A Local Reformulation of the Williams Cycle

Gereon Müller, Universität Leipzig

Abstract

Asymmetries between movement types have standardly been derived by theories of improper movement that preclude certain configurations where different kinds of movement steps are mixed in the course of displacement of a single item. However, closer inspection reveals that none of the existing accounts of improper movement 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 (2001)). In view of this state of affairs, the present paper pursues a fairly modest goal: It implements a specific constraint against improper movement going back to Williams (1974; 2003) – viz, what I will refer to the *Williams Cycle* – in a local way, without a need for backtracking or look-ahead.

1. Introduction: Improper Movement

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 targeted 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 be applied long-distance; and movement types that have landing sites which are high in the clausal structure (e.g., SpecC) typcally can be applied long-distance. Thus, (1-ab) shows that scrambling in German is clause-bound; unlike, 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 is shown for object shift of nonpronominal DPs in Icelandic; see (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). (The same goes for pronominal object shift in the Scandinavian languages.)

(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 by the pair of examples 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 finite CP boundary.

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

(6) a. [$_{DP}$ A review t$_1$] will appear [$_{PP_1}$ of his new book]

b. *John always maintains [$_{CP}$ that [$_{DP}$ a review t$_1$] will appear shortly] whenever he is asked about it [$_{PP_1}$ 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. This role can be played by the Subjacency Condition (if movement must not cross two bounding nodes, and TP qualifies as a bounding node; see Chomsky (1977; 1986)), or by the Phase Impenetrability Condition (PIC) in (7) (given that CP is a phase, and phrasal movement cannot target a head position like C; see Chomsky (2001)).

(7) *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.

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 (8-a) (where raising is followed by wh-movement), or indeed (8-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.

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

2. Existing Analyses

2.1 Principle C

According to the highly influential account developed in May (1979) and adopted in Chomsky (1981), improper movment emerges as an instance of a Principle C effect. The account relies on two central 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 – i.e., its local binder – 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 (9).

(9) *Mary$_1$ seems [$_{CP}$ t_1' that t_1 likes John]

The initial trace t_1 here qualifies as a variable because it is locally A-bar bound by the intermediate trace t_1' in the embedded SpecC position; however, t_1 is then illegitimately also A-bound from the matrix SpecT position (assuming that this latter position qualifies as 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 locally A-bar bound; 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 is argued not to exhibit the typical properties of A-movement – it licenses parasitic gaps, it does not lead to new licensing options for reflexives and reciprocals, it gives rise to weak crossover effects (at least for some speakers, and in certain contexts), and so on; and (b) there are asymmetries between uncontroversial A-bar movement types as well, e.g., topicalization vs. wh-movement in German. The asymmetry between topicalization from a wh-island in German (which typically produces results that are fairly acceptable with argument displacement for most speakers) and wh-movement from a wh-island in German (which leads to strict illformedness irrespective of the status of the moved item as argument or adjunct) that was first noted by Fanselow (1987) is a case in point; see (10-a) vs. (10-b).

(10) a. *Welches Radio$_1$ weißt du nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_1$ t$_2$ repariert]] ?
which radio know you not how one fixes

b. ?Radios$_1$ weiß ich nicht [$_{CP}$ wie$_2$ C [$_{TP}$ man t$_1$ t$_2$ repariert]]
radios know I not how one fixes

The analysis of the contrast in (10) in Müller & Sternefeld (1993) rests on two assumptions. First, different movement types are defined by targetting different landing sites: Thus, it can be assumed for German that scrambling targets SpecV or Specv; raising targets SpecT; and topicalization and wh-movement target different projections in a split left periphery, viz., SpecTop and SpecC, respectively.[1] Second, there is a constraint on uniform chains that makes use of these differences in landing sites. This constraint is called the Principle of Unambiguous Binding (PUB), and it is formulated as in (11).

(11) *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 (in this technical, purely syntactic sense) 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 all 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 (8)) is correctly predicted to be unproblematic because the original trace does not qualify as a variable (as in the original May/Chomsky approach based on Principle C). Furthermore, the analysis can be extended to topicalization/wh-movement asymmetries as in (10), assuming that the embedded SpecC position is uniformly blocked because of the presence of the wh-phrase creating the wh-island: Then, topicalization may use an additional embedded SpecTop escape hatch here that is unavailable for wh-movement, because of the PUB.

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 in Sternefeld (1992), Grewendorf (2003; 2004), Abels (2008), and Neeleman & van de Koot (2010), 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 the *same* kind of domain, or a *higher* domain, in the matrix clause, but not to

[1] As a matter of fact, with multiple specifiers not yet an option, the vP/VP divide not yet established, and IP rather than TP acting as the projection providing derived subject positions, the original assumption for German scrambling in Müller & Sternefeld (1993) was that it targets an adjunction position of VP or IP, and that raising is movement to SpecI. Furthermore, the two separate functional projections in a split left periphery of the clause (generalized in Rizzi (1997) and much subsequent work) providing landing sites for topicalization and wh-movement, respectively, were originally labelled TP (for 'topic phrase') and CP, rather than TopP and CP. None of these issues is important in the present context.

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 > Pred > 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 section 1. 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]

(12) *Generalized Ban on Improper Movement* (GBOIM; Williams (2003)):
Given a Pollock/Cinque-style 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.

As noted above, the Williams Cycle has been adopted in some version in various analyses covering improper movement (and sometimes other phenomena).[4] A particularly explicit version of the Williams Cycle, with far-reaching empirical consequences that go beyond instances of improper movement, has been proposed by Abels (2008); see (13).

(13) *Generalized Prohibition against Improper Movement* (GENPIM; Abels (2008)):
No constituent may undergo movement of type τ if it has been affected by movement of type σ, where $\tau < \sigma$ under UCOOL.

(13) requires a clarification of what it means for a constituent to have been *affected* by movement, and how

---

[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' (arguably 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 paper.

[4] See, for instance, Sternefeld (1992) for a formulation of the Williams Cycle that is very similar to GBOIM in (12). Sternefeld

UCOOL (a *Universal Constraint on Operational Ordering in Language*) encodes an order < among movement types. As for the latter, Abels (2008) assumes an order of movement operations in (14), which is similar to the hierarchies of movement types employed in other versions of the Williams Cycle (including Williams (1974; 2003)).

(14) *The Universal Constraint on Operational Ordering in Language* (UCOOL):
θ < scrambling < A-movement < wh < topicalization

As for the notion of *affectedness* relevant for Abels's (2008) version of the Williams Cycle, it is defined as in (15).

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

In the simplest case, a constituent is affected by a movement operation if it *undergoes* the movement operation (hence the postulation of *reflexive* dominance in (15)). However, Abels argues that the more complex notion of affectedness is required because the GenPIM in (13) is supposed to restrict not only the interaction of movement operations applying to a single item, but also the interaction of movement operations applying to two different items that are base-generated in a dominance relation – more specifically, he takes the Williams Cycle to also restrict combinations of two movement operations in a base structure $[_\beta \ldots \alpha \ldots]$, where either β moves first and α subsequently moves to a higher position (freezing configurations), or α moves out of β first, and β then undergoes movement to a higher position (remnant movement configurations; based on the terminology introduced in Sauerland (1996), the difference is that between a 'surfing' path and a 'diving' path). I will disregard these latter issues in what follows, though, focussing on cases where an item is affected because itself undergoes the movement operation throughout.[5]

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 to a lower kind of domain in the matrix clause. There is a proviso, though. The fatal first movement step to the

calls the constraint *Prinzip der hierarchiekonformen Bewegung* ('Principle of Hierarchy-Compliant Movement'); compare his original (German) definition in (i).

(i) *Prinzip der hierarchiekonformen Bewegung* (Sternefeld (1992)):
Bewegung von α nach β ist nur möglich, wenn der Typ(α) in der Hierarchie [Topik > SpecC > Adj > SpecI > SpecV] tiefer ist als der Typ(β).
(Movement of α to β is possible only when type(α) is lower than type(β) on the hierarchy [topic > SpecC > Adj > SpecI > SpecV].)

[5] I would like to contend that the relevant data considered by Abels (2008) can be derived independently in many cases, without recourse to a theory of improper movement. In addition, at least some of the data instantiating (legitimate and illegitimate) surfing and divings paths advanced by Abels look potentially controversial. (See Müller (1998; 2011) on both these issues.) Given that this may then mean that there is no uniform behaviour of cases where affectedness implies identity of the moved item, where affectedness implies surfing paths, and where affectedness implies diving paths, this will then threaten to undermine a homogeneous approach based a single notion of affectedness of constituents.

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 (2012) 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 (16-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 (16-b) showing that long-distance scrambling is impossible in German is excluded. ((16-b) = (1-b), with the intermediate traces added that are required by the PIC if vP and CP are phases.)

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

b. *dass Karl [$_{\text{vP}}$ das Buch$_1$ glaubt [$_{\text{CP}}$ t_1'' dass [$_{\text{vP}}$ t_1' keiner t_1 liest]]]
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$ reads

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

2.4 The Activity Condition

In Chomsky (2000, 123), Chomsky (2001) and much 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), here repeated again in (17)). 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.

(17) *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). Here is one empirical argument against this constraint raised by Nevins: The Activity Condition is empirically problematic because it 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 (as assumed throughout the present paper). 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 (18).

(18) $[_{\text{CP}}$ 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 (19), 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.

(19) *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)/(17): 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.[6]

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 account and the 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 movement to a criterial position listed in section 1. 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 movmeent, clitic climbing, object shift, and extraposition (if there

[6] Note incidentally that a similar consequence arises in the approach to improper movement developed in Adger (2003, 388); Adger stipulates that "only wh-features are visible in the specifier of CP".

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 derivaitonal 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 obeys Principle C or not); and the position of the moved item in the final landing site that induces the constraint violation.[7]

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, larges pieces of structure must be considered: Under the formulation in (12), this is evident because the restriction explicitly holds for "a movement that *spans a matrix and an embedded clause*" (my emphasis).[8] Under the formulation in (13), the legitimacy of a movement step is checked by inspecting whether the moved item "*has been affected by movement*" (again, my emphasis) of a different type earlier in the derivation, which in the simplest case implies simultaneously taking into account the base position, an intermediate position, and the final landing site, and potentially (given that "to be affected" by movement may apply to many more nodes than "to have undergone" a movement) many more intervening categories. 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 strictly 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).

2.6.3 Promiscuity

The third, and arguably most pressing, 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

[7] 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 non-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).

[8] Non-locality is also an inherent property of the Lec from which the Gboim in (12) is derived as a theorem in Williams (2003); see footnote 3.

up to Chomsky (1986)).[9] 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), and Müller (2011), among many others.[10] 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 (16-a) (legitimate long-distance wh-movement) and (16-b) (illegitimate long-distance scrambling).

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 certainly 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.[11]

---

[9] A version of this problem is also mentioned in Neeleman & van de Koot (2010, 346-347).

[10] At this point, it does not matter whether the iterated intermediate movement steps are then required by the PIC (which would imply that all XPs are phases), or by some other (perhaps additional) locality constraint (in which case it can be maintained that only certain kinds of categories qualify as phases).

[11] Abels (2012) 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 (2012) adduces potentially interesting evidence from Tagalog to support such an approach. However, I will not consider this approach

2.6.4 So?

It seems that many cases where improper movement has been invoked can in fact be derived differently, without recourse to a specific constraint on improper movement. Concerning the phenomena tackled in Müller & Sternefeld (1993) by invoking the PUB, this holds, e.g., for the asymmetry with topicalization from wh-islands vs. wh-movement from wh-islands in German (and other languages), as in (10). (This is analyzed as a maraudage effect under the Intermediate Step Corollary (see below) in Müller (2011), such that a first-moved wh-phrase on its way out of the clause, and to its final landing site in the matrix clause, 'maraudes' the embedded C's features that were needed for a second-moved wh-phrase that is supposed to end up in the embedded wh-clause's SpecC position, whereas a first-moved topic on its way out of the clause does not maraude the embedded C's features because it is equipped with fewer movement-related features to begin with.) It also holds for an asymmetry with extraction from verb-second clauses in German that is derived in terms of improper movement in Haider (1984), von Stechow & Sternefeld (1988, ch. 11.7), Sternefeld (1992), Müller & Sternefeld (1993), and Williams (2003). (This is analyzed as a CED effect derivable from the PIC in Müller (2011).)

However, this conclusion does not hold for all cases. In particular, it does not hold for the core cases in (1-b)–(6-b) (super-raising, long-distance scrambling, etc.): There is no maraudage here (because there is no competing moved item to begin with), and there are no CED islands involved (other items, like wh-phrases or topics, can be extracted into the matrix clause in otherwise identical contexts).

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.[12]

3. 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

---

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 paper (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 (2012) are assumed to accompany, rather than replace, the Williams Cycle that it would seem to strongly resemble. 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 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.

[12] Abels (2008) remarks that "GenPIM [...] cannot be understood directly as a constraint on derivations (unless the standard assumption is given up that successive cyclic movement is launched before the target of movement is merged into the tree)", and states that he makes "no attempt to reformulate GenPIM in derivational terms". Essentially, this is what I set out to do in what follows.

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), 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), no restrictions on improper movement can be imposed in the domain of edge features.[13] Therefore, I assume that edge feature insertion is not free. In the approach to movement developed in Müller (2011), constraints on edge feature insertion play a decisive role in deriving MLC and CED effects from the PIC. In what follows, I will adopt this approach as a general background for a theory of improper movement.[14]

The approach rests on four main assumptions. First, all phrases are phases. Second, all syntactic operations are driven by designated features: There are structure-building features ([•F•]) that trigger internal and external Merge operations (movement and base-concatenation, respectively), and there are probe features ([∗F∗]) that trigger Agree operations. Third, operation-inducing features are ordered; they show up on stacks, with a Last Resort condition demanding that only the topmost feature on a given stack can be discharged (and thereby deleted). Fourth and finally, edge feature insertion that is required for effecting intermediate movement steps (given the second assumption) is restricted by an Edge Feature Condition that is a modification of Chomsky's original proposal. Chomsky (2000; 2001) suggests that the head X of phase XP may be assigned an edge feature *after* the phase XP is otherwise complete, *but only if that has an effect on outcome*. In Müller (2011), it is argued that the italicized parts of the condition should be changed, such that the head X of phase XP may be assigned an edge feature *before* the phase XP is otherwise complete (i.e., only as long as the phase head is still active, and has not yet become completely inert), *but only if there is no other way to produce a balanced phase* (this last requirement can be viewed as a way to encode the 'effect on outcome' condition by inspecting movement-inducing features of the numeration and comparing them with the potentially available material matching these features in the current derivation, in a way that does not require actual look-ahead; see Heck & Müller (2000)). The resulting version of the Edge Feature Condition is given in (20).

(20) *Edge Feature Condition*:
An edge feature [•X•] can be assigned to the head γ of a phase, ending up on the top of γ's stack of structure-building features, only if (a) and (b) hold:
a. γ has not yet discharged all its structure-building or probe features.
b. The phase headed by γ is otherwise not balanced.

As shown in Müller (2011), given these assumptions, MLC and CED effects follow from the PIC, and there is no need to invoke specific constraints to derive them anymore. In particular, MLC effects follow because the higher one of two items competing for movement to the domain of a movement-inducing head (i.e., the

[13] Unless, that is, one assumes edge features to be flavoured in the sense of Abels (2012); recall the discussion in footnote 11.
[14] To a large extent, this is solely 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. In a few areas, however, the specicic choice of frame of reference does matter; I will discuss these issues when they arise.

item that is merged later) ensures phase balance without edge feature insertion, which is therefore blocked; and subsequent movement of the lower item violates the PIC. On the other hand, CED effects follow because edge feature insertion cannot take place for an item that is included in a *last-merged specifier* of a phase head, with the phase head qualifying as inert at this point; subsequent movement of such an item included in a last-merged specifier then also violates the PIC (given non-recursive phases edges); in contrast, non-last-merged specifiers and non-last-merged complements are predicted to be fully transparent (where specifiers are non-first-merged items, and complements are first-merged items), and last-merged complements are predicted to be sometimes transparent (when additional probe features show up on the phase heads, an option that does not arise with specifiers for systematic reasons related to cyclicity and the c-command requirement for Agree).

A side effect of this approach is that intermediate movement steps must take place before regular specifiers are merged; this is referred to in Müller (2011) as the Intermediate Step Corollary; see (21).

(21) *Intermediate Step Corollary*:
Intermediate movement steps to specifiers of X (as required by the PIC) must take place before a final specifier is merged in XP.

The Intermediate Step Corollary is argued to have interesting consequences for a residue of MLC effects that cannot be subsumed under the PIC in Müller (2011) (cf. the selective nature of wh-islands for topicalization vs. wh-movement; see (10) above). For the approach to improper movement to be developed in what follows it will not be directly relevant; still, it will be worth keeping in mind when one looks at the order of operations in the derivations that will be given below.

To sum up so far, edge features play a central role in the approach to movement developed in Müller (2011), and to the extent that the restrictions on their insertion make it possible to derive a number of different locality effects in a unified way, they can arguably be viewed as well motivated empirically. However, from a minimalist perspective, there is a very 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.[15]

A possible solution to this problem is advanced in Lahne (2009). Lahne suggests that edge features do not exist as such; rather, there is just an edge property (or a structure-building instruction: [• •]) that can be assigned to *some feature(s)* of a phase head, thereby creating an edge feature. Discharge (and deletion) of such derivative edge features then accounts for a generalization concerning the morphological form of intermediate reflexes of successive-cyclic movement; by deleting derivative edge features, contexts for (late) morphological insertion are impoverished, thereby effecting a "retreat to the general case" (see Halle & Marantz (1993)), which then implies that reflexes of successive-cyclic movement will always qualify as morphological default exponents.[16] However, there is a problem with this approach: The newly derived edge feature may need to attract and check a moved item on which a matching feature is not found. For instance, an edge feature

[15] Also, given that edge feature insertion is heavily constrained by the Edge Feature Condition, such features cannot plausibly be assumed to all be present in numerations since the latter would have to anticipate the exact number of edge features needed for convergence – not one too many, not one too few – if one wants to maintain a model approaching crash-proofness in which it is not the case that the vast majority of derivations will eventually fail.
[16] More specifically, the generalization looks as in (i) (see Lahne (2009, 60)).

[•V•] must be assumed to be able to attract a DP with a conflicting categorial feature in transitive irrealis contexts in Chamorro (where the morphological reflex of successive-cyclic movement necessitates deletion of [V]); see Lahne (2009, 70). Similarly, an edge feature [•voice:–ag•] may need to attract a wh-item with a conflicting feature value (viz., [voice:+ag]) in nominalization constructions in Chamorro (see Lahne (2009, 80)); and so on.

I conclude from all this that the idea to construct edge features from existing material on phase heads (rather than insert them out of nowhere) is on the right track because it helps to avoid the problem with the Inclusiveness Condition.[17] However, the newly formed edge features cannot be assumed to directly correspond to features that show up on phase heads: A phase head v, for instance, must be able to attract a DP with which it does not share a single feature (categorial or other).[18]

4. A Reformulation of the Williams Cycle

4.1 Assumptions

I would like to suggest that edge features are ***defective copies*** of categorial features of phase heads; 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. Such information is maintained for a while in derivations, but is deleted as soon as information of the same type is encountered. Finally, when an eventual (criterial) target position is reached, the *functional sequence* (f-seq) of heads (see Starke (2001)) must be respected on such lists.

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 ([• •]). The copy mechanism is given in (22-a) (with γ a variable over category labels), and it is illustrated for some phase heads in (22-b).

(22) a. $[\gamma] \rightarrow [\gamma]$, $[\bullet X_{\gamma} \bullet]$
b. (i) $[V] \rightarrow [V]\ [\bullet X_{V} \bullet]$

(i) *Lahne's Generalization*:
Morphological reflexes of successive-cyclic movement involve elsewhere markers: If a language has different exponents in contexts with and without movement, then the marker that shows up in the movement context is less specific than the marker that shows up in the non-movement context.

This generalization is derived as follows in Lahne's (2009) approach: First, intermediate movement steps require an edge feature. Second, edge features are not inserted as such; rather, an *edge feature property* [• •] is assigned (by edge property insertion rules) to a feature (or features) of the phase head. Third, the features that are thus affected trigger structure-building (i.e., intermediate movement) in syntax and undergo deletion. Fourth, the features that are deleted are not available for post-syntactic morphological realization anymore; i.e., there is *impoverishment* in the syntax (also see Keine (2010), where this option is systematically pursued). Finally, as a consequence of syntactic impoverishment, a less specific, default exponent is chosen by spell-out.

[17] Depending on the exact interpretation of what it means to assign a "property" to a given feature, one might argue that an Inclusiveness Condition violation is not fully avoided either in this approach. However, even on such a strict view, there would still be an obvious minimization of a potential violation of the constraint.

[18] Of course, additional assumptions are then required to capture Lahne's Generalization (assuming that it can indeed be upheld beyond the relatively small sample of languages that motivates it).

(ii) $[\text{v}] \rightarrow [\text{v}], [\bullet \text{X}_{\text{v}} \bullet]$

(iii) $[\text{T}] \rightarrow [\text{T}], [\bullet \text{X}_{\text{T}} \bullet]$

(iv) $[\text{C}] \rightarrow [\text{C}], [\bullet \text{X}_{\text{C}} \bullet]$

As shown in (22), 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 anymore to merge an item with the exact same categorial feature as that of the phase head). 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.[19] As before (see the last section), as many edge features can be generated (by copying the categorial feature of the phase head) as are needed to effect intermediate movement steps of items, in accordance with the Edge Feature Condition. The revised version of the Edge Feature Condition is given in (23) (compare (20)).

(23) *Edge Feature Condition (EFC, revised)*:
An edge feature $[\bullet \text{X}_\gamma \bullet]$ can be generated by defective copying of the categorial feature of a head γ of a phase, and can be assigned to the top of γ's stack of structure-building features, only if (a) and (b) hold:

a. γ has not yet discharged all its structure-building or probe features.

b. The phase headed by γ is otherwise not balanced.

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. 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 (24), with □ representing an initially empty list.[20]

(24) a. [Σ:□] (scrambling)

b. [wh:□] (wh-movement)

c. [top:□] (topicaliztion)

d. [rel:□] (relativization)

e. [EPP:□] (raising to SpecT)

---

[19] One can think of this in terms of a vaccination analogy: A virus is rendered inactive by destroying its contents, but the protein shell uniquely identifying the virus is preserved. – Note incidentally that it would in principle be possible to assume that only the defective copy is in fact retained in (22); or, in fact, that edge feature generation involves only the assignment of the edge property and the demotion of the original categorial content, with no copying involved in the first place. To make this work, one would have to assume that the categorial information of the phase head is either irrelevant in the remainder of the derivation (see Collins (2002), Heck (2010)), or has been projected before edge feature generation applies.

[20] Whereas [wh:□] and [rel:□] features for wh-movement and relativization, respectively, can be assumed to obligatorily show up on an item, [Σ:□] and [top:□] features for scrambling and topicalization must be optional on items. Similarly, the movement-inducing features [●wh●] and [●rel●] show up obligatorily on interrogative and relative C, respectively (in languages of the English or German type), whereas [●Σ●] is optional on v/V in scrambling languages, and [●top●] 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 [(●)Σ(●)] 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 yet something else.

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. This is shown in (25) for an abstract derivation in which a wh-phrase undergoes successive-cyclic movement via all intervening phases edges to the embedded SpecC position (which is not its ultimate target position because, by assumption, C is declarative here and lacks [•wh•]).

(25) a. Merge(V:[$\bullet X_{\text{V}}\bullet$], DP:[wh:$\square$]) $\rightarrow$ V DP:[wh:$\boxed{\text{V}}$]
b. Merge(v:[$\bullet X_{\text{v}}\bullet$], DP:[wh:$\boxed{\text{V}}$]) $\rightarrow$ v DP:[wh:$\boxed{\text{vV}}$]
c. Merge(T:[$\bullet X_{\text{T}}\bullet$], DP:[wh:$\boxed{\text{vV}}$]) $\rightarrow$ T DP:[wh:$\boxed{\text{TvV}}$]
d. Merge(C:[$\bullet X_{\text{C}}\bullet$], 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.[21] Such a deletion of parts of the derivational record in feature value lists is depicted in (26), which continues the derivation in (25) into the matrix clause (the last step that merges C and DP is repeated here); 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.

(26) a. Merge(C:[$\bullet X_{\text{C}}\bullet$], DP:[wh:$\boxed{\text{TvV}}$]) $\rightarrow$ C DP:[wh:$\boxed{\text{CTvV}}$]
b. Merge(V:[$\bullet X_{\text{V}}\bullet$], DP:[wh:$\boxed{\text{CTvV}}$]) $\rightarrow$ V DP:[wh:$\boxed{\text{VCTv}\cancel{\text{V}}}$]
c. Merge(v:[$\bullet X_{\text{v}}\bullet$], 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 (27).

(27) a. *Valuation*:
Merge(Y:[$\bullet X_{\gamma}\bullet$], Z:[F:$\boxed{(\delta_1)\ ...\ (\delta_n)}$]) $\rightarrow$ Y Z:[F:$\boxed{\gamma(\delta_1)\ ...\ (\delta_n)}$], where F is a movement-related feature and $\boxed{(\delta_1)\ ...\ (\delta_n)}$ is a (possibly empty) list of pieces of categorial information.
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, and $\delta_1\ ...\ \delta_n$ and γ are pieces of categorial information.[22]

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., [$\bullet \text{wh}_{\text{C}}\bullet$]; this ensures deletion of the earlier $\boxed{\text{C}}$ information in the case at hand; see (28) (again, the last relevant step of the earlier derivation is repeated here for convenience).

(28) a. Merge(v:[$\bullet X_{\text{v}}\bullet$], DP:[wh:$\boxed{\text{VCTv}\cancel{\text{V}}}$]) $\rightarrow$ v DP:[wh:$\boxed{\text{vVCT}\cancel{\text{vV}}}$]

---

[21] 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, or by parsing considerations (Josef Bayer, p.c.). Also cf. Heck (2010), where it is argued that direct recursion is not available in syntax. The conceptual underpinning for this latter view (a kind of ambiguity avoidance) and its formal implementation look very similar to the present restriction, but I will refrain from trying to suggest a unified approach here.

[22] This formulation 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, so as to maximize perspicuity.

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

c. Merge(C:[$\bullet$wh$_{\text{C}}\bullet$], DP:[wh:$\boxed{\text{TvVC}\cancel{\text{TV}}}$]) $\rightarrow$ C DP:[wh:$\boxed{\text{CTvV}\cancel{\text{CTvV}}}$]

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); cf. (29).

(29) a. Merge(C:[$\bullet$wh$_{\text{C}}\bullet$], DP:[wh:$\boxed{\text{TvVC}\cancel{\text{TV}}}$]) $\rightarrow$ C DP:[wh:$\boxed{\text{CTvV}\cancel{\text{CTvV}}}$]

b. Check C DP:[wh:$\boxed{\text{CTvV}\cancel{\text{CTvV}}}$] $\rightarrow \surd$

This is the new version of the Williams Cycle, which can be formulated as in (30).

(30) *Williams Cycle*:

Categorial 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).

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

4.2 Clause-Bound Wh-Movement

Consider first a simple case of clause-bound wh-movement, as in (31) in English.

(31) (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 (32). On the VP level, the wh-object *what* first needs to undergo movement to SpecV because of the PIC (cf. (7)), given that first-merged items (i.e., complements) are not yet part of the edge domain of a phase. An edge feature [$\bullet$X$_{\text{V}}\bullet$] can be generated by defective copying here in accordance with the Edge Feature Condition in (23), 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: DP$_{\text{wh:}\boxed{\text{V}}}$; see (32-b).[23] Next, on the vP level, a new edge feature [$\bullet$X$_{\text{v}}\bullet$] is generated and discharged by movement of DP$_{\text{wh:}\boxed{\text{V}}}$ to Specv, which further values the wh-feature on the moved item: DP$_{\text{wh:}\boxed{\text{vV}}}$. Given the Intermediate Step Corollary, such movement must take place before the external argument DP is merged with v; the latter operation therefore creates an outer specifier; see (32-de). The pattern is repeated on the TP level, where [$\bullet$X$_{\text{T}}\bullet$] is first generated and then discharged by movement of DP to SpecT, thereby valuing its wh-feature with the newly encountered syntactic context; see DP$_{\text{wh:}\boxed{\text{TvV}}}$ in (32-g). (Again, the Intermediate Step Corollary ensures that EPP-driven movement of the subject DP comes later in the derivation, and creates a higher specifier position of T; see (32-h)). Finally, on the CP level, the inherent movement-inducing feature [$\bullet$wh$_{\text{C}}\bullet$] on the interrogative C head triggers movement of DP$_{\text{wh:}\boxed{\text{TvV}}}$ to SpecC, valuing the wh-feature by adding the categorial information and thereby producing DP$_{\text{wh:}\boxed{\text{CTvV}}}$; cf. (32-j). Since an inherent structure-building feature has been checked at this point, the Williams Cycle in (30) 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.

[23] See Müller (2011) for discussion. 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.

(32) a. $[_{V'}\ [_{V}\ \text{said}\ ]_{[\bullet X_V \bullet]}\ \text{what}_{[\text{wh}:\square]}\ ]$

b. $[_{VP}\ \text{what}_{[\text{wh}:\boxed{V}]}\ [_{V'}\ [_{V}\ \text{said}\ ]]]$

c. $[_{v'}\ v_{[\bullet X_v \bullet] \succ [\bullet D \bullet]}\ [_{VP}\ \text{what}_{[\text{wh}:\boxed{V}]}\ [_{V'}\ [_{V}\ \text{said}\ ]\ ]]]$

d. $[_{v'}\ \text{what}_{[\text{wh}:\boxed{vV}]}\ [_{v'}\ v_{[\bullet D\bullet]}\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]$

e. $[_{vP}\ \text{she}\ [_{v'}\ \text{what}_{[\text{wh}:\boxed{vV}]}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]$

f. $[_{T'}\ T_{[\bullet X_T\bullet]\succ[\bullet\text{epp}\bullet]}\ [_{vP}\ \text{she}\ [_{v'}\ \text{what}_{[\text{wh}:\boxed{vV}]}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]$

g. $[_{T'}\ \text{what}_{[\text{wh}:\boxed{TvV}]}\ T_{[\bullet\text{epp}\bullet]}\ [_{vP}\ \text{she}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]$

h. $[_{TP}\ \text{she}\ [_{T'}\ \text{what}_{[\text{wh}:\boxed{TvV}]}\ T\ [_{vP}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]]$

i. $[_{C'}\ C_{[\bullet\text{wh}\bullet]}\ [_{TP}\ \text{she}\ [_{T'}\ \text{what}_{[\text{wh}:\boxed{TvV}]}\ T\ [_{vP}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]]]$

j. $[_{CP}\ \text{what}_{[\text{wh}:\boxed{CTvV}]}\ [_{C'}\ C\ [_{TP}\ \text{she}\ [_{T'}\ T\ [_{vP}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]]]]$ ($\surd$f-seq)

4.3 Long-Distance Wh-Movement

Consider next the case of long-distance wh-movement from a declarative CP, as in the English example in (33).

(33) 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 (31); see (34). As before, extremely local movement from the complement position of V to a specifier of V is required by the PIC (which values the wh-feature on the DP with the symbol V), and the intermediate movement steps on the vP and TP levels precede external Merge and internal Merge of the subject DP, respectively, because of the Intermediate Step Corollary. The only relevant difference to (31) 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 (23), by defective copying of the categorial feature of the phase head. As a consequence, valuation of the wh-feature in (34-j) does not activate the Williams Cycle: Movement has not yet targetted a criterial position.

(34) *Embedded clause*:

a. $[_{V'}\ [_{V}\ \text{said}\ ]_{[\bullet X_V \bullet]}\ \text{what}_{[\text{wh}:\square]}\ ]$

b. $[_{VP}\ \text{what}_{[\text{wh}:\boxed{V}]}\ [_{V'}\ [_{V}\ \text{said}\ ]\ ]]$

c. $[_{v'}\ v_{[\bullet X_v \bullet] \succ [\bullet D \bullet]}\ [_{VP}\ \text{what}_{[\text{wh}:\boxed{V}]}\ [_{V'}\ [_{V}\ \text{said}\ ]\ ]]]$

d. $[_{v'}\ \text{what}_{[\text{wh}:\boxed{vV}]}\ [_{v'}\ v_{[\bullet D\bullet]}\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]$

e. $[_{vP}\ \text{she}\ [_{v'}\ \text{what}_{[\text{wh}:\boxed{vV}]}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]$

f. $[_{T'}\ T_{[\bullet X_T\bullet]\succ[\bullet\text{epp}\bullet]}\ [_{vP}\ \text{she}\ [_{v'}\ \text{what}_{[\text{wh}:\boxed{vV}]}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]$

g. $[_{T'}\ \text{what}_{[\text{wh}:\boxed{TvV}]}\ T_{[\bullet\text{epp}\bullet]}\ [_{vP}\ \text{she}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]$

h. $[_{TP}\ \text{she}\ [_{T'}\ \text{what}_{[\text{wh}:\boxed{TvV}]}\ T\ [_{vP}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]]$

i. $[_{C'}\ C_{[\bullet X_C\bullet]}\ [_{TP}\ \text{she}\ [_{T'}\ \text{what}_{[\text{wh}:\boxed{TvV}]}\ T\ [_{vP}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]]]$

j. $[_{CP}\ \text{what}_{[\text{wh}:\boxed{CTvV}]}\ [_{C'}\ C\ [_{TP}\ \text{she}\ [_{T'}\ T\ [_{vP}\ [_{v'}\ [_{v'}\ v\ [_{VP}\ [_{V'}\ [_V\ \text{said}\ ]\ ]]]]]]]]]$

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 (35-l) 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 (27-b)); movement to matrix Specv in

(35-n) adds v on top of the list and deletes v at the bottom; and movement to matrix SpecT in (35-q) does the same with T. In all three cases, a feature list results that does not conform to f-seq (viz., $\boxed{\text{VCTv}}$, $\boxed{\text{vVCT}}$, and $\boxed{\text{TvVC}}$); 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 (35-t) *is* triggered by an inherent movement-inducing feature of the matrix interrogative C head, viz., $[\bullet\text{wh}_{\text{C}}\bullet]$, 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. Therefore, the Williams Cycle is (non-vacuously) satisfied by the final movement step.

(35) *Matrix clause*:

k. $[_{\text{V}'}$ $[_{\text{V}}$ think$]_{[\bullet\text{X}_{\text{V}}\bullet]}$ $[_{\text{CP}}$ what$_{[\text{wh:}\boxed{\text{CTvV}}]}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]
l. $[_{\text{VP}}$ what$_{[\text{wh:}\boxed{\text{VCTv}\cancel{\text{V}}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ think] $[_{\text{CP}}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]]
m. $[_{\text{v}'}$ v$_{[\bullet\text{X}_{\text{v}}\bullet]\succ[\bullet\text{D}\bullet]}$ $[_{\text{VP}}$ what$_{[\text{wh:}\boxed{\text{VCTv}\cancel{\text{V}}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ think] $[_{\text{CP}}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]]]
n. $[_{\text{v}'}$ what$_{[\text{wh:}\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ $[_{\text{v}'}$ v$_{[\bullet\text{D}\bullet]}$ $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ think] $[_{\text{CP}}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]]]]
o. $[_{\text{vP}}$ you $[_{\text{v}'}$ what$_{[\text{wh:}\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ think] $[_{\text{CP}}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she said]]]]]]]]
p. $[_{\text{T}'}$ T$_{[\bullet\text{X}_{\text{T}}\bullet]\succ[\bullet\text{epp}\bullet]}$ $[_{\text{vP}}$ you $[_{\text{v}'}$ what$_{[\text{wh:}\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]
q. $[_{\text{T}'}$ what$_{[\text{wh:}\boxed{\text{TvVC}\cancel{\text{TvV}}}]}$ $[_{\text{T}'}$ T$_{[\bullet\text{epp}\bullet]}$ $[_{\text{vP}}$ you $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]]
r. $[_{\text{TP}}$ you $[_{\text{T}'}$ what$_{[\text{wh:}\boxed{\text{TvVC}\cancel{\text{TvV}}}]}$ $[_{\text{T}'}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]]]
s. $[_{\text{C}'}$ C$_{[\bullet\text{wh}\bullet]}$ $[_{\text{TP}}$ you $[_{\text{T}'}$ what$_{[\text{wh:}\boxed{\text{TvVC}\cancel{\text{TvV}}}]}$ $[_{\text{T}'}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ think she said]]]]]]]]
t. $[_{\text{CP}}$ what$_{[\text{wh:}\boxed{\text{CTvV}\cancel{\text{CTvV}}}]}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ you think she said]]] ($\checkmark$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.

4.4 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 ($[\Sigma{:}\square]$) features on the attracting V or v head and the moved item, respectively. In (36) (= (1-a)), scrambling must target a Specv position, with the subject DP *keiner* ('no-one') staying in situ.

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

The derivation is straightforward, and shown in (37). Extremely local movement to SpecV (which is also string-vacuous, given the SOV nature of German) takes place at first (see (37-b)). This movement step is brought about by an edge feature $[\bullet\text{X}_{\text{V}}\bullet]$ 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 (37-e)). 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 (37-fg).

(37) 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. $[_{v'}$ $[_{VP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{V}]}$ $[_{V'}$ $[_{V}$ liest]]] $v_{[\bullet D\bullet]\succ[\bullet\Sigma\bullet]}$]
d. $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{V}]}$ $[_{V'}$ $[_{V}$ liest]]] $v_{[\bullet\Sigma\bullet]}$]]
e. $[_{vP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]] (√f-seq)
f. $[_{TP}$ $[_{vP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]] T]
g. $[_{CP}$ $[_{C}$ dass] $[_{TP}$ $[_{vP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ keiner $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]] T]]

4.5 The Ban on Long-Distance Scrambling

Things are different with illegitimate long-distance scrambling in German. A relevant example is repeated here as (38) (= (1-b)).

(38) *dass Karl das Buch glaubt $[_{CP}$ dass keiner t_1 liest]
that $Karl_{nom}$ the $book_{acc}$ thinks that $no\text{-}one_{nom}$ 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.–j. of (34). 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 (39-a)–(39-i). The list $\boxed{CTvV}$ 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).

(39) *Embedded clause*:

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. $[_{v'}$ $[_{VP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{V}]}$ $[_{V'}$ $[_{V}$ liest]]] $v_{[\bullet X_v\bullet]\succ[\bullet D\bullet]}$]
d. $[_{v'}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] $v_{[\bullet D\bullet]}$]]
e. $[_{vP}$ keiner $[_{v'}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]]
f. $[_{T'}$ $[_{vP}$ keiner $[_{v'}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{vV}]}$ $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]] $T_{[\bullet X_T\bullet]}$]
g. $[_{TP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{TvV}]}$ $[_{T'}$ $[_{vP}$ keiner $[_{v'}$ $[_{v'}$ $[_{VP}$ $[_{V'}$ $[_{V}$ liest]]] v]]] T]]
h. $[_{C'}$ $[_{C}$ dass $]_{[\bullet X_C\bullet]}$ $[_{TP}$ $[_{DP}$ das Buch $]_{[\Sigma:\boxed{TvV}]}$ $[_{T'}$ $[_{vP}$ keiner liest v] T]]]
i. $[_{CP}$ $[_{DP}$ 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 (35); see (40-k), which gives rise to a Σ-feature list $\boxed{VCTv}$, which is at variance with f-seq but 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 (35), is fatal; see (40-n): Movement to Specv gives rise to a list $\boxed{vVCT}$ 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 betweem f-seq and the feature list on the moved DP a violation of the Williams Cycle. The derivation given here also includes further steps: CP extraposition, with triggering features ignored to simplify exposition;

Merge of T; optional EPP-driven movement of the matrix subject DP to SpecT (cf. Grewendorf (1989)); and Merge of C; see (40-o)–(40-r). However, these steps are given here merely for the sake of clarity, to show what would have subsequently happened had there not been a fatal step of improper movement: The derivation crashes after the scrambled item has reached its criterial position in the matrix vP, and the Williams Cycle is violated.

(40) *Matrix clause*:

j. [$_{V'}$ [$_{CP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{CTvV}}]}$ [$_{C'}$ [$_{C}$ dass] [$_{TP}$ keiner liest v T] [$_{V}$ glaubt]$_{[\bullet X_V \bullet]}$]]]

k. [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{VCTv}\cancel{\text{V}}}]}$ [$_{V'}$ [$_{CP}$ [$_{C'}$ [$_{C}$ dass] [$_{TP}$ keiner liest v T] [$_{V}$ glaubt]]]]]

l. [$_{v'}$ [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{VCTv}\cancel{\text{V}}}]}$ [$_{V'}$ [$_{CP}$ dass keiner liest] [$_{V}$ glaubt]]] v$_{[\bullet D\bullet]\succ[\bullet\Sigma\bullet]}$]

m. [$_{v'}$ Karl [$_{v'}$ [$_{VP}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{VCTv}\cancel{\text{V}}}]}$ [$_{V'}$ [$_{CP}$ dass keiner liest] [$_{V}$ glaubt]]] v$_{[\bullet\Sigma\bullet]}$]]

n. [$_{v'}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ [$_{v'}$ Karl [$_{v'}$ [$_{VP}$ [$_{V'}$ [$_{CP}$ dass keiner liest] [$_{V}$ glaubt]]] v]]]

(*f-seq → *Williams Cycle → crash)

o. [$_{vP}$ [$_{v'}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ [$_{v'}$ Karl [$_{v'}$ [$_{VP}$ [$_{V'}$ [$_{V}$ glaubt]]] v] [$_{CP}$ dass keiner liest]]]]

p. [$_{T'}$ [$_{vP}$ [$_{v'}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ [$_{v'}$ Karl [$_{v'}$ glaubt dass keiner liest]]]] T$_{[\bullet epp\bullet]}$]

q. [$_{TP}$ Karl [$_{T'}$ [$_{vP}$ [$_{v'}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ [$_{v'}$ [$_{v'}$ glaubt dass keiner liest]]]] T]]

r. [$_{CP}$ [$_{C}$ dass] [$_{TP}$ Karl [$_{T'}$ [$_{vP}$ [$_{v'}$ [$_{DP}$ das Buch]$_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ [$_{v'}$ [$_{v'}$ glaubt dass keiner liest]]]] T]]]

Note that the same consequence arises if long-distance scrambling targets SpecV rather than Specv (which might be an option yielding the same string (38) 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 scrambling cannot feed subsequent wh-movement 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 $[\bullet\Sigma_v\bullet]$ gives rise to certain properties, like absence of weak crossover effects and absence of superiority effects, whereas checking of a pure edge feature $[\bullet X_v\bullet]$ (in the same position), or checking of $[\bullet wh_C\bullet]$ does not.[24]

Thus, on this view, clause-bound wh-movement in (41-a) does not induce a superiority effect, and clause-bound wh-movement in (41-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 (41-a), and *wen*$_1$ ('whom')

[24] The question of why exactly scrambling – conceived of as checking of $[\bullet\Sigma_{v/V}\bullet]$ – has these consequences is immaterial in the present context. Still, for weak crossover effects, one may assume that checking of $[\bullet\Sigma_{v/V}\bullet]$ 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 (independently of what ultimately derives these effects; see Müller (2011) and above).

in (41-b)), has undergone an intermediate movement step to a Specv position in the same clause by virtue of an optional inherent feature [$\bullet\Sigma_{v}\bullet$] on v (and a matching movement-related [Σ]-feature on the DP); and not by virtue of [$\bullet X_{v}\bullet$] on v.

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

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

Given that discharge of a movement-inducing (edge or inherent) feature on the VP and vP levels in (41-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 (42-a)) and the existence of weak crossover effects with long-distance wh-movement (see (42-b)) follow without further ado; see Frey (1993), Büring & Hartmann (1994), Fanselow (1996), Heck & Müller (2000), and Pesetsky (2000) for discussion of this phenomenon. It is the presence of a criterial ([$\bullet\Sigma_{v/V}\bullet$]-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 in German (because of the Williams Cycle), superiority effects and weak crossover effects cannot be circumvented.

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

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

4.6 Super-Raising

The ban on super-raising can be derived in a similar way as the ban on long-distance scrambling. A relevant example is repeated here as (43) (cf. (4-b)).

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

By assumption, the relevant movement-related feature on *Mary* is [EPP]; matrix T bears the corresponding structure-building feature [$\bullet$EPP$\bullet$]. 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 [$\bullet$EPP$\bullet$] 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 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 discussed in the previous section. The example in (19) is repeated here as (44-a), and accompanied by a slightly different construction instantiating the same problem.

(44) a. *Who$_1$ seems [$_{CP}$ C t$_1$ will leave] ?

b. *What$_1$ seems [$_{CP}$ that it was said t$_1$] ?

The analysis is straightforward: EPP-driven movement to matrix SpecT gives rise to a fatal 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 (there is no back-tracking or look-ahead).

4.7 Other Local Movement Types

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 paper is derived.

5. Extensions

Finally, I will discuss a number of possible extensions and modifications of the present analysis.[25] I begin with a movement type that affects a DP-internal position.

5.1 DP-Internal PP Preposing

German has a movement type 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 (45).

(45)	$[_{\mathrm{DP}_2}$ $[_{\mathrm{PP}_1}$	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$\underline{\mathrm{DNP}}$. This reflects the fact that C (rather than D) is the root node, and that nominals are typically parts of clauses. Under this assumption, DP-internal PP preposing in local contexts, as in (45), 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 $[\omega{:}\square]$ on the PP), the list on the movement-related feature on the moved PP in the criterial SpecD target position respects the Williams Cycle: $[\omega{:}\boxed{\mathrm{DN}}]$. 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.

However, things should be different for 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 4.5).

These predictions are borne out. Consider first (46), an instance of long-distance topicalization of a PP from an argument CP embedded in an object DP. The example has an intermediate status, as is typical of

[25] I hasten to add that some of 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 – is possible.

CNPC violations with argument extraction, which are standard cases of weak islands.[26] However, (46) 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.

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

Consider next (47), 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 weak locality (CNPC) effect were involved; in particular, the contrast to (46) is striking.[27] 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 (47) fails to do this; consequently, the movement-related feature, which is valued as $\boxed{\text{DNCTvV}}$ in the criterial position, violates the Williams Cycle, and the derivation crashes.

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

Finally, in (48), 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.

(48) *Es kennt [$_{\text{PP}_1}$ über die Liebe] keiner [$_{\text{D}'}$ das/ein Gerücht [$_{\text{CP}}$ dass sie ein Buch t_1 geschrieben hat]]]
EXPL knows about the love no-one the/a rumour that she a book written has
*[Σ: $\boxed{\text{vVDNCT}\cancel{\text{vVDN}}}$]

5.2 Relativization

As noted by Bayer & Salzmann (2009) (also Plank (1983) for some preliminary remarks in this direction), many speakers of German do not permit long-distance relativization (in contrast to wh-movement or topicalization); see, e.g., (49-ab).

(49) a. *Das ist einer [$_{\text{RelP}}$ der$_1$ ich glaube [$_{\text{CP}}$ dass t_1 das schaffen wird]]
this is one who I believe that this manage will

b. *der Mann [$_{\text{RelP}}$ den$_1$ ich glaube [$_{\text{CP}}$ dass Maria t_1 liebt]]
the man who I believe that Maria loves

Given the present analysis, this follows as an instance of improper movement from the Williams Cycle if C c-commands a functional head Rel that provides the landing site for relativization in German according to

[26] For present purposes, I leave open how CNPC island effects should be derived, and whether they should be derived by invoking concepts of core grammar in the first place (rather than, e.g., parsing difficulties, as in Sag et al. (2008)).

[27] 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.*

f-seq. Under this assumption, the relative pronouns in (49) bear a feature [rel] in the criterial position that violates the Williams Cycle: *[rel:$\boxed{\text{RelTvVC}}$].[28]

5.3 Exceptional Case Marking Constructions

Abels (2008) observes that a strict interpretation of the Williams Cycle (like the one adopted in the present paper) 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)). 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 $\boxed{\text{vVT}}$ (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 list would violate the Williams Cycle.[29] 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 (50-abc) (where (50-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 (50-c)). This is an argument for exceptional case marking, and against raising to object.

(50) a. John promised repeatedly to leave
b. #John believed Mary repeatedly to have left (She left repeatedly.)
c. *John believed Mary sincerely to have left

On the other hand, Lasnik (1999) points out that sentences like (51) in English permit anaphoric binding of *each other*$_1$ 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*$_1$*'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.

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

Taken together, though, it seems fair to conclude that there is no uncontroversial case for raising to object

[28] Ultimately, if this analysis is to be more than a rough sketch, the central assumption about the structure of the left periphery required here would have to be substantiated by independent empirical evidence. Boef (2012) notes that varieties of Dutch permit a co-occurrence of a relative pronoun and a right-adjacent C item *dat*; interestingly, long-distance relativization also seems to be entirely unproblematic in all Dutch varieties. Varieties of German that permit a co-occurrence of a relative pronoun and some other head element to its right in the left periphery (or, indeed, the sole presence of such an element) always seem to use a form for this latter item that is very different from standard C elements (viz., *wo* ('where'), never uncontroversial C elements like *dass* ('that') or *ob* ('whether')); *wo* is arguably best analyzed as a Rel element. Thus, as far as I can see, the question of whether C is higher or lower than Rel cannot be decided by simple empirical evidence involving a complementizer *dass* and the relative pronoun because these two items cannot co-occur in varieties of German.

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

to be made yet. Therefore, at present it is unclear whether ECM constructions pose a problem for a strict interpretation of the Williams Cycle.

5.4 Languages that Permit Long-Distance Scrambling, Languages that Permit Super-Raising

The issue of ECM constructions notwithstanding, 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 (i.e., across an XP that contains a complementizer) 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 data from (colloquial) Russian taken from Zemskaja (1973) illustrate long-distance scrambling.

(52) a. Ty [$_{DP}$ doktor]$_1$ videl [$_{CP}$ kogda t$_1$ pod"ezžal] ?
you doctor$_{nom}$ saw when came

b. Vy [$_{DP}$ pocylku]$_1$ videli [$_{CP}$ 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 (53) (see Perlmutter & Soames (1979, 156)):

(53) [$_{DP}$ I kopeles]$_1$ fenonde [$_{CP}$ 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 valuation and/or deletion of movement-related features on moved items can be employed in certain constructions and languages.[30]

For concreteness, I would like to suggest that a key to a solution of the problem posed by data such as those in (52) and (53) 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 (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

[30] Also see Obata & Epstein (2011) for this general strategy; and also note that these counter-examples also raise problems for virtually all other existing analyses of improper movement, at least if taken in their entirety.

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 [±Π], where Π stands for propositionality (in an extended sense): C and v are characterized by [+Π], and T and V are characterized by [–Π].[31] The crucial assumption now is that deletion in the lists of movement-related features may not have to apply under full identity; "categorial information" in the sense of (27-b) may refer only to a small (but fundamental) part of the category label, viz., information related to the [±Π] status of the phase head.[32]

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 paper: 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, the symbol v deletes an earlier v, and so forth). Second, another option is that only the (most basic) feature [±Π] 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 feature list (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 (54-abc), for long-distance movement to the VP, vP , and TP domains, respectively.[33]

(54) a. (i) * $\boxed{\text{VC}\cancel{\text{TvV}}}$ *f-seq

(ii) *[$_{\text{VP}}$ DP$_1$ .. [$_{\text{CP}}$ t_1'''' .. [$_{\text{TP}}$ t_1''' .. [$_{\text{vP}}$ t_1'' .. [$_{\text{VP}}$ t_1' .. t_1 ..]]]]]
V([–Π]) C([+Π]) ~~T([–Π])~~ ~~v([+Π])~~ ~~V([–Π])~~

b. (i) $\boxed{\text{vV}\cancel{\text{CTvV}}}$ √f-seq

(ii) [$_{\text{vP}}$ DP$_1$ .. [$_{\text{VP}}$ t_1''''' .. [$_{\text{CP}}$ t_1'''' .. [$_{\text{TP}}$ t_1''' .. [$_{\text{vP}}$ t_1'' .. [$_{\text{VP}}$ t_1' .. t_1 ..]]]]]]
v([+Π]) V([–Π]) ~~C([+Π])~~ ~~T([–Π])~~ ~~v([+Π])~~ ~~V([–Π])~~

c. (i) $\boxed{\text{Tv}\cancel{\text{VCTvV}}}$ √f-seq

(ii) [$_{\text{TP}}$ DP$_1$ .. [$_{\text{vP}}$ t_1'''''' .. [$_{\text{VP}}$ t_1''''' .. [$_{\text{CP}}$ t_1'''' .. [$_{\text{TP}}$ t_1''' .. [$_{\text{vP}}$ t_1'' .. [$_{\text{VP}}$ t_1' ..
T([–Π]) v([+Π]) ~~V([–Π])~~ ~~C([+Π])~~ ~~T([–Π])~~ ~~v([+Π])~~ ~~V([–Π])~~
t_1 ..]]]]]

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. (However, 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 (which requires full categorial identity) but still more restrictive than the second option. Now C and v delete one another, but T and V do not. The effects are shown in (55-abc), again for long-distance movement to the VP, vP , and TP domains.

(55) a. (i) * $\boxed{\text{VCT}\cancel{\text{vV}}}$ *f-seq

---

[31] Of course, 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. (To ensure that they do not play a role, one might postulate a geometric organization of the primitive features for categories as it has been proposed by Harley & Ritter (2002) for ϕ-features, with [±Π] at the top; accompanied by a restriction that either top-node identity or full identity may lead to deletion.) – 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.

[32] In what follows, I focus on improper movement effects in the clausal domain. I leave open how to extend the analysis in terms of primitive features and underspecification to the nominal domain, but there should be no severe problems with this.

[33] In each case I present the feature list as it looks in the criterial position first (i); (ii) then illustrates the effects of the actual derivation on each level. (Here I add traces as mnemonic devices; traces (or copies) as such play no role in the present approach.)

(ii) $^*[_{\text{VP}}$ DP$_1$ .. $[_{\text{CP}}$ t_1'''' .. $[_{\text{TP}}$ t_1''' .. $[_{\text{vP}}$ t_1'' .. $[_{\text{VP}}$ t_1' .. t_1 ..]]]]]
V C([+Π]) T $\cancel{\text{v([+Π])}}$ $\cancel{\text{V}}$

b. (i) $^*\boxed{\text{vV}\cancel{\text{C}}\text{T}\cancel{\text{v}}\cancel{\text{V}}}$ *f-seq

(ii) $^*[_{\text{vP}}$ DP$_1$ .. $[_{\text{VP}}$ t_1''''' .. $[_{\text{CP}}$ t_1'''' .. $[_{\text{TP}}$ t_1''' .. $[_{\text{vP}}$ t_1'' .. $[_{\text{VP}}$ t_1' .. t_1 ..]]]]]
v([+Π]) V $\cancel{\text{C([+Π])}}$ T $\cancel{\text{v([+Π])}}$ $\cancel{\text{V([–Π])}}$

c. (i) $\boxed{\text{TvV}\cancel{\text{C}}\cancel{\text{T}}\cancel{\text{v}}\cancel{\text{V}}}$ √f-seq

(ii) $[_{\text{TP}}$ DP$_1$ .. $[_{\text{vP}}$ t_1'''''' .. $[_{\text{VP}}$ t_1''''' .. $[_{\text{CP}}$ t_1'''' .. $[_{\text{TP}}$ t_1''' .. $[_{\text{vP}}$ t_1'' .. $[_{\text{VP}}$ t_1' .. t_1 ..]]]]]
T v([+Π]) V $\cancel{\text{C([+Π])}}$ $\cancel{\text{T}}$ $\cancel{\text{v([+Π])}}$ $\cancel{\text{V}}$

Finally, the fourth option is 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. (56-abc).

(56) a. (i) $^*\boxed{\text{VC}\cancel{\text{T}}\text{v}\cancel{\text{V}}}$ *f-seq

(ii) $^*[_{\text{VP}}$ DP$_1$ .. $[_{\text{CP}}$ t_1'''' .. $[_{\text{TP}}$ t_1''' .. $[_{\text{vP}}$ t_1'' .. $[_{\text{VP}}$ t_1' .. t_1 ..]]]]]
V([–Π]) C $\cancel{\text{T([–Π])}}$ v $\cancel{\text{V([–Π])}}$

b. (i) $^*\boxed{\text{vVC}\cancel{\text{T}}\cancel{\text{v}}\cancel{\text{V}}}$ *f-seq

(ii) $^*[_{\text{vP}}$ DP$_1$ .. $[_{\text{VP}}$ t_1''''' .. $[_{\text{CP}}$ t_1'''' .. $[_{\text{TP}}$ t_1''' .. $[_{\text{vP}}$ t_1'' .. $[_{\text{VP}}$ t_1' .. t_1 ..]]]]]
v V([–Π]) C $\cancel{\text{T([–Π])}}$ $\cancel{\text{v}}$ $\cancel{\text{V([–Π])}}$

c. (i) $^*\boxed{\text{Tv}\cancel{\text{V}}\text{C}\cancel{\text{T}}\cancel{\text{v}}\cancel{\text{V}}}$ *f-seq

(ii) $^*[_{\text{TP}}$ DP$_1$ .. $[_{\text{vP}}$ t_1'''''' .. $[_{\text{VP}}$ t_1''''' .. $[_{\text{CP}}$ t_1'''' .. $[_{\text{TP}}$ t_1''' .. $[_{\text{vP}}$ t_1'' .. $[_{\text{VP}}$ t_1' .. t_1 ..]]]]]
T([–Π]) v $\cancel{\text{V([–Π])}}$ C $\cancel{\text{T([–Π])}}$ $\cancel{\text{v}}$ $\cancel{\text{V([–Π])}}$

Assuming that the choice among the four options is fixed once and for all in a given language, languages like Greek that permit super-raising (but not long-distance scrambling) could choose the option in (55) (with shared [+Π] permitting deletion in addition to full identity).[34] Languages like Russian could choose the option in (54) (with shared [+Π] and [–Π] permitting deletion in addition to full identity).[35] However, the option in (55) also implies that long-distance movement to the TP domain should be possible. Stepanov (2007) argues that this is not the case in Russian since there are no legitimate cases of super-raising. If he is right about the reanalysis of constructions that might at first sight suggest an availability of super-raising in Russian, and if there is no reason to look for an alternative account, this could be taken to imply that the choice among the four options of deletion in feature lists varies among indivudual movement-related features (in the case at hand: [Σ:□] vs. [EPP:□] as F in (27-b)), rather than among languages.

As for option (56), at present 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 paper). A relevant case to look at in this context is, again ECM constructions (recall section 5.3). Recall that if the raising to object (i.e., movement to Specv) analysis is adopted, then English ECM constructions, analyzed

---

[34] 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 potentially problematic for the feature splitting approach.

[35] It follows that all long-distance scrambling minimally has to target Specv, which looks plausible.

via TP embedding, will give rise to unwanted violations of the Williams Cycle because of 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. However, if option (56) 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.)

Similar conclusions might be drawn independently of the issue of raising to object for scrambling from ECM complements in German if one assumes that ECM complements are TPs in this language. If they are, scrambling from ECM complements as in (57) poses a potential problem for the present analysis because a feature list $\boxed{\text{vVT}}$ will come into being after movement of DP to the matrix VP or vP domain (or, at any rate, a matrix TP-internal domain) if deletion in feature lists only takes place under full identity.

(57) a. dass der Kollege [$_{\text{DP}}$ den Antrag]$_1$ [$_{\text{XP}}$ seine Mitarbeiter t$_1$ gerade schreiben lässt]
that the colleague$_{nom}$ the proposal$_{acc}$ his co-workers$_{acc}$ currently write lets

b. dass [$_{\text{DP}}$ den Antrag]$_1$ der Kollege [$_{\text{XP}}$ seine Mitarbeiter t$_1$ gerade schreiben lässt]
that the proposal$_{acc}$ the colleague$_{nom}$ his co-workers$_{acc}$ currently write lets

The same conclusion can be drawn for pronoun fronting from ECM complements in German; see (58).

(58) a. dass er es$_1$ [$_{\text{XP}}$ den Jungen t$_1$ lesen sah]
that he$_{nom}$ it$_{acc}$ the boy$_{acc}$ read saw

b. dass er es$_1$ [$_{\text{XP}}$ den Jungen t$_1$ machen ließ]
that he$_{nom}$ it$_{acc}$ the boy$_{acc}$ make let

This potential problem disappears if the option in (56) is in fact adopted in German. On this view, super-raising and scrambling from finite clauses (more generally, from CPs) will be excluded as before (because of the fatal presence of C in the feature list which cannot be deleted before matrix C is reached), but scrambling and pronoun fronting from a TP complement to a TP-internal domain are predicted to be possible, yielding a feature list $\boxed{\text{vV}}$ that conforms to the Williams Cycle because the problematic T symbol has been removed by discharge of the edge feature on the matrix VP cycle. All that said, it is by no means clear that such a move is required for German: As argued by Fanselow (1991, 120) Wurmbrand (2001, 216), and Haider (2010, 309ff), there is evidence against a TP level for the ECM complements in German labelled as XP in (57) and (58); Fanselow and Wurmbrand suggest a vP/VP embedding, which renders the data in (57) and (58) unproblematic from the perspective of the Williams Cycle even if the original deletion operation in feature lists is adopted that relies on full categorial information.[36]

[36] Note that there are indeed restrictions on scrambling from ECM constructions in German; in particular, as observed by Höhle (1978, 56-57), Thiersch (1978, 168-169), and Grewendorf (1989, 150), dative DPs cannot be scrambled or pronoun-moved from these complements; see (i-ab).

(i) a. *dass keiner [$_{\text{DP}}$ dieser Frau]$_1$ [$_{\text{XP}}$ den Jungen t$_1$ helfen sah/ließ]
that no-one$_{nom}$ this woman$_{dat}$ the boy$_{acc}$ help saw/let

b. *dass er [$_{\text{DP}}$ ihm]$_1$ [$_{\text{XP}}$ den Jungen t$_1$ helfen sah/ließ]
that he$_{nom}$ him$_{dat}$ the boy$_{acc}$ help saw/let

One might think that the illformedness of (i-ab) might be traced back to the Williams Cycle, based on a TP analysis of XP and the assumption that only full categorial identity can lead to deletion in feature lists. However, this restriction cannot be derived by invoking the theory of improper movement. For one thing, accusative DPs are not affected by it (as just seen in the main text); for another, as also noted by Höhle (1978), wh-movement of dative DPs to the matrix SpecC is also not possible, 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 proposed that languages (or perhaps even movement types) can choose whether only 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. If the latter is the case, three options arise: Either [+Π] is the relevant subfeature, or [–Π] is, or both are ([±Π]). Ultimately, the question to what extent individual languages make use of these 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 paper.

6. Conclusion

The main result of the present study 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 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), and Neeleman & van de Koot (2010). In all existing analyses where a version of the Williams Cycle is put to use, it is 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. In contrast, in the reformulation that I have suggested, all relevant pieces of information are locally available; no more structure needs to be considered than the attracting phase head and the moved item in its specifier. 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 it passes through is successively picked up and registered in a buffer by a 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

---

which would be completely unexpected under an account in terms of the Williams Cycle; cf. (ii).

(ii) *Wem$_1$ sah/ließ Karl [$_{XP}$ den Jungen t$_1$ helfen] ?
whom$_{dat}$ saw/let Karl$_{nom}$ the boy$_{acc}$ help

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 associated, as it is typically assumed for other kinds of feature valuation (that said, instances of case stacking in the world's languages arguably also require multiple valuation of a single feature).

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.[37] Still, for the time being, the approach chosen here strikes me as the most straightforward one because it is both fine-grained and potentially flexible, and (not least of all) because it structurally assimilates the operation to other syntactic operations (in particular, to Agree under matching).

Needless to say, the proposed analysis of improper movement also gives rise to a number of further questions.[38] For reasons of space and coherence, I will not try to pursue them here.

[37] For instance, one could devise a minimally different approach in which there is no actual deletion in feature lists at all; or one could devise another minimally different approach in which there is much more deletion in feature lists. A conceptually somewhat more radical departure might be to give up the assumption that every phrase is a phase (see section 3.), 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, 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.

[38] Here is one area about which much more would eventually have to be said. It concerns other functional heads in the clausal and nominal domains than just C, T, v, and D. Suppose, for instance, that the embedded clause contains a negation, whereas the matrix clause does not (Peter Biskup, p.c.). If there is a NegP as part of the clausal spine, and if every phrase is a phase, successive-cyclic movement to the matrix clause must go through this position, registering the categorial information on the moved item. The Williams Cycle ultimately requires a deletion of this information, and the question then is what phase head of the matrix clause can do this. Several options arise. For instance, there could be a Pol(arity)P throughout, and Pol:Pos in the matrix clause can delete Pol:Neg in the embedded clause; or there could be a functional head that can simultaneously delete two sufficiently similar pieces of information (for instance, T deletes T and Neg); or NegP is in fact not part of the functional spine; etc.

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 (2012): 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.

Bailyn, John Frederick (2001): On Scrambling: A Reply to Bošković and Takahashi, *Linguistic Inquiry* 32, 635–658.

Bayer, Josef & Martin Salzmann (2009): That-Trace Effects and Resumption as the Result of Improper Movement. Ms., Universität Konstanz.

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.

Boef, Eefje (2012): Long-Distance Relativization in Varieties of Dutch. In: *Proceedings of Console 17.* Leiden, pp. 77–100.

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.

Büring, Daniel & Katharina Hartmann (1994): The Dark Side of Wh-Movement, *Linguistische Berichte* 149, 56–74.

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 (1973): Conditions on Transformations. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle.* Academic Press, New York, pp. 232–286.

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 (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 (2008): On Phases. In: R. Freidin, C. Otero & M. L. Zubizarreta, eds., *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, Mass., pp. 133–166.

Collins, Chris (2002): Eliminating Labels. In: S. D. Epstein & T. D. Seely, eds., *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 42–64.

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 (1996): The Proper Interpretation of the Minimal Link Condition. Ms. Universität Potsdam.

Fischer, Silke (2006): Matrix Unloaded: Binding in a Local Derivational Approach, *Linguistics* 44, 913–935.

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.

Frey, Werner (1993): *Syntaktische Bedingungen für die Interpretation*. Akademieverlag, Berlin.

Grewendorf, Günther (1988): *Aspekte der deutschen Syntax*. Narr.

Grewendorf, Günther (1989): *Ergativity in German*. Foris, Dordrecht.

Grewendorf, Günther (2003): Improper Remnant Movement, *Gengo Kenkyo: The Journal of the Linguistic Society of Japan* 123, 47–94.

Grewendorf, Günther (2004): Die linke Bewegung von Restkategorien. Ms., Universität Frankfurt/Main.

Grewendorf, Günther & Joachim Sabel (1999): Scrambling in German and Japanese, *Natural Language and Linguistic Theory* 17, 1–65.

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 (1984): Topic, Focus, and V-Second, *Groninger Arbeiten zur Germanistischen Linguistik* 25, 72–120.

Haider, Hubert (2010): *The Syntax of German*. Cambridge University Press, Cambridge.

Halle, Morris & Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. In: K. Hale & S. J. Keyser, eds., *The View from Building 20*. MIT Press, Cambridge, Mass., pp. 111–176.

Harley, Heidi & Elisabeth Ritter (2002): Person and Number in Pronouns: A Feature-Geometric Analysis, *Language* 78, 482–526.

Heck, Fabian (2010): Against Direct Recursion in Syntax. Ms., Universität Leipzig.

Heck, Fabian & Gereon Müller (2000): Successive Cyclicity, Long-Distance Superiority, and Local Optimization. In: R. Billerey & B. D. Lillehaugen, eds., *Proceedings of WCCFL*. Vol. 19, Cascadilla Press, Somerville, MA, pp. 218–231.

Heim, Irene & Angelika Kratzer (1998): *Semantics in Generative Grammar*. Blackwell, Oxford.

Höhle, Tilman (1978): *Lexikalistische Syntax: Die Aktiv-Passiv-Relation und andere Infinitkonstruktionen im Deutschen*. Niemeyer, Tübingen.

Keine, Stefan (2010): *Case and Agreement from Fringe to Core. Impoverishment Effects on Agree*. Linguistische Arbeiten, Mouton de Gruyter, Berlin.

Lahne, Antje (2009): Where There is Fire There is Smoke. Local Modelling of Successive-Cyclic Movement. PhD thesis, Universität Leipzig.

Lasnik, Howard (1999): *Minimalist Analysis*. Blackwell, Oxford.

Lindauer, Thomas (1995): *Genitivattribute: Eine morphosyntaktische Untersuchung zum deutschen DP/NP-System*. Niemeyer, Tübingen.

May, Robert (1979): Must Comp-To-Comp Movement Be Stipulated?, *Linguistic Inquiry* 10, 719–725.

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 (2011): *Constraints on Displacement. A Phase-Based Approach*. Vol. 7 of *Language Faculty and Beyond*, Benjamins, Amsterdam.

Müller, Gereon & Wolfgang Sternefeld (1993): Improper Movement and Unambiguous Binding, *Linguistic Inquiry* 24, 461–507.

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.

Obata, Miki & Samuel David Epstein (2011): Feature-Splitting Internal Merge: Improper Movement, Intervention, and the A/A′ Distinction, *Syntax* 14, 122–147.

Perlmutter, David & Scott Soames (1979): *Syntactic Argumentation and the Structure of English*. The University of California Press, Berkeley.

Pesetsky, David (2000): *Phrasal Movement and Its Kin*. MIT Press, Cambridge, Mass.

Plank, Frans (1983): Transparent vs. Functional Encoding of Grammatical Relations: A Parameter for Synctactic Change and Typology, *Linguistische Berichte* 86, 1–13.

Postal, Paul (1974): *On Raising*. MIT Press, Cambridge, Massachusetts.

Reuland, Eric (2001): Primitives of Binding, *Linguistic Inquiry* 32, 439–492.

Richards, Marc (2007): On Feature Inheritance, *Linguistic Inquiry* 38, 563–572.

Rizzi, Luigi (1997): The Fine Structure of the Left Periphery. In: L. Haegeman, ed., *Elements of Grammar*. Kluwer, 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, Philip Hofmeister & Neal Snider (2008): Processing Complexity in Subjacency Violations: The Complex Noun Phrase Constraint. Ms., Stanford University. To appear in *Proceedings of CLS*.

Saito, Mamoru (1985): Some Asymmetries in Japanese and Their Theoretical Implications. PhD thesis, MIT, Cambridge, Mass.

Sauerland, Uli (1996): The Interpretability of Scrambling. In: M. Koizumi, M. Oishi & U. Sauerland, eds., *Formal Approaches to Japanese Linguistics 2*. MIT Working Papers in Linguistics, MITWPL, Cambridge, Mass., pp. 213–234.

Sauerland, Uli (1999): Erasability and Interpretation, *Syntax* 3, 161–188.

Sportiche, Dominique (1989): Le Mouvement Syntaxique: Contraintes et Paramètres, *Langages* 95, 35–80.

Starke, Michal (2001): Move Dissolves Into Merge: A Theory of Locality. PhD thesis, University of Geneva.

Stepanov, Arthur (2007): On the Absence of Long-Distance A-Movement in Russian, *Journal of Slavic Linguistics* 15, 81–108.

Sternefeld, Wolfgang (1992): Transformationstypologie und strukturelle Hierarchie. Ms., Universität Tübingen.

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.

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.

Thiersch, Craig (1978): Topics in German Syntax. PhD thesis, MIT, Cambridge, Mass.

Vikner, Sten (2005): Object Shift. In: H. v. Riemsdijk & M. Everaert, eds., *The Blackwell Companion to Syntax*. Blackwell, Oxford, pp. 392–436.

von Stechow, Arnim & Wolfgang Sternefeld (1988): *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, Opladen.

Williams, Edwin (1974): Rule Ordering in Syntax. PhD thesis, MIT, Cambridge, Mass.

Williams, Edwin (2003): *Representation Theory*. MIT Press, Cambridge, Mass.

Wurmbrand, Susanne (2001): *Infinitives. Restructuring and Clause Structure*. Mouton de Gruyter, Berlin.

Zemskaja, E.A. (1973): *Russkaja Razgovornaja Reč'*. Nauka, Moskva.