); and that XP_1 therefore c-commands XP_2 rather than the other way round in the resulting respresentation. Still, this has no consequences whatsoever because XP_2 will still have to violate the Williams Cycle. In addition, some minimal adjustments might be necessary for cases where there is a criterial initial step (see section 2.3.6. of chapter 3). Finally, one area where additional assumptions might prove necessary concerns one of the two possible accounts of data such as those in (51-d) in chapter 2 that I have addressed in appendix A.1.2.2., viz., the analysis according to which extremely local criterial scrambling of XP may in fact permit extraction from XP *after* XP movement in accordance with the CED. Here, criterial movement needs to precede intermediate movement, as it stands.

While I take none of these consequences to be alarming, it is worth noting that that there is a straightforward minimal mofication of the assumptions made about edge features and intermediate movement steps in Müller (2011) laid out in the main text above, a modification that will leave all the results regarding islands intact and at the same time offer unqualified compatibility with *all* (i.e., including minor) results of chapter 3: Suppose that edge features do not end up on the stack of inherent structure-building features of a phase head, but do in fact go on a separate stack; and the crossing of what would otherwise be an island possible.[22] The symbol $\bullet 1 \bullet$ is discharged as a result of this operation. This is an instance of what has been called *Münchhausen* movement.[23] (36) shows how an XP_1 that undergoes resumptive movement (and hence, has the symbol $\bullet 1 \bullet$ as part of the value of its movement-related feature γ) can be extracted from what would normally be a barrier (a last-merged specifier WP, in the case at hand) to the phase edge of the next-higher category ZP (so as to avoid a PIC violation on the ensuing cycle, outside of ZP) even though Z has already been rendered inactive at the point where movement of XP_1 must take place, and therefore cannot be assigned an edge feature attracting XP_1 anymore: $\bullet 1 \bullet$ on XP_1 functions as an instruction to merge a category with index 1 anew.

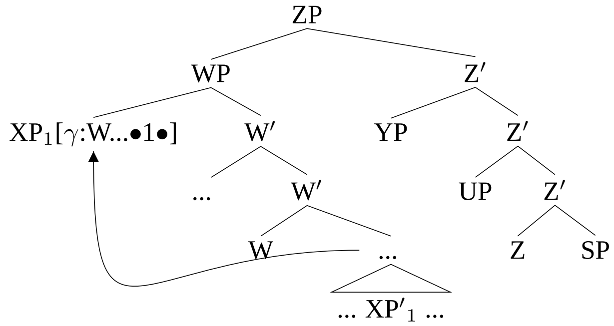
suppose further that presence of an edge feature alone cannot keep a phase head active in the sense of (28). It then follows that edge features do not have to be discharged immediately (but delaying discharge of edge features assigned earlier will not offer a new way of voiding islands for last-merged items), and all orders in cases of multiple movement (whether criterial or intermediate) will fully be regulated by (17) of chapter 3.

[22] This approach is somewhat reminiscent of Assmann's (2012) movement-based analysis of parasitic gaps, which is also framed against the background of the approach to locality constraints developed in Müller (2011). In this approach, an item (i.e., what is sometimes conceived of as an empty operator) can be moved from the base position (i.e., the position containing the "parasitic gap") across an intervening island (adjunct, subject, etc.) because a defective, complementary part of the empty operator (-like element) keeps the phase head active at the crucial stage of the derivation where an edge feature needs to be generated. Still, there is an important difference: In Assmann's analysis, a phase head is exceptionally kept active, so that an edge feature can be assigned; in the present approach, the moved item brings its own edge feature.

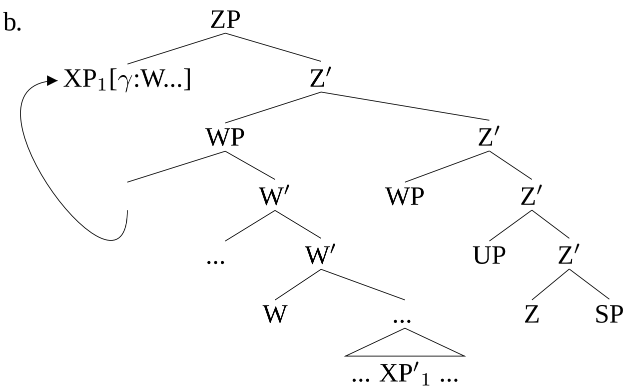
[23] Baron Münchhausen escapes from a swamp (where he is trapped on the back of his horse) by pulling himself up by his hair. The use of the name 'Münchhausen' in syntactic theory for operations that resemble such an escape from a swamp arguably goes back to Sternefeld (1991); also see Fanselow (2003) on head movement by reprojection.

(36) *Circumvention of island effects with resumptive movement*

a.



b.



Suppose that all the well-formed examples involving resumption in German discussed above involve islands that can be reduced to inactive phase heads (cf. relativization in subject island contexts in (5-b), relativization in what otherwise acts as a bridge environment in (6-c), relativization in the presence of matrix negation in (6-d), relativization in non-bridge contexts in (6-e), relativization from a CNPC island in (9-a), (12-b) and (17-b), and relativization from an adjunct island in (9-b) and (17-a)). Then it follows that resumptive movement is possible here whereas non-resumptive movement is not. And indeed, as argued in Müller (2011), in all these contexts there is a phase head that is inactive at the stage of the derivation where the phrase containing the moved item (in its left edge) is merged with it, with one proviso: To derive the island status of CP complements embedded under bridge verbs for the movement type relativization (see (6-c), (7-c), and (8-c)) – but not, say, for topicalization or wh-movement – by invoking a ban on edge feature insertion (due to an inactive phase head) and the PIC, it seems that is has to be assumed that there can be no probe feature for the last-merged CP in just this context that would keep the matrix V phase head active and accessible. At least from a purely technical point of view, this does not pose an insurmountable problem in the approach developed in Müller (2011): It is possible to postulate an incompatibility of a probe feature on V (required for edge feature generation for an item contained in a last-merged CP complement) and a moved item that needs to undergo an intermediate step in the course of relativization.[24]

This analysis has two immediate consequences, both of which turn out to be confirmed by the evidence from resumption in German given above. First, •*n*• can only be used to circumvent *one* island, not multiple islands. And second, •*n*• needs to find an island in order to be deleted from the buffer, which in turn is ultimately required by the Williams Cycle. I address these two consequences in the next two subsections.

4.3.3. Multiple Islands

Given that, like regular structure-building features, •*n*• on a buffer of a moved item is discharged once it has brought about a structure-building (Münchhausen) operation, the prediction arises that from this point onwards, an item undergoing resumptive movement is actually not distinguishable anymore from other kinds of moved items. Consequently, crossing of more than one island by resumptive movement should result in ungrammaticality. Perhaps somewhat surprisingly, this prediction seems to be confirmed for German. Consider the following examples. In (37-a), there is resumptive movement across two islands: First, a CP island is crossed (part of a CNPC context), and second, a subject DP island is crossed. This produces ungrammaticality; there is a striking contrast between (37-a) (with resumptive movement from a subject DP)

[24] The reason is that the latter information is locally available at this point, assuming a solution to the look-ahead problem involved here along the lines of the approaches mentioned in footnote 3 of chapter 1. However, the technical viability of such an account of course still leaves open the more fundamental question *why* relativization behaves differently from wh-movement and topicalization in this respect.

and (37-b) (= (9-a), with resumptive movement from an object DP).[25]

(37) a. *Das ist ein Buch [$_{CP}$ OP$_1$ [$_C$ wo] [$_{DP}$ ein Mann t$_{CP}$] die
this is a book where a man$_{nom}$ the
Maria getroffen hat [$_{CP}$ der es$_1$ gelesen hat]]
Maria$_{acc}$ met has who it read has

b. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich [$_{DP}$ einen Mann t$_{CP}$]
this is a book where I a man$_{acc}$
getroffen habe [$_{CP}$ der es$_1$ gelesen hat]]
met have who it read has

The same goes for cases of resumptive movement combining first an adjunct island and then a CNPC island, as in (38-a); again, there is a (subtle, but clear) contrast with bare resumption across an adjunct island, as in (38-b) (= (9-b)).[26]

(38) a. ?*Das ist ein Buch [$_{CP}$ OP$_1$ [$_C$ wo] ich einen Mann getroffen
this is a book where I a man$_{acc}$ met
habe [$_{CP}$ der eingeschlafen ist [$_{CP}$ nachdem er es$_1$ gelesen
have who fallen asleep has after he it read
hat]]
has

b. Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich eingeschlafen bin [$_{CP}$
this is a book where I fallen asleep have
nachdem ich es$_1$ gelesen habe]]
after I it read have

Still, it might be that there are instances of resumptive movement in the world's languages that do not exhibit this restrictive pattern, but actually permit multiple circumvention of islands. Polish might be a case in point; see (39) (Joanna

---

[25] This account of (37-a) presupposes that extraction of the relative operator takes place before CP extraposition, thereby producing a counter-feeding interaction of operations (i.e., extraposition would feed resumptive movement by making it possible to circumvent the subject DP phrase but comes too late to have this effect).

[26] This account of (37-a) presupposes that extraction of the relative operator takes place before CP extraposition, thereby producing a counter-feeding interaction of operations (i.e., extraposition would feed resumptive movement by making it possible to circumvent the subject DP phrase but comes too late to have this effect).

Zaleska, p.c.).

(39) a. To jest książ-k-a [$_{CP}$ któr-ą$_1$ [$_{DP}$ chłopak [$_{CP}$ któr-y
this is book-DIM$_{nom}$ which$_{acc}$ boy$_{nom}$ who$_{nom}$
ją$_1$ przeczatał]] zadzwoniłdo Ani]
it$_{acc}$ read called Ania$_{gen}$

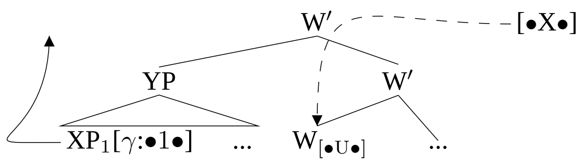
b. To jest książ-k-a [$_{CP}$ któr-ą$_1$ spotkałam [$_{DP}$ chłopaka
this is book-DIM$_{nom}$ which$_{acc}$ I.met boy$_{acc}$
[$_{CP}$ któr-y zasnął [$_{CP}$ kiedy ją$_1$ przeczytał]]]
who$_{nom}$ fell asleep when it$_{acc}$ read

In (39-a), resumptive movement crosses both a CNPC island and a subject island; and in (39-b), there is resumptive movement across both an adjunct island and a CNPC island. To accomodate such conflicting pieces of empirical evidence, it will suffice to postulate that the symbol •*n*• on a buffer of a moved item can also be treated differently from regular structure-building features in languages (perhaps as a marked option), such that it does in fact not (necessarily; see below) disappear after effecting an intermediate movement step.

4.3.4. Required Islands

So far, we have seen that, in the current system, a natural way of expressing the fact that a copy of a moved item with index *n* has been generated is to assume that this is registered by a symbol •*n*• on the moved item's buffer, and that this symbol can be used to bring about an intermediate movement step of the moved item in cases where no edge feature is available (i.e., in island contexts, given that these are reducible to the PIC via an absence of edge features). However, it is clear that such a symbol is not quite a proper edge feature, even if it can fulfill the latter's tasks as a last resort. Thus, a natural conclusion would seem to be that a symbol •*n*• on a moved item cannot normally be used to bring about intermediate movement, in contexts where an edge feature would also be available; it provides a last resort when all else fails. This means that in a situation like the one depicted in (40), where XP$_1$ in the specifier of Y needs to undergo movement to a specifier of the next phase head W (which is active, as signalled by [•U•]), the derivation can only proceed by generation of an edge feature [•X$_W$•], not by discharging the special symbol •1• recording the presence of a resumptive pronoun in XP's base position.

(40)



As a matter of fact, a preference for category-neutral edge features over category-specific (index-sensitive) structure-building symbols on buffers of moved items follows automatically if the Edge Feature Condition in (27) is minimally strengthened in such a way that edge feature generation is viewed as obligatory rather than optional (in contexts where the phase head is still active, and where there is an item that needs to be moved to the next higher phase head, i.e., where this "has an effect on outcome"); see (41).

(41) *Edge Feature Condition* (revised):
An edge feature [•X$_\pi$•] is generated by defective copying of the categorial feature of a head π of a phase iff (a) π is still active and (b) this has an effect on outcome.

(41) implies that an edge feature is generated when it can be generated, and given that an unchecked edge feature would lead to a crash of the derivation in the same way that other structure-building features do, it must be discharged instantaneously. A symbol registering the creation of a resumptive pronoun on the buffer of a moved item, on the other hand, does not immediately lead to a crash of the derivation; it can be tolerated by the derivation in intermediate movement steps. However, the presence of such a symbol •*n*• on a moved item will lead to a violation of the Williams Cycle if it is not discharged before a criterial position is reached: •*n*• can never be part of a proper functional sequence (f-seq), in exactly the same way that the index of some other category that is placed on the buffer of a moved item in α-over-β (remnant movement) constructions can never be part of an f-seq (see chapter 3). This, then, derives the last resort nature of resumption in German: A symbol •*n*• on a buffer must be discharged before a criterial position is reached, and the only way to delete it is to use it in a context where a regular edge feature cannot be generated – i.e., in an island context.

Still, as noted above (see page 138), there are also languages where resumption is not a last resort phenomenon, in the sense that resumption can circumvent islands but does not have to. For languages where resumption is possible in non-island contexts, it can be assumed that the feature can be deleted on buffers if a Williams Cycle violation would otherwise be unavoidable.[27] Taken together, the space for cross-linguistic variation in the realm of symbols registering resumption on buffers then comprises the option of deleting or maintaining the symbol after it has effected an intermediate movement step, and the option of maintaining or deleting the symbol in cases where it has not effected an intermediate step, with the latter choices arguably emerging as the more marked ones (from a conceptual point of view at least, if not based on the actual distribution of the patterns among the world's languages). Since the variation would seem to be empirically well established, it is not clear whether further restrictions could – or should – be established; at any rate, the current approach locates the variation in a low-level domain (manipulation of symbols on syntactic buffers), and not in deeply embedded parameters that yield several further consequences in potentially unrelated domains.

4.3.5. Other Islands

Wh-islands in German pose an interesting challenge for the approach developed above. As shown in (42), resumptives are permitted with long-distance relativization across wh-islands in German.

(42) Das ist ein Buch [$_{CP}$ Op$_1$ [$_{C}$ wo] ich nicht weiß [$_{CP}$ wie man es$_1$
this is a book where I not know how one it
bestellen kann]]
order can

Movement without resumption leads to illformedness, as one would expect; cf. the examples in (43-a) (with an overt relative pronoun) and (43-b) (with an empty operator, a strategy that is available in Standard German but stigmatized, as noted above).

---

[27] This symbol would thus behave similarly to what is assumed for probe features in general in Preminger (2011).

(43) a. *Das ist ein Buch [$_{CP}$ das$_1$ [$_C$ (wo)] ich nicht weiß [$_{CP}$ wie
this is a book where I not know how
man t$_1$ bestellen kann]]
one order can

b. *Das ist ein Buch [$_{CP}$ Op$_1$ [$_C$ wo] ich nicht weiß [$_{CP}$ wie man
this is a book where I not know how one
t$_1$ bestellen kann]]
order can

The reason why an example like (42) is potentially problematic is that unlike all the other island types discussed so far, wh-islands are actually not derived via inactive phase heads in the approach developed in Müller (2011, ch. 5) (but in some other way that involves the "maraudage" of movement-inducing features on interrogative C by the long-distance-moved item, making subsequent creation of the wh-clause by wh-movement impossible; also cf. page 19 above). It is therefore not clear how a symbol like •1• on a buffer of a moved item after resumption (copying) has taken place in situ can help circumvent the wh-island.

Taking a step back, it would seems to be reasonable to distinguish theoretically between strong (invariant) islands and weak (partially transparent, operator-induced) islands; see Cinque (1990) (although it can be noted that Cinque groups factive islands together with wh-islands, whereas the present approach groups them together with other strong islands). Furthermore, as noted by Boeckx (2003, 110-113), there are some languages (Scottish Gaelic, Greek, Romanian) where strong islands (like adjunct islands and CNPC islands) block resumption whereas wh-islands permit resumption, which might also be taken to indicate that the two island types should be treated differently.[28] These considerations notwithstanding, it seems clear that something will eventually have

[28] Of course, the question arises of why it is that some languages have resumption constructions that cannot cross (strong) islands. In footnote 13, I speculated that where resumption cannot cross any islands whatsoever (as in Vata), this might be due to a very different syntactic process underlying the phenomenon. The observation that the present analysis predicts that resumption can (in the unmarked case) cross one island but not more (since there is only one relevant symbol present on the buffer of a moved item that can effect Münchhausen movement) potentially offers another solution to this problem: It might be that where resumption cannot cross strong islands in a language, this could be due to the fact that there is actually more than one island (i.e., more than one item that is last-merged in its phase) present in these constructions.

to be said against the backdrop of the present analysis as to why wh-islands permit resumption in German – a language where, as we have seen, resumptive movement is characterized by contexts where an item has to cross exactly one last-merged XP that would block regular movement.

I will not attempt to advance a full-fledged solution at this point, but there are various possible ways out of this conundrum. Perhaps most obviously, it might simply be the case that the existence of a factive interpretation of the embedded clause, which is also present in embedded wh-clauses in German, is sufficient to derive the island status of the construction, at least with relativization – even verbs like *sagen* ('say') which are not factive when they embed a declarative clause, trigger factivity when they embed a wh-clause.[29] A second alternative that strikes me as viable in principle would be to assume that wh-islands can be given the same kind of explanation as other islands in the theory developed in Müller (2011), i.e., that they could be reduced to the impossibility of edge-feature insertion, and thus to a violation of the PIC, after all. While there are various technical issues that would have to be solved in such an account, it is worth pointing out that CED-based approaches to wh-island effects are by no means unheard of. As a matter of fact, the analysis of wh-islands put forward in Chomsky (1986) is of exactly this type; and so is the optimality-theoretic approach developed in Legendre, Smolensky & Wilson (1998).

5. Conclusion

At this point, let me draw a conclusion. Based on empirical evidence involving resumptive movement in German, I set out to address three problems that arise for a local–derivational approach: First and foremost, there is the backtracking problem: How can the information that a resumptive pronoun occupies the base position be made accessible on the moved item at later stages of the derivation, where it is required? Second, there is the problem that resumption raises with respect to movement theory: How *can* movement in these resumption constructions circumvent an island? Finally, third, there is the last resort problem: Why does movement in these resumption constructions *have to* cross

[29] This would presuppose that CP complements of one and the same verb may either enter the structure as the final operation brought about by inherent features of the phase head (in the case of wh-complements), or as a non-final operation triggered by the phase head (in the case of declarative complements).

an island? I have proposed a specific solution to the first problem in terms of buffers, conceived of as lists that act as the values of movement-related features (like, in the case at hand, [rel]), and that temporarily store minimal pieces of syntactic information (like, in the case at hand, the information that a copy has been generated of the moved item in its in situ position). Given the approach to movement developed in Müller (2011), it has turned out that this is virtually all that needs to be said, provided that the information that a copy has been split off the moved item, and is thus "missing", is encoded by a symbol like •1• (with *n* the index of both the original item and the resumptive copy): Under this assumption, the two remaining problems are then automatically covered as well because (a) •1• *can* be used to circumvent an island (by attracting the item that bears index 1 to the next specifier, which would otherwise not be possible), and (b) •1• *must* be used to circumvent an island because it has to be removed from the buffer before a criterial position is reached (otherwise the Williams Cycle will be violated). Nevertheless, I would like to emphasize that the main goal of the present chapter is to motivate the encoding of resumption on the buffer of a moved item; island circumvention as a last resort could then in principle also be expressed in a less direct way against the background of other approaches to (locality constraints on) movement.[30]

[30] Also note that there are alternative approaches to island circumvention by resumptive movement in the literature; see in particular Boeckx (2003) (who suggests that only those items can leave islands whose movement does not involve an Agree relation with respect to ϕ-features that is brought about by the attracting head, and that this holds for all legitimate cases of resumptive movement across an island) and Klein (2013) (who proposes that moved items that leave behind a resumptive pronoun are small enough to squeeze through barriers, metaphorically speaking). These approaches are potentially interesting as they would seem to make it possible to address the backtracking problem without buffers: Both approaches stipulate that the moved item looks differently in cases with and without resumption (see footnote 16 above). I will not attempt here to discuss these approaches in any detail; but I would like to contend that neither approach is completely unproblematic as it stands, and that the backtracking problem that motivates the present chapter can still safely be assumed to be real. As shown by Salzmann (2006, 292-294), Boeckx's (2003) account essentially amounts to a restatement of the facts; also, the basic premise that ϕ-feature Agree is generally involved in A-bar movement is most likely misguided (see Salzmann (2006, 293), Müller (2011, 99-101)). As for Klein's (2013) approach, it seems clear that the hypothesis that constituent size matters for locality constraints on movement, while original and intrinsically interesing, will ultimately require a complete rethinking of both movement and locality theory, which has not yet been undertaken.

Chapter 5

Conclusion

This monograph has addressed three backtracking problems that arise with movement in local–derivational approaches to syntax in general, and with phase-based approaches in particular. First, there is the issue of backtracking in *improper movement*: To determine whether movement to, say, a specifier of v counts as improper or not, access to information about what kinds of phase edges a moved item has gone through on the way to its ultimate landing site is required. Given phase-based syntactic derivation (more generally, a local–derivational approach), such information is not accessible at the point where it is needed. Second, *remnant movement* gives rise to a backtracking problem: To determine whether movement of some item from which extraction has taken place earlier in the derivation is legitimate or not, access to information is required that specifies whether the extracted item has already reached a criterial position or not. Again, in a phase-based approach such information is not per se available at the point where it is needed. Third, a similar backtracking problem shows up with *resumptive movement*: To determine whether movement of some item can (and, indeed, must) cross an island, it must be known whether a resumptive pronoun occupies the base position or not. Since the relevant island can show up at much later stages of the derivation (in fact, separated from the base posiiton by an arbitrary number of intervening clauses), we are again dealing with a kind of information that is not available at the point where it is needed.

Since in all these cases it is *contextual* information present at an earlier stage of the derivation, rather than *inherent* information associated with the moved item, the obvious conclusion is that a mechanism must exist that makes this type of information accessible on the moved item, and one would expect this mechanism to be as minimal and economical as possible. Against this background, the main substantial hypothesis of the present monograph is that languages accomplish this feat by resorting to *buffers*. Buffers are lists of symbols encoding minimal aspects of earlier stages of the derivation, and technically they are simply viewed as values of movement-related features (like [wh], [top], [Σ], [rel]) of moved items. The single main difference to standard assumptions about feature valuation that is needed in this approach is that valuation is not disjunctive: The value of a movement-related feature on a moved item is a *queue*-like list to which information is successively added, and from which information is also constantly deleted, as the derivation unfolds. This makes it possible to solve the backtracking problems and maintain strictly local analyses of improper movement, remnant movement, and resumptive movement.

In all three cases, a local version of the Williams Cycle requiring the symbol list on a buffer to conform to f-seq has proven crucial: Non-clause bound movement, remnant movement, and resumptive movement all give rise to defects (viz., incompatibilities with f-seq) that languages ideally want to avoid, that they can tolerate temporarily in the derivation, and that must eventually be remedied before a criterial position is reached.

At this point, it may be worth pointing out that the buffer-based approach developed in this monograph potentially has much wider ramifications, far beyond solving backtracking problems with improper movement, remnant movement, and resumptive movement: It offers the beginnings of a theory of *markedness* of movement operations, and, concomitantly, of constructions involving displacement. Thus, first, non-clause bound movement does not come for free; by contaminating the buffer of a moved item with a category symbol that is at variance with f-seq, it gives rise to temporary illformedness, which must be undone before a criterial position is reached if improper movement is to be avoided. Second, remnant movement does not come for free; movement from another item that will also undergo movement itself leaves an index on the latter's buffer that contaminates it and must be removed before a criterial position is reached. And third, resumption does not come for free either: The generation of a copy in the base position leaves a designated structure-building symbol on the moved item's buffer that contaminates it, and that must also be removed before a criterial movement step is carried out. Thus, we end up with clear predictions as to what kinds of displacement count as marked, and what kinds of displacement count as unmarked: On this view, in the unmarked case, movement is clause-bound; movement does not involve incomplete categories (where there is extraction from the moved item); and movement does not leave a resumptive pronoun behind. This makes clear, and potentially interesting, predictions for areas of linguistic research where markedness plays a role: language change, acquisition, and processing.

Similarly, the consequences of the buffer-based approach to movement developed here might also be relevant from a typological perspective. In particular, they would seem to be relevant for the concept of canonicity argued for in Corbett (2005) and much related work, in the sense that they directly provide a theoretical underpinning for the concept of canonicity in displacement constructions: Canonical movement is clause-bound, complete (no extraction from the moved item), and unique (only one position in the dependency is phonologically realized). Deviations from the canonical ideal are expected to occur freely, but to be identifiable as such.

References

Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Connecticut, Storrs, Connecticut.

Abels, Klaus (2008): Towards a Restrictive Theory of (Remnant) Movement, *Linguistic Variation Yearbook* 7, 53–120.

Abels, Klaus (2012a): *Phases. An Essay on Cyclicity in Syntax*. Vol. 543 of *Linguistische Arbeiten*, De Gruyter, Berlin.

Abels, Klaus (2012b): Syntax with a Gilded Edge; Or, Why Edge Features Come in Flavours. Talk, Workshop on Structure Building, Konstanz. Ms., University College London.

Adger, David (2003): *Core Syntax*. Oxford University Press, Oxford, New York.

Agbayani, Brian (1998): Feature Attraction and Category Movement. PhD thesis, UC Irvine.

Alexiadou, Artemis & Elena Anagnostopoulou (2002): Raising Without Infinitives and the Nature of Agreement. In: *Dimensions of Movement: From Remnants to Features*. Benjamins, Amsterdam, pp. 17–30.

Alexiadou, Artemis, Tibor Kiss & Gereon Müller (2012): Local Modelling of Non-Local Dependencies in Syntax: An Introduction. In: A. Alexiadou, T. Kiss & G. Müller, eds., *Local Modelling of Non-Local Dependencies in Syntax*. De Gruyter, Berlin, pp. 1–49.

Alexopoulou, Theodora & Frank Keller (2003): Linguistic Complexity, Locality and Resumption. In: G. Garding & M. Tsujimura, eds., *Proceedings of the 22th WCCFL*. Cascadilla Press, Somerville, Mass., pp. 15–28.

Aoun, Joseph, Lina Choueiri & Norbert Hornstein (2001): Resumption, Movement, and Derivational Economy, *Linguistic Inquiry* 32, 371–403.

Arregi, Karlos & Andrew Nevins (2012): *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Springer, Heidelberg.

Arsenijević, Boban & Wolfram Hinzen (2012): On the Absence of X-Within-X Recursion in Human Grammar, *Linguistic Inquiry* 43, 423–440.

Assmann, Anke (2012): Deriving Parasitic Gaps by Fission and Fusion. In: E. Boone, K. Linke & M. Schulpen, eds., *Proceedings of ConSOLE XIX*. Groningen, pp. 49–75.

Assmann, Anke (2014): Die parasitäre-Lücken-Konstruktion im Deutschen. PhD thesis, Universität Leipzig. Forthcoming.

Assmann, Anke & Fabian Heck (2011): Reply to Fanselow (2001). Ms., Universität Leipzig.

Assmann, Anke & Fabian Heck (2013): Opaque Intervention. In: *Rule Interaction in Grammar*. Vol. 90 of *Linguistische Arbeits Berichte*, Institut für Linguistik, Universität Leipzig, pp. 463–509.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller & Philipp Weisser (2012): Ergatives Move Too Early. Ms., Universität Leipzig (*Linguistische Arbeits Berichte* 90, 2013). To appear in *Syntax*.

Assmann, Anke, Fabian Heck, Johannes Hein, Stefan Keine & Gereon Müller (2010): Does Chain Hybridization in Irish Support Movement-Based Approaches to Long-Distance Dependencies?. In: S. Müller, ed., *The Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford, pp. 27–46.

Asudeh, Ash (2004): Resumption as Resource Managment. PhD thesis, Stanford University.

Bader, Monika (2011): Constraining Operations. A Phase-Based Account of Improper Movement and Anaphoric Binding. PhD thesis, CASTL, Tromsø University., Tromsø.

Bailyn, John Frederick (2001): On Scrambling: A Reply to Bošković and Takahashi, *Linguistic Inquiry* 32, 635–658.

Barbiers, Sjef (2002): Remnant Stranding and the Theory of Movement. In: *Dimensions of Movement*. Benjamins, Amsterdam, pp. 47–67.

Barss, Andrew (1984): Chain Binding. Ms., MIT, Cambridge, Mass. (Talk at GLOW).

Barss, Andrew (1986): Chains and Anaphoric Dependence. Ph.d. thesis, MIT, Cambridge, Mass.

Bayer, Josef & Martin Salzmann (2009): That-Trace Effects and Resumption as the Result of Improper Movement. Ms., Universität Konstanz.

Bayer, Josef, Markus Bader & Michael Meng (2001): Morphological Underspecification Meets Oblique Case: Syntactic and Processing Effects in German, *Lingua* 111, 465–514.

Beck, Sigrid (1997): Wh-Constructions and Transparent Logical Form. PhD thesis, Universität Tübingen.

Béjar, Susana & Milan Řezáč (2009): Cyclic Agree, *Linguistic Inquiry* 40, 35–73.

Belletti, Adriana & Luigi Rizzi (1988): Psych-Verbs and Θ-Theory, *Natural Language and Linguistic Theory* 6, 291–352.

Bhatt, Rajesh & Veneeta Dayal (2007): Rightward Scrambling as Rightward Remnant Movement, *Linguistic Inquiry* 38, 287–301.

Boeckx, Cedric (2003): *Islands and Chains. Resumption as Stranding*. Benjamins, Amsterdam.

Boeckx, Cedric & Kleanthes K. Grohmann (2007): Putting Phases in Perspective, *Syntax* 10, 204–222.

Boeckx, Cedric, Norbert Hornstein & Jairo Nunes (2010): *Control as Movement*. Cambridge University Press, Cambridge.

Börjesson, Kristin & Gereon Müller (2014): Long-Distance Agreement and Locality: A Reprojection Approach. Ms., Universität Leipzig.

Bouma, Gosse, Robert Malouf & Ivan Sag (2001): Satisfying Constraints on Extraction and Adjunction, *Natural Language and Linguistic Theory* 19, 1–65.

Bošković, Željko (1997): *The Syntax of Nonfinite Complementation. An Economy Approach*. MIT Press, Cambridge, Mass.

Bošković, Željko (2002): A-Movement and the EPP, *Syntax* 5, 167–218.

Bošković, Željko (2007): On the Locality and Motivation of Move and Agree: An Even More Minimal Theory, *Linguistic Inquiry* 38, 589–644.

Brandner, Ellen & Claudia Bucheli (2014): Alternative Strategien zur langen W-Extraktion: Evidenz aus dem Alemannischen. Ms., Universität Konstanz.

Branigan, Phil (2013): Cyclicity and the Approach the Probe Principle. Ms., Memorial University of Newfoundland.

Bresnan, Joan (1972): Theory of Complementation in English Syntax. PhD thesis, MIT, Cambridge, Mass.

Bresnan, Joan (1976a): Evidence for a Theory of Unbounded Transformations, *Linguistic Analysis* 2, 353–399.

Bresnan, Joan (1976b): On the Form and Functioning of Transformations, *Linguistic Inquiry* 7, 3–40.

Bresnan, Joan (2001): *Lexical-Functional Syntax*. Blackwell, Oxford.

Broihier, Kevin (1995): Optimality Theoretic Rankings with Tied Constraints: Slavic Relatives, Resumptive Pronouns and Learnability. Ms., MIT, Cambridge, Mass.

Browning, Marguerite (1987): Null Operator Constructions. PhD thesis, MIT, Cambridge, Mass.

Browning, Marguerite (1991): Bounding Conditions on Representations, *Linguistic Inquiry* 22, 541–562.

Büring, Daniel (2005): *Binding Theory*. Cambridge University Press, Cambridge.

Büring, Daniel & Katharina Hartmann (1994): The Dark Side of Wh-Movement, *Linguistische Berichte* 149, 56–74.

Büring, Daniel & Katharina Hartmann (1997): Doing the Right Thing, *The Linguistic Review* 14, 1–42.

Butt, Miriam (1995): *The Structure of Complex Predicates in Urdu*. CSLI Publications, Standford, California.

Butt, Miriam (2008): Revisiting Long-Distance Agreement in Urdu. Ms., Universität Konstanz.

Campana, Mark (1992): A Movement Theory of Ergativity. PhD thesis, McGill University.

Chandra, P. (2005): Hindi-Urdu Long Distance Agreement: Agree, AGREE or Spec-Head?. Ms, University of Maryland, College Park.

Chandra, Pritha (2007): Long-Distance Agreement in Tsez: A Reappraisal. In: A. Conroy, C. Jing, C. Nakao & E. Takahashi, eds., *University of Maryland Working Papers in Linguistics*. Vol. 15, University of Maryland, College Park, pp. 47–72.

Chomsky, Noam (1951): Morphophonemics of Modern Hebrew. Master's thesis, University of Pennsylvania.

Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, Noam (1970): Remarks on Nominalization. In: R. Jacobs & P. Rosenbaum, eds., *Readings in English Transformational Grammar*. Ginn and Company, Waltham, Mass., pp. 184–221.

Chomsky, Noam (1972): Some Empirical Issues in the Theory of Transformational Grammar. In: S. Peters, ed., *Goals of Linguistic Theory*. Prentice-Hall, Englewood Cliffs.

Chomsky, Noam (1973): Conditions on Transformations. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle*. Academic Press, New York, pp. 232–286.

Chomsky, Noam (1975): *The Logical Structure of Linguistic Theory*. Plenum Press, New York.

Chomsky, Noam (1977): On Wh-Movement. In: P. Culicover, T. Wasow & A. Akmajian, eds., *Formal Syntax*. Academic Press, New York, pp. 71–132.

Chomsky, Noam (1981): *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam (1986): *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1993): A Minimalist Program for Syntactic Theory. In: K. Hale & S. J. Keyser, eds., *The View from Building 20*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. In: R. Martin, D. Michaels & J. Uriagereka, eds., *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (2005): Three Factors in Language Design, *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam (2007): Approaching UG from Below. In: U. Sauerland & H.-M. Gärtner, eds., *Interfaces + Recursion = Language?*. Mouton de Gruyter, Berlin, pp. 1–31.

Chomsky, Noam (2008): On Phases. In: R. Freidin, C. Otero & M. L. Zubizarreta, eds., *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, Mass., pp. 133–166.

Chomsky, Noam (2013): Problems of Projection, *Lingua* 130, 33–49.

Chomsky, Noam & Howard Lasnik (1993): Principles and Parameters Theory. In: J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann, eds., *Syntax. Ein internationales Handbuch zeitgenössischer Forschung*. Vol. 1, de Gruyter, Berlin, pp. 506–569.

Cinque, Guglielmo (1990): *Types of A-bar Dependencies*. MIT Press, Cambridge, Mass.

Collins, Chris (1994): Economy of Derivation and the Generalized Proper

Binding Condition, *Linguistic Inquiry* 25, 45–61.

Collins, Chris & Joachim Sabel (2007): An LF Interface Constraint on Remnant Movement. Ms., NYU and Université Catholique de Louvain.

Coon, Jessica, Pedro Mateo Pedro & Omer Preminger (2011): The Role of Case in A-Bar Extraction Asymmetries: Evidence from Mayan. Ms., McGill University.

Corbett, Greville (2005): The Canonical Approach in Typology. In: Z. Frajzyngier, A. Hodges & D. Rood, eds., *Linguistic Diversity and Language Theories*. John Benjamins, Amsterdam, pp. 25–49.

Culicover, Peter & Susanne Winkler (2010): Freezing: A Conspiracy. Ms., Ohio State University and Universität Tübingen.

Dalrymple, Mary (2001): *Lexical Functional Grammar*. Vol. 34 of *Syntax and Semantics*, Academic Press, San Diego.

De Kuthy, Kordula & Walt Detmar Meurers (2001): On Partial Constituent Fronting in German, *Journal of Comparative Germanic Linguistics* 3(3), 143–205.

Deal, Amy Rose (2014): Cyclicity and Connectivity in Nez Perce Relative Clauses. Ms., University of California, Santa Cruz.

den Besten, Hans & Gert Webelhuth (1987): Adjunction and Remnant Topicalization in the Germanic SOV-Languages. Paper presented at the GLOW-Conference, Venice.

den Besten, Hans & Gert Webelhuth (1990): Stranding. In: G. Grewendorf & W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 77–92.

Engdahl, Elisabet (1985): Parasitic Gaps, Resumptive Pronouns, and Subject Extractions, *Linguistics* 23(1), 3–44.

Epstein, Samuel David & T. Daniel Seely (2002): Rule Applications as Cycles in a Level-Free Syntax. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65–89.

Erlewine, Michael (2014): Anti-Locality and Optimality in Kaqchikel Agent Focus. Ms., MIT.

Fanselow, Gisbert (1987): *Konfigurationalität*. Narr, Tübingen.

Fanselow, Gisbert (1990): Scrambling as NP-Movement. In: G. Grewendorf & W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 113–140.

Fanselow, Gisbert (1991): Minimale Syntax. Habilitation thesis, Universität

Passau.

Fanselow, Gisbert (1992a): Deplazierte Argumente. Ms., Universität Passau.

Fanselow, Gisbert (1992b): The Return of the Base Generators. Ms., Universität Passau.

Fanselow, Gisbert (1995): A Minimalist Approach to Free Constituent Order. Ms., Universität Potsdam.

Fanselow, Gisbert (1996): The Proper Interpretation of the Minimal Link Condition. Ms. Universität Potsdam.

Fanselow, Gisbert (2001): Features, Theta-Roles, and Free Constituent Order, *Linguistic Inquiry* 32, 405–436.

Fanselow, Gisbert (2002): Against Remnant VP Movement. In: *Dimensions of Movement*. Benjamins, Amsterdam, pp. 91–125.

Fanselow, Gisbert (2003): Münchhausen-Style Head Movement and the Analysis of Verb-Second. In: A. Mahajan, ed., *Head Movement and Syntactic Theory*. Vol. 3 of *Syntax at Sunset*, UCLA & Universität Potsdam Working Papers in Linguistics, Los Angeles & Potsdam, pp. 40–76.

Fiengo, Robert (1977): On Trace Theory, *Linguistic Inquiry* 8, 35–61.

Fischer, Silke (2006): Matrix Unloaded: Binding in a Local Derivational Approach, *Linguistics* 44, 913–935.

Fitzpatrick, Justin (2002): On Minimalist Approaches to the Locality of Movement, *Linguistic Inquiry* 33, 443–463.

Fox, Danny (1995): Untitled Ms.. Unpublished, MIT.

Fox, Danny (2000): *Economy and Semantic Interpretation*. MIT Press, Cambridge, Mass.

Fox, Danny & David Pesetsky (2005): Cyclic Linearization of Syntactic Structure, *Theoretical Linguistics* 31, 1–45.

Frampton, John & Sam Gutmann (2002): Crash-Proof Syntax. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 90–105.

Frank, Robert, Young-Suk Lee & Owen Rambow (1992): Scrambling as Non-Operator Movement and the Special Status of Subjects. In: S. Barbiers, M. den Dikken & C. Levelt, eds., *Proceedings of the Third Leiden Conference for Junior Linguists*. University of Leiden, Leiden, pp. 135–154.

Freidin, Robert (1978): Cyclicity and the Theory of Grammar, *Linguistic Inquiry* 19, 519–549.

Frey, Werner (1993): *Syntaktische Bedingungen für die Interpretation*. Akademieverlag, Berlin.

Frey, Werner (2000): Über die syntaktische Position der Satztopiks im Deutschen. In: *ZAS Papers in Linguistics*. Vol. 20, pp. 137–172.

Frey, Werner (2004): A Medial Topic Position for German, *Linguistische Berichte* 198, 153–190.

Frey, Werner & Thilo Tappe (1991): Zur Interpretation der X-bar-Theorie und zur Syntax des Mittelfeldes. Grundlagen eines GB-Fragmentes. Ms., Universität Stuttgart.

Gallego, Ángel (2007): Phase Theory and Parametric Variation. PhD thesis, Universitat Autónoma de Barcelona, Barcelona.

Gazdar, Gerald (1981): Unbounded Dependencies and Coordinate Structure, *Linguistic Inquiry* 12, 155–184.

Gazdar, Gerald (1982): Phrase Structure Grammar. In: *The Nature of Syntactic Representation*. Reidel, Dordrecht, pp. 131–186.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag (1985): *Generalized Phrase Structure Grammar*. Blackwell, Oxford.

Georgi, Doreen (2013): Opaque Interaction of Internal Merge and Agree. In: *Rule Interaction in Grammar*. Vol. 90 of *Linguistische Arbeits Berichte*, Institut für Linguistik, Universität Leipzig, pp. 413–461.

Georgi, Doreen (2014): Opaque Interactions of Merge and Agree. PhD thesis, Universität Leipzig.

Graf, Thomas (2013): Local and Transderivational Constraints in Syntax and Semantics. PhD thesis, UCLA.

Grewendorf, Günther (1988): *Aspekte der deutschen Syntax*. Narr.

Grewendorf, Günther (1989): *Ergativity in German*. Foris, Dordrecht.

Grewendorf, Günther (2002): *Minimalistische Syntax*. Francke/UTB, Tübingen and Basel.

Grewendorf, Günther (2003): Improper Remnant Movement, *Gengo Kenkyo: The Journal of the Linguistic Society of Japan* 123, 47–94.

Grewendorf, Günther (2004): Die linke Bewegung von Restkategorien. Ms., Universität Frankfurt/Main.

Grewendorf, Günther (2013): Problems of Remnant Movement. Ms., Universität Frankfurt/Main.

Grewendorf, Günther & Joachim Sabel (1994): Long Scrambling and Incorporation, *Linguistic Inquiry* 25, 263–308.

Grewendorf, Günther & Joachim Sabel (1999): Scrambling in German and

Japanese, *Natural Language and Linguistic Theory* 17, 1–65.
Grimshaw, Jane (1997): Projection, Heads, and Optimality, *Linguistic Inquiry* 28, 373–422.
Grimshaw, Jane (1998): Constraints on Constraints in Optimality Theoretic Syntax. Ms., Rutgers University, New Brunswick, New Jersey.
Grohmann, Kleanthes (1997): German Superiority, *Groninger Arbeiten zur Germanistischen Linguistik* 40, 97–107.
Grohmann, Kleanthes K. (2003): *Prolific Domains: On the Anti-Locality of Movement Dependencies*. John Benjamins, Amsterdam & Philadelphia.
Haider, Hubert (1993): *Deutsche Syntax – generativ*. Narr, Tübingen.
Haider, Hubert (2000a): Branching and Discharge. In: P. Coopmans, M. Everaert & J. Grimshaw, eds., *Lexical Specification and Insertion*. Benjamins, Amsterdam, pp. 135–164.
Haider, Hubert (2000b): Towards a Superior Account of Superiority. In: U. Lutz, G. Müller & A. von Stechow, eds., *Wh-Scope Marking*. Benjamins, Amsterdam, pp. 231–248.
Haider, Hubert (2004): The Superiority Conspiracy: Four Constraints and a Processing Effect. In: A. Stepanov, G. Fanselow & R. Vogel, eds., *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 147–175.
Haider, Hubert (2010): *The Syntax of German*. Cambridge University Press, Cambridge.
Haider, Hubert & Inger Rosengren (1998): *Scrambling*. Vol. 49, Sprache und Pragmatik Arbeitsberichte, Lund.
Hale, John & Géraldine Legendre (2004): Phrase Impenetrability and Wh-Intervention. In: A. Stepanov, G. Fanselow & R. Vogel, eds., *Minimal Links, Remnant Movement, and (Non-) Derivational Grammar*. Mouton de Gruyter, Berlin, pp. 177–203.
Heck, Fabian (2008): *On Pied-Piping. Wh-Movement and Beyond*. Vol. 98 of *Studies in Generative Grammar*, Mouton de Gruyter, Berlin.
Heck, Fabian (2010): Against Direct Recursion in Syntax. Ms., Universität Leipzig.
Heck, Fabian & Anke Assmann (2012): Barss' Generalization and the Strict Cycle at LF. Ms., Universität Leipzig.
Heck, Fabian & Gereon Müller (2000a): Repair-Driven Movement and the Local Optimization of Derivations. Ms., Universität Stuttgart and IDS Mannheim.

Heck, Fabian & Gereon Müller (2000b): Successive Cyclicity, Long-Distance Superiority, and Local Optimization. In: R. Billerey & B. D. Lillehaugen, eds., *Proceedings of WCCFL*. Vol. 19, Cascadilla Press, Somerville, MA, pp. 218–231.

Heck, Fabian & Gereon Müller (2003): Derivational Optimization of Wh-Movement, *Linguistic Analysis* 33, 97–148. (Volume appeared 2007).

Heck, Fabian & Gereon Müller (2010): Lokale Modellierung nicht-lokaler Abhängigkeiten in der Syntax. Ms., Universität Leipzig.

Heck, Fabian & Gereon Müller (2013): On Accelerating and Decelarating Movement: From Minimalist Preference Principles to Harmonic Serialism. In: *Rule Interaction in Grammar*. Vol. 90 of *Linguistische Arbeits Berichte*, Institut für Linguistik, Universität Leipzig, pp. 511–558.

Heim, Irene & Angelika Kratzer (1998): *Semantics in Generative Grammar*. Blackwell, Oxford.

Hornstein, Norbert (2001): *Move. A Minimalist Theory of Construal*. Blackwell, Oxford.

Hornstein, Norbert (2009): *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press, Cambridge.

Huang, Cheng-Teh James (1982): Logical Relations in Chinese and the Theory of Grammar. PhD thesis, MIT, Cambridge, Mass.

Huybregts, Riny (2009): The Minimalist Program: Not a Bad Idea. Ms., Universiteit Utrecht.

Kayne, Richard (1982): *Connectedness and Binary Branching*. Foris, Dordrecht.

Keine, Stefan (2008): Long-Distance Agreement und Zyklisches Agree. Ms., Universität Leipzig.

Keine, Stefan (2014): Improper Agree. Ms., University of Massachusetts, Amherst.

Kim, Shin-Sook & Wolfgang Sternefeld (1997): Superiority vs. Crossover. Ms., Universität Tübingen.

Kiparsky, Paul (1973): Abstractness, Opacity and Global Rules. In: O. Fujimura, ed., *Three Dimensions in Linguistic Theory*. TEC, Tokyo, pp. 57–86.

Kitahara, Hisatsugu (1994): Restricting Ambiguous Rule-Application. In: M. Koizumi & H. Ura, eds., *Formal Approaches to Japanese Linguistics*. Vol. 1, MITWPL, Cambridge, Mass., pp. 179–209.

Kitahara, Hisatsugu (1997): *Elementary Operations and Optimal Derivations*.

MIT Press, Cambridge, Mass.

Klein, Timo (2013): Agree > Move = Resumption. Ms., Universität Leipzig.

Kobele, Greg (2006): Generating Copies. PhD thesis, UCLA, Los Angeles.

Kobele, Greg (2010): Without Remnant Movement, MGs are Context-Free. In: C. Ebert, G. Jäger & J. Michaelis, eds., *Mathematics of Language*. Vol. 10/11, Springer, Heidelberg, pp. 160–173.

Kobele, Greg (2012): Importing Montagovian Dynamics into Minimalism. In: D. Béchet & A. Dikovsky, eds., *Logical Aspects of Computational Linguistics*. Vol. 7351 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 103–118.

Koizumi, Masatoshi (1995): Phrase Structure in Minimalist Syntax. PhD thesis, MIT, Cambridge, Mass.

Koopman, Hilda (1984): *The Syntax of Verbs*. Foris, Dordrecht.

Koopman, Hilda & Anna Szabolcsi (2000): *Verbal Complexes*. MIT Press, Cambridge, Mass.

Koster, Jan (1978): *Locality Principles in Syntax*. Foris, Dordrecht.

Koster, Jan (1987): *Domains and Dynasties*. Foris, Dordrecht.

Lahne, Antje (2009): Where There is Fire There is Smoke. Local Modelling of Successive-Cyclic Movement. PhD thesis, Universität Leipzig.

Lahne, Antje (2012): Specificity-driven Syntactic Derivation. In: M. Uribe-Etxebarria & V. Valmala, eds., *Ways of Structure Building*. Oxford University Press, Oxford, pp. 271–296.

Lasnik, Howard (1999): *Minimalist Analysis*. Blackwell, Oxford.

Lasnik, Howard & Mamoru Saito (1984): On the Nature of Proper Government, *Linguistic Inquiry* 15, 235–289.

Lasnik, Howard & Mamoru Saito (1992): *Move* α. MIT Press, Cambridge, Mass.

Lavine, James (2003): Resumption in Slavic: Phases, Cyclicity, and Case. In: W. Browne, ed., *Formal Approaches to Slavic Linguistics 11*. Bloomington, Indiana, pp. 355–372.

Lechner, Winfried (1998): Two Kinds of Reconstruction, *Studia Linguistica* 52, 276–310.

Legate, Julie Anne (2005): Phases and Cyclic Agreement, *MITWPL* 49, 147–156. Perspectives on Phases.

Legendre, Géraldine, Paul Smolensky & Colin Wilson (1998): When is Less More? Faithfulness and Minimal Links in Wh-Chains. In: P. Barbosa,

D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 249–289.

Lenerz, Jürgen (1977): *Zur Abfolge nominaler Satzglieder im Deutschen*. Narr, Tübingen.

Lightfoot, David (1994): Degree-0 Learnability. In: B. Lust, G. Hermon & J. Kornfilt, eds., *Syntactic Theory and First Language Acquisition: Crosslinguistic Perspectives. Vol. 2: Binding Dependency and Learnability*. Erlbaum, Hillsdale, NJ, pp. 453–472.

Lindauer, Thomas (1995): *Genitivattribute: Eine morphosyntaktische Untersuchung zum deutschen DP/NP-System*. Niemeyer, Tübingen.

Longobardi, Giuseppe (1985): Connectedness and Island Constraints. In: J. Guéron, H.-G. Obenauer & J.-Y. Pollock, eds., *Grammatical Representation*. Foris, Dordrecht, pp. 169–185.

Maling, Joan & Annie Zaenen (1982): A Phrase Structure Account of Scandinavian Extraction Phenomena. In: P. Jacobson & G. Pullum, eds., *The Nature of Syntactic Representation*. Reidel, Dordrecht, pp. 229–282.

May, Robert (1979): Must Comp-To-Comp Movement Be Stipulated?, *Linguistic Inquiry* 10, 719–725.

McCawley, James (1984): Exploitation of the Cyclic Principle as a Research Strategy in Syntax. In: W. de Geest & Y. Putseys, eds., *Sentential Complementation*. Foris, Dordrecht, pp. 165–183.

McCawley, James (1998): *The Syntactic Phenomena of English*. University of Chicago Press, Chicago.

McCloskey, James (1988): Syntactic Theory. In: F. Newmeyer, ed., *Linguistics: The Cambridge Survey. Vol. I. Linguistic Theory: Foundations*. Cambridge University Press, Cambridge, pp. 18–59.

McCloskey, James (1999): On the Right Edge in English, *Syntax* 2, 189–209.

McCloskey, James (2002): Resumptives, Successive Cyclicity, and the Locality of Operations. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 184–226.

McCloskey, James (2006): Resumption. In: M. Everaert & H. v. Riemsdijk, eds., *The Blackwell Companion to Syntax*. Blackwell, Oxford, pp. 94–117.

Müller, Gereon (1993): On Deriving Movement Type Asymmetries. PhD thesis, Universität Tübingen.

Müller, Gereon (1995): *A-bar Syntax*. Mouton/de Gruyter, Berlin.

Müller, Gereon (1996): On Extraposition and Successive Cyclicity. In: U. Lutz

& J. Pafel, eds., *On Extraction and Extraposition in German*. Benjamins, Amsterdam, pp. 213–243.

Müller, Gereon (1998): *Incomplete Category Fronting*. Kluwer, Dordrecht.

Müller, Gereon (1999): Optimality, Markedness, and Word Order in German, *Linguistics* 37, 777–818.

Müller, Gereon (2000a): Das Pronominaladverb als Reparaturphänomen, *Linguistische Berichte* 182, 139–178.

Müller, Gereon (2000b): Zur Ableitung der NP-Adv-V-Stellung im Deutschen. In: J. Bayer & C. Römer, eds., *Von der Philologie zur Grammatiktheorie*. Niemeyer, Tübingen, pp. 117–137.

Müller, Gereon (2002): Two Types of Remnant Movement. In: A. Alexiadou, E. Anagnostopoulou, S. Barbiers & H.-M. Gärtner, eds., *Dimensions of Movement*. John Benjamins, Amsterdam, pp. 209–241.

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. In: K. K. Grohmann, ed., *Explorations of Phase Theory. Features and Arguments*. Mouton de Gruyter, Berlin, pp. 269–308.

Müller, Gereon (2010): On Deriving CED Effects from the PIC, *Linguistic Inquiry* 41, 35–82.

Müller, Gereon (2011): *Constraints on Displacement. A Phase-Based Approach*. Vol. 7 of *Language Faculty and Beyond*, Benjamins, Amsterdam.

Müller, Gereon (2013): On the Order of Syntactic Operations. Ms., Universität Leipzig. (Lecture Notes for a Compact Course, University of Cambridge).

Müller, Gereon & Wolfgang Sternefeld (1993): Improper Movement and Unambiguous Binding, *Linguistic Inquiry* 24, 461–507.

Müller, Stefan (2007): *Head-Driven Phrase Structure Grammar: Eine Einführung*. Stauffenburg, Tübingen.

Murasugi, Kumiko (1992): Crossing and Nested Paths. PhD thesis, MIT, Cambridge, Mass.

Neeleman, Ad & Hans van de Koot (2010): A Local Encoding of Syntactic Dependencies and its Consequences for the Theory of Movement, *Syntax* 13, 331–372.

Nevins, Andrew (2004): Derivations Without the Activity Condition. In: M. McGinnis & N. Richards, eds., *Proceedings of the EPP/Phase Workshop*. MITWPL, Cambridge, Mass., pp. 287–310.

Nordlinger, Rachel (1998): *Constructive Case: Evidence from Australian Languages*. CSLI, Stanford.

Obata, Miki & Samuel David Epstein (2011): Feature-Splitting Internal Merge: Improper Movement, Intervention, and the A/A′ Distinction, *Syntax* 14, 122–147.

Overfelt, Jason (2013): Unbounded Successive-Cyclic Rightward DP Movement. Ms., University of Massachusetts, Amherst.

Perlmutter, David & Scott Soames (1979): *Syntactic Argumentation and the Structure of English*. The University of California Press, Berkeley.

Pesetsky, David (1985): Morphology and Logical Form, *Linguistic Inquiry* 16, 193–246.

Pesetsky, David (1987): Wh-in-Situ: Movement and Unselective Binding. In: E. Reuland & A. ter Meulen, eds., *The Representation of (In)Definiteness*. MIT Press, Cambridge, Mass, pp. 98–129.

Pesetsky, David (1997): Optimality Theory and Syntax: Movement and Pronunciation. In: D. Archangeli & T. Langendoen, eds., *Optimality Theory. An Overview*. Blackwell, Oxford, pp. 134–170.

Pesetsky, David (1998): Some Optimality Principles of Sentence Pronunciation. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 337–383.

Pesetsky, David (2000): *Phrasal Movement and Its Kin*. MIT Press, Cambridge, Mass.

Pesetsky, David (2012): Phrasal Movement and its Discontents: Diseases and Diagnoses. Ms., MIT, Cambridge, Mass. (To appear in Lisa Cheng and Norbert Corver (eds.), *Diagnostics in Syntax*. Oxford University Press.).

Phillips, Colin (2003): Linear Order and Constituency, *Linguistic Inquiry* 34, 37–90.

Plank, Frans (1983): Transparent vs. Functional Encoding of Grammatical Relations: A Parameter for Synctactic Change and Typology, *Linguistische Berichte* 86, 1–13.

Polinsky, Maria (2003): Non-Canonical Agreement is Canonical, *Transactions of the Philological Society* 101.

Polinsky, Maria & Eric Potsdam (2001): Long-Distance Agreement and Topic in Tsez, *Natural Language and Linguistic Theory* 19, 583–646.

Pollard, Carl & Ivan Sag (1992): Anaphors in English and the Scope of Binding Theory, *Linguistic Inquiry* 23, 261–303.

Pollard, Carl J. & Ivan A. Sag (1994): *Head-Driven Phrase Structure Gram-*

mar. University of Chicago Press, Chicago.

Postal, Paul (1974): *On Raising*. MIT Press, Cambridge, Massachusetts.

Preminger, Omer (2009): Breaking Agreements: Distinguishing Agreement and Clitic Doubling by Their Failures, *Linguistic Inquiry* 40, 619–666.

Preminger, Omer (2011): Agreement as a Fallible Operation. PhD thesis, MIT, Cambridge, Mass.

Prince, Alan & Paul Smolensky (2004): *Optimality Theory. Constraint Interaction in Generative Grammar*. Blackwell, Oxford.

Pullum, Geoffrey (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.

Pullum, Geoffrey (1992): The Origins of the Cyclic Principle. In: *Proceedings of the Regional Meeting of the Chicago Linguistic Society*. Vol. 28.2, pp. 209–236.

Pullum, Geoffrey (1997): The Morpholexical Nature of English to-Contraction, *Language* 73, 79–102.

Putnam, Michael, ed. (2009): *Towards a Derivational Syntax*. John Benjamins, Amsterdam.

Rackowski, Andrea & Norvin Richards (2005): Phase Edge and Extraction, *Linguistic Inquiry* 36, 565–599.

Reinhart, Tanya & Eric Reuland (1993): Reflexivity, *Linguistic Inquiry* 24, 657–720.

Reuland, Eric (2001): Primitives of Binding, *Linguistic Inquiry* 32, 439–492.

Richards, Marc (2007): On Feature Inheritance, *Linguistic Inquiry* 38, 563–572.

Richards, Marc (2011): Deriving the Edge: What's in a Phase?, *Syntax* 14, 74–96.

Richards, Norvin (1998): The Principle of Minimal Compliance, *Linguistic Inquiry* 29, 599–629.

Richards, Norvin (2001): *Movement in Language*. Oxford University Press, Oxford.

Richards, Norvin (2012): Lardil "Case Stacking" and the Timing of Case Assignment, *Syntax* 16, 42–76.

Riemsdijk, Henk van & Edwin Williams (1981): NP-Structure, *The Linguistic Review* 1, 171–217.

Riemsdijk, Henk van (1978): *A Case Study in Syntactic Markedness: The Binding Nature of Prepositional Phrases*. Foris, Dordrecht.

Riemsdijk, Henk van (1989): Swiss Relatives. In: *Sentential Complementation*. Foris, Dordrecht, pp. 343–354.

Rizzi, Luigi (1982): *Issues in Italian Syntax*. Foris, Dordrecht.

Rizzi, Luigi (2007): On Some Properties of Criterial Freezing, *CISCL Working Papers on Language and Cognition* 1, 145–158.

Ross, John (1967): Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, Mass.

Sag, Ivan & Thomas Wasow (1999): *Syntactic Theory. A Formal Introduction*. CSLI Publications, Stanford University.

Saito, Mamoru (1985): Some Asymmetries in Japanese and Their Theoretical Implications. PhD thesis, MIT, Cambridge, Mass.

Saito, Mamoru (1992): Long Distance Scrambling in Japanese, *Journal of East Asian Linguistics* 1, 69–118.

Saito, Mamoru (2003): Some Asymmetries in Japanese and Their Theoretical Implications, *Lingua* 113, 481–518.

Salzmann, Martin (2006): Resumptive Prolepsis. A Study in Indirect A′-Dependencies. PhD thesis, Universiteit Leiden.

Salzmann, Martin (2012): On Three Types of Variation in Resumption: Evidence in Favor of Violable and Ranked Constraints. In: H. Broekhuis & R. Vogel, eds., *Linguistic Derivations and Filtering*. Equinox, Sheffield, pp. 76–108.

Sanders, Gerald (1974): Precedence Relations in Language, *Foundations of Language* 11, 361–400.

Sauerland, Uli (1999): Erasability and Interpretation, *Syntax* 3, 161–188.

Sauerland, Uli & Paul Elbourne (2002): Total Reconstruction, PF Movement, and Derivational Order, *Linguistic Inquiry* 33, 283–319.

Sells, Peter (1984): Syntax and Semantics of Resumptive Pronouns. PhD thesis, University of Massachusetts, Amherst.

Shlonsky, Ur (1992): Resumptive Pronouns as a Last Resort, *Linguistic Inquiry* 23, 443–468.

Sportiche, Dominique (1989): Le Mouvement Syntaxique: Contraintes et Paramètres, *Langages* 95, 35–80.

Stabler, Edward (1996): Acquiring and Parsing Languages with Movement. Book ms., UCLA.

Stabler, Edward (1997): Derivational Minimalism. In: C. Retoré, ed., *Logical Aspects of Computational Linguistics*. Springer, Heidelberg, pp. 68–95.

Stabler, Edward (1999): Remnant Movement and Complexity. In: G. Bouma, G.-J. Kruijff, E. Hinrichs & R. Oehrle, eds., *Constraints and Resources in Natural Language Syntax and Semantics*. CSLI Publications, Stanford University, pp. 299–326.

Starke, Michal (2001): Move Dissolves Into Merge: A Theory of Locality. PhD thesis, University of Geneva.

Stechow, Arnim von (1992): Kompositionsprinzipien und grammatische Struktur. In: P. Suchsland, ed., *Biologische und soziale Grundlagen der Sprache*. Niemeyer, Tübingen, pp. 175–248.

Stechow, Arnim von & Wolfgang Sternefeld (1988): *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, Opladen.

Stepanov, Arthur (2007): On the Absence of Long-Distance A-Movement in Russian, *Journal of Slavic Linguistics* 15, 81–108.

Sternefeld, Wolfgang (1991): *Syntaktische Grenzen. Chomskys Barrierentheorie und ihre Weiterentwicklungen*. Westdeutscher Verlag, Opladen.

Sternefeld, Wolfgang (1992): Transformationstypologie und strukturelle Hierarchie. Ms., Universität Tübingen.

Sternefeld, Wolfgang (2006): *Syntax*. Stauffenburg, Tübingen. Two volumes.

Stiebels, Barbara (2006): Agent Focus in Mayan Languages, *Natural Language and Linguistic Theory* 24, 501–570.

Stowell, Tim (1981): Origins of Phrase Structure. PhD thesis, MIT, Cambridge, Mass.

Stowell, Tim (1991): Small Clause Restructuring. In: R. Freidin, ed., *Principles and Parameters in Comparative Grammar*. MIT Press, Cambridge, Mass., pp. 182–218.

Stroik, Thomas (2009): *Locality in Minimalist Syntax*. MIT Press, Cambridge, Mass.

Szczegielniak, Adam (2005): Two Types of Resumptive Pronouns in Polish Relative Clauses, *Linguistic Variation Yearbook* 5, 165–185.

Takahashi, Daiko (1993): Movement of Wh-phrases in Japanese, *Natural Language and Linguistic Theory* 11(4), 655–678.

Takahashi, Daiko (1994): Minimality of Movement. PhD thesis, University of Connecticut.

Takano, Yuji (1994): Unbound Traces and Indeterminacy of Derivation. In: M. Nakamura, ed., *Current Topics in English and Japanese*. Hituzi Syobo, Tokyo, pp. 229–253.

Takano, Yuji (2000): Illict Remnant Movement: An Argument for Feature-Driven Movement, *Linguistic Inquiry* 31, 141–156.

Thiersch, Craig (1985): VP and Scrambling in the German Mittelfeld. Ms., University of Tilburg.

Thoms, Gary & George Walkden (2013): VP Preposing Without Remnant Movement. Ms., University of Manchester.

Toman, Jindřich (1998): A Discussion of Resumptives in Colloquial Czech. In: Ž. Bošković, S. Franks & W. Snyder, eds., *Formal Approaches to Slavic Linguistics*. Ann Arbor, pp. 303–318.

Trommer, Jochen (2006): Plural Insertion is Constructed Plural. In: G. Müller & J. Trommer, eds., *Subanalysis of Argument Encoding in Distributed Morphology*. Vol. 84 of *Linguistische Arbeitsberichte*, Universität Leipzig, pp. 197–228.

Unger, Christina (2010): A Computational Approach to the Syntax of Displacement and the Semantics of Scope. PhD thesis, Universiteit Utrecht, LOT.

Vaillette, Nathan (2002): Irish Gaps and Resumptive Pronouns in HPSG. In: *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford, pp. 284–299.

van Koppen, Marjo (2005): One Probe – Two Goals. Aspects of Agreement in Dutch Dialects. PhD thesis, Leiden University.

Vikner, Sten (2005): Object Shift. In: H. v. Riemsdijk & M. Everaert, eds., *The Blackwell Companion to Syntax*. Blackwell, Oxford, pp. 392–436.

Webelhuth, Gert (1992): *Principles and Parameters of Syntactic Saturation*. Oxford University Press, Oxford.

Wexler, Ken & Peter Culicover (1980): *Formal Principles of Language Acquisition*. MIT Press, Cambridge, Mass.

Williams, Edwin (1974): Rule Ordering in Syntax. PhD thesis, MIT, Cambridge, Mass.

Williams, Edwin (2003): *Representation Theory*. MIT Press, Cambridge, Mass.

Witkoś, Jacek (1995): Wh-Extraction from Clausal Complements in Polish: A Minimality/Locality Account, *Folia Linguistica* 29, 223–264.

Zemskaja, E.A. (1973): *Russkaja Razgovornaja Reč′*. Nauka, Moskva.

Index