LINGUISTISCHE ARBEITS BERICHTE

Rule Interaction in Grammar

Edited by

Fabian Heck & Anke Assmann

90

Institut für Linguistik
Universität Leipzig
2013

Herausgeber: Institut für Linguistik
Universität Leipzig
Beethovenstr. 15
D-04107 Leipzig
www.uni-leipzig.de/~asw

ISSN 0138-4597

Druck: MERKUR DRUCK
Salomonstr. 20
D-04103 Leipzig

Vertrieb: Sabine Tatzelt
Institut für Linguistik
Universität Leipzig
Beethovenstr. 15
D-04107 Leipzig

Tel.: +341 97 37 610
Fax.: +341 97 37 609
E-mail: infl@uni-leipzig.de

Contents

Preface

Fabian Heck & Anke Assmann*

The notion of "rule" has always been at the heart of grammatical theory. Within transformational grammar of the 1960s and 70s rules were thought of as transformations, mapping representations onto other representations in both phonology and syntax (Chomsky 1965, Chomsky and Halle 1968, Anderson 1969, 1974, Kenstowicz and Kisseberth 1977, 1979, Williams 1974, Kayne 1975, Pullum 1979, Perlmutter and Soames 1979, among many others). It soon became clear that if the generation of a representation requires the application of more than one rule, then potential cases of rule interaction arise (Chomsky 1951). In order to describe interactions between rule applications, Kiparsky (1971, 1976) introduced a taxonomy of rule interaction that distinguishes transparent interactions (feeding, bleeding) and opaque ones (counter-feeding, counter-bleeding). Feeding describes an interaction where the application of a rule R_1 provides the context for the application of a rule R_2. Bleeding, on the other hand, is an interaction where application of a rule R_1 destroys the context for the application of a rule R_2. Counter-feeding is a term for an interaction where the feeding relation that could, in principle, apply between two rules does not arise. On the surface, it looks like a rule did not apply although its context was created by the application of another rule. Similarly, counter-bleeding describes an interaction where a bleeding relation that could show up between two rules is not found on the surface: a rule applied although its context was destroyed by another rule.

Extending and generalizing Kiparsky's (1971, 1976) taxonomy, one may distinguish interactions that involve a sequential application of rules from interactions that involve parallel rule application; alongside with this, there is a distinction between interactions that come about by rules that strengthen/support one

---

[*] We would like to thank all contributors for their cooperation. Particular thanks go to Sebastian Bank, who supported us typesetting this LAB volume.

Rule Interaction in Grammar, v–xiii
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

another (the interaction is excitatory) and those where rules weaken/restrain one another (the interaction is inhibitory). Cross-classifying these two binary oppositions, bleeding and counter-feeding (as they arise in rule based theories) fall under the notion of inhibitory sequential rule interaction; accordingly, feeding and counter-bleeding can be comprised under the notion of excitatory sequential rule interaction (see below for parallel rule interaction).

In the 1980s, there was a general shift towards representationalism in syntactic theory (Government and Binding, cf. Chomsky 1981; Generalized Phrase Structure Grammar, cf. Gazdar et al. 1985; Head Driven Phrase Structure Grammar, cf. Pollard and Sag 1994). The general approach consisted in formulating principles and filters that restrict the number freely generated representations to the well-formed ones. As a consequence, the interest in rule interactions was temporarily reduced. In the 1990s, the derivational view was resurrected within the syntactic framework of the Minimalist Programm (Chomsky 1995, et seq.; but see also Brody 1995, 2002, where a representational variant of this framework is pursued). In its context, it became common to assume that the rules, which were thought of as theoretical primitives before, are actually composed of more elementary operations, such as Merge, Move, Agree, etc. (see, e.g., Kitahara 1997, Řezáč 2004). This assumption created further room for potential rule interaction, which could now also involve elementary operations.

At about the same time, Optimality Theory (Prince and Smolensky 1993, 2004) became a popular stream of research in phonology. The standard variant of this theory involves generation of representations plus their subsequent evaluation by grammatical constraints. It is thus highly representational. Although one can think of the grammatical constraints in an optimality theoretical grammar as "triggering" rules in a certain sense, the architecture of parallel rule application (or constraint evaluation) does not allow for the full range of rule interaction known from derivational frameworks: Opaque interactions cannot be derived without additional assumptions. This is even the case in a branch of optimality theoretic research called Harmonic Serialism (McCarthy 2000, 2007), which reintroduces certain derivational traits into the theory (see also Hermans and van Oostendorp 2000 for related issues). There is, however, another sense in which Optimality Theory contributes to the topic of rule interaction. Namely, one of its basic tenets is that grammatical constraints can be in conflict with one another. The conflict is resolved by enforcing the fulfillment of the higher-ranked constraint at the expense of the non-fulfillment of the lower-ranked one. In this way, the theory introduces a new type of rule interaction included within the taxonomy mentioned above: inhibitory simultaneous rule interaction. And also cases of exhibitory simultaneous rule interaction can be found within optimality theoretic work: Analyses that involve the notion of "local conjunction" (Smolensky 1995) employ the idea that constraints may strengthen/support each other (see, e.g., Alderete 1997, Łubowicz 2005 for phonology and Legendre et al. 1998, Aissen 1999, 2003, Fischer 2001 for syntax).

The present volume of Linguistische Arbeitsberichte collects contributions that focus on the way rules interact within and across different grammatical components. It starts with papers that involve rule interaction across components. A common trait of the analyses they contain is that the independently motivated relative order of the components imposes an intrinsic order on the interaction of rules across these components.

The paper by **Daniela Thomas** deals with the question as to how split ergativity in subordination contexts in Sierra Popoluca and Jacaltec is derived. The leading idea is that there are filters that apply to the numeration, a component which precedes the syntax. One such filter, dubbed the *Constraint on Case Assignment in Intransitive Contexts*, determines which functional head assigns case in intransitive clauses and which does not. The application of this filter to the numeration leads to bleeding of absolutive case assignment in intransitive clauses in the syntax. Thus, the the assumption that the numeration is formed before the syntactic derivation proper starts induces an order among the constraints and operations that are associated with the numeration and the derivation, respectively.

Interactions between operations that apply in the syntactic component and those that apply post-syntactically are investigated in the paper by **Anke Assmann, Svetlana Edygarova, Doreen Georgi, Timo Klein & Philipp Weisser**. The proposal derives the case split on possessors in Udmurt by assuming that the syntax creates case stacking configurations that serve as an input for the post-syntactic morphological component, which then computes what case marker has to be inserted on the possessor. Thus, syntactic case stacking feeds the application of the post-syntactic, morphological rules fusion and vocabulary insertion.

The paper by **Martin Salzmann** focuses on the interaction between syntactic rules and PF rules. It presents a solution to an extraposition paradox (an observation due to Hubert Haider), which arises in the context of verb cluster formation in the Germanic languages, and an explanation for the difference in behaviour of the infinitival markers *te/zu* in Dutch and German, respectively. The first part of the proposal consists of ordering the syntactic rule of extraposition before the phonological rule of cluster formation. The second part is concerned with different rule orders within PF. Descriptively speaking, placement of the infinitival marker applies early in Dutch, thereby transparently interacting with PF-inversion within the verb-cluster (feeding); in German, lowering of the infinitival marker enters into an opaque interaction with inversion (counter-feeding) because it applies late.

We now introduce the papers of the present volume that deal with potential morphology-phonology interactions. Such an interaction is sometimes argued to be involved in the linearization of affixes. The paper by **Eva Zimmermann and Jochen Trommer** argues against such a view on the basis of the linearization of abstract mora affixes, resulting in morphological lengthening at the surface. It is suggested that all instances of bona fide mora affixation that are attested cross-linguistically, and only those, can be derived by assuming that moras are assigned to a small array of fixed positions within morphological representations. In contrast, theories that derive the ultimate positioning of mora affixes by the interaction of morphological and phonological rules are argued to overgenerate and undergenerate at the same time.

A case of superficial counter-bleeding in morphology-phonology interaction is discussed in the paper by **Barbara Stiebels**. It concerns the interaction of the phonological rule of dorsal fricative fronting and affixation of the diminutive *-ske* in Kleverlandish (a German dialect). On the one hand, affixation of *-ske* seems to be contingent on the feature [+VELAR] on the stem final consonant. On the other hand, affixation indirectly triggers dorsal fricative fronting of the stem final consonant, thereby destroying the context for affixation. Upon closer inspection, however, it turns out that the alleged opaque relation becomes transparent once it is assumed that the phonological context for diminutive affixation involves the feature [+DORSAL]. As Stiebels argues, this assumption is motivated on independent grounds. A further issue that is addressed in the paper concerns a rule order paradox between cluster simplification and diminutive affixation.

An underlying theme of both the paper by Aaron Doliana and the one by Anke Assmann is the idea that one and the same rule may show interaction within different components. Thus, **Aaron Doliana**'s paper explores a new argument for the idea that the operation of Impoverishment, which is standardly assumed to be morphological, may also interact with syntactic rules, in particular the operation Agree. The proposal consists of splitting up Agree into the more primitive operations Copy and Check. This creates an additional window for the application of Impoverishment in the syntax, leading to a feeding relation between Copy and Impoverishment and a bleeding relation between Impoverishment and Check. As a result, a broader range of typologically attested variants of the Person Case Constraint can be accounted for.

In a similar vein, the paper by **Anke Assmann** explores the possibility for pre-syntactic operations that create new contexts for the syntactic derivation. In order to capture the behavior of *wh*-phrases in free relative clauses, it is proposed that the operation Copy, which is usually assumed to be syntactic, may also apply pre-syntactically within the lexical array. More precisely, pre-syntactic copying affects the *wh*-head and creates a partial copy of it, which in turn allows for the creation of a new lexical item. The existence of this item, then, gives rise to new possibilities for the syntactic operations Merge and Agree; thus, pre-syntactic copying feeds syntactic rules.

Finally, we introduce the various proposals in the present volume that deal with the interaction of rules in narrow syntax. To begin with, the paper by **Gereon Müller** provides a derivational reformulation of a constraint called the Williams Cycle, which is concerned with the ban on improper movement: Ā-movement bleeds subsequent A-movement of the same category; but A-movement counter-bleeds subsequent Ā-movement. Standard formulations of the Williams Cycle are incompatible with a locality theory that is based on the Phase Impenetrability Condition. In order to remedy this situation, the proposal is made that every moved category stores aspects of its derivational history. What then violates the Williams cycle is not Ā-movement followed by A-movement as such but rather Ā-movement that is not ultimately terminated by Ā-movement (but by A-movement).

A closer inspection of the interaction of Move and Merge is presented in **Philipp Weisser**'s paper. It deals with the Left Subordinating *and*-Construction in English, which differs from other coordinations in that it does not obey the Coordinate Structure Constraint. To account for this exceptional behavior, it is proposed that clauses in this construction start out as subordinate adjuncts and then become coordinates of their superordinate clause via movement. This leads to a counter-bleeding relation because, usually, coordination bleeds asymmetric extraction. As often, opacity is accounted for by rule ordering:

Extraction is possible after all in this case because it takes place prior to the movement that creates coordination in the first place.

The paper by **Marie-Luise Popp** discusses the interaction of two Agree operations. In the languages Itomana and Basque, the goal for person agreement is determined by a person hierarchy. The difference between the two languages is that a combination of two arguments bearing local person results in inverse marking in Basque while it leads to direct marking in Itomana. This is captured in the analysis by assuming that one functional head agrees with both arguments. The order of the two Agree operations is then determined by a language-specific parameter, leading to direct marking if subject agreement precedes object agreement and inverse marking if the order is vice versa.

The goal of the paper by **Anke Assmann, Doreen Georgi, Fabian Heck, Gereon Müller, and Philipp Weisser** is to present a new account of the ban on ergative movement in morphologically ergative languages. The underlying assumption is that in such languages Merge/Move applies before Agree, thereby creating new configurations for Agree between the moved ergative argument and the head which triggered movement. This in turn bleeds case Agree between the head and the absolutive argument in the structure, ultimately leading to a case filter violation. In contrast, in accusative systems, where Agree applies prior to Merge, a counter-bleeding relation between both operations holds and no restriction on moving the accusative argument arises.

Closely related to this is the paper by **Doreen Georgi**. Its main claim is that a distinction between intermediate steps of successive cyclic movement on the one hand and movement to a final position (often called criterial movement) on the other hand is necessary. The distinction allows for variable orders of Agree and Move within one language, thus accounting for possible instances of opaque interactions between these two operations: Whereas criterial movement precedes Agree and may hence feed or bleed Agree, intermediate movement takes place after Agree, which may create instances of counter-feeding and counter-bleeding.

The interaction of intermediate movement steps and Agree also plays a role in the paper by **Anke Assmann and Fabian Heck**. It deals with cases of opaque and transparent intervention. On the one hand, there are cases where an antecedent cannot enter into Agree with its associate (a floating quantifier or a parasitic gap) because another potential antecedent intervenes at the surface: transparent intervention. On the other hand, one can observe cases where an intervention effect occurs although no argument intervenes on the surface.

Opaque configurations such as these are then accounted for by assuming that intervention actually takes place at some earlier point of the derivation and is then undone by movement at some later step.

Finally, the paper by **Fabian Heck and Gereon Müller** deals with cases where an interaction between Move and Agree is expected but exceptionally does not show up. It is argued that the ban on moving ergative arguments (assumed to be due to Move preceding Agree) fails to hold in particular cases where movement is procrastinated. This avoids an otherwise expected bleeding of Agree, resulting in counter-bleeding instead. Also, it is proposed that movement of dative arguments in German, which otherwise follows Agree, exceptionally applies early (before Agree) in particular contexts. This gives rise to bleeding of Agree (instead of the expected counter-bleeding relation) and thereby derives a ban against extraction of dative arguments out of ECM complements in German.

References

Aissen, Judith (1999): 'Markedness and Subject Choice in Optimality Theory', *Natural Language and Linguistic Theory* 17, 673–711.

Aissen, Judith (2003): 'Differential Object Marking: Iconicity vs. Economy', *Natural Language and Linguistic Theory* 21, 435–483.

Alderete, John (1997): Dissimilation as Local Conjunction. *In:* K. Kusumoto, ed., *Proceedings of NELS 27*. GLSA, Amherst, Massachusetts, pp. 17–32.

Anderson, Stephen R. (1969): West Scandinavian Vowel Systems and the Ordering of Phonological Rules. PhD thesis, MIT, Cambridge, Massachusetts.

Anderson, Stephen R. (1974): *The Organization of Phonology*. Academic Press, New York.

Brody, Michael (1995): *Lexico-Logical Form: A Radically Minimalist Theory*. MIT Press, Cambridge, Massachusetts.

Brody, Michael (2002): On the Status of Representations and Derivations. *In:* S. D. Epstein and T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 326–342.

Chomsky, Noam (1951): Morphophonemics of Modern Hebrew. Master's thesis, University of Pennsylvania, Philadelphia, Pennsylvania.

Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. MIT Press, Cambridge.

Chomsky, Noam (1981): *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Chomsky, Noam and Morris Halle (1968): *The Sound Pattern of English*. Harper & Row, New York.

Fischer, Silke (2001): On the Integration of Cumulative Effects into Optimality Theory. *In:* G. Müller and W. Sternefeld, eds, *Competition in Syntax*. Mouton de Gruyter, Berlin, pp. 151–173.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag (1985): *Generalized Phrase Structure Grammar*. Blackwell, Oxford.

Hermans, Ben and Mark van Oostendorp, eds (2000): *The Derivational Residue in Phonological Optimality Theory*. Benjamins, Amsterdam.

Kayne, Richard (1975): *French Syntax*. MIT Press, Cambridge, Massachusetts.

Kenstowicz, Michael and Charles Kisseberth (1977): *Topics in Phonological Theory*. Academic Press, NY.

Kenstowicz, Michael and Charles Kisseberth (1979): *Generative Phonology*. Academic Press, New York.

Kiparsky, Paul (1971): Historical Linguistics. *In:* W. O. Dingwall, ed., *A Survey of Linguistic Science*. University of Maryland Linguistics Program, College Park, pp. 576–642.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Kitahara, Hisatsugu (1997): *Elementary Operations and Optimal Derivations*. MIT Press, Cambridge, Massachusetts.

Legendre, Géraldine, Paul Smolensky and Colin Wilson (1998): When is Less More? Faithfulness and Minimal Links in WH-Chains. *In:* P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, eds, *Is the Best Good Enough? Optimality and Competition in Syntax*. MIT Press, pp. 249–289.

Łubowicz, Anna (2005): Locality of Conjunction. *In:* J. Alderete, C. hye Han and A. Kochetov, eds, *Proceedings of WCCFL 24*. Cascadilla Press, Sommerville, Massachusetts, pp. 254–262.

McCarthy, John (2000): Harmonic Serialism and Parallelism. *In:* M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, eds, *Proceedings of NELS 30*. GLSA, Amherst, Massachussets, pp. 501–524.

McCarthy, John (2007): *Hidden Generalizations. Phonological Opacity in Optimality Theory*. Equinox, London.

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

Pollard, Carl and Ivan A. Sag (1994): *Head driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Prince, Alan and Paul Smolensky (1993): *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers. Technical Report.

Prince, Alan and Paul Smolensky (2004): Constraint Interaction in Generative Grammar. *In:* J. J. McCarthy, ed., *Optimality Theory in Phonology: A Reader*. Blackwell, Oxford, pp. 3–71.

Pullum, Geoffrey K. (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.

Řezáč, Milan (2004): Elements of Cyclic Syntax: Agree and Merge. PhD thesis, University of Toronto.

Smolensky, Paul (1995): On the Internal Structure of Con, the Constraint Component of UG. Ms., Johns Hopkins University.

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

Split Ergativity in Subordination

Daniela Thomas[*]

Abstract

The aim of this paper is to derive split ergativity in subordination in the framework of the Minimalist Program. Data from Jacaltec and Sierra Popoluca show that this phenomenon is triggered by aspectlessness. Moreover, the only difference between the split system and the regular ergative system lies in intransitive contexts. To derive these encoding irregularities the *Constraint on case assignment in intransitive contexts* is introduced. It applies to the pre-syntactic numeration and guides the distribution of case features, leading to the assignment of the correct case. Furthermore, the functioning of the new constraint is demonstrated within the framework of optimality theory.

1. Introduction

This paper aims at deriving split ergativity in subordination in the framework of the Minimalist Program. Data illustrating this phenomenon are presented from the languages Sierra Popoluca and Jacaltec. The split in these languages is always triggered by the same condition, which is aspectlessness. In order to derive the irregularities in the encoding system, the analysis of ergativity in Müller (2009) based on Murasugi (1992) is extended with a new principle. This principle guides case assignment in intransitive contexts.

The present analysis is linked to the main topic of "Rule Interaction in Grammar" because Müller's (2009) analysis derives the difference between ergative and accusative encoding system by ordering the elementary operation Merge and Agree. Furthermore, the new *Constraint on case assignment in intransitive contexts* applies to the numeration, i.e., before syntactic structure is built, and eventually leads to the bleeding of absolutive case assignment in the derivation.

---

[*] The research documented here was carried out in the DFG project *Argument encoding in Morphology and Syntax*. I am grateful to Gereon Müller and Doreen Georgi for discussion and intellectual stimulus.

Rule Interaction in Grammar, 1–21
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

The paper is structured as follows: section 2 presents language data that illustrate the phenomenon of split ergativity in subordination; it ends with generalisations about them. Section 3 contains the description of the syntactic basis as well as a new approach to split ergativity. It is completed by an optimality theoretic approach to the deletion of features in the numeration. Furthermore, an alternative theory is presented and discussed. Finally, in section 4 I draw some conclusions and mention directions for further research.

2. Language Data

In this section, data from the two unrelated languages Sierra Popoluca and Jacaltec are presented. The examples will illustrate the phenomenon of split ergativity in subordination.

2.1. Sierra Popoluca

Sierra Popoluca, a Mixe-Zoquean language spoken in Veracruz, Mexico, exhibits argument encoding via verbal agreement that is driven by a hierarchical system. The hierarchy ranks speech act participants (first and second person) above third person. That means that in transitive cases only the argument ranked higher on the hierarchy is realised on the verb. If a relation between two speech act participants is expressed, a special set of markers, the local set, is used.

The encoding system is an ergative one, marking the single argument of an intransitive verb and the internal argument of a transitive verb in the same way and the external argument of a transitive verb with a different set of markers. This can be seen in (1)[1]:

(1) a. sɨʔɨp ta=wiʔk-pa=ʔam
now 1INCL.ABS=eat-INC=ALR
‘Now we eat.’

---

[1] The following abberviations are used: 1/2/3 - first/second/third person, ABS - absolutive, ALR - ‘already’, ASP - aspect, C - consonant, CL - classifier, CMP - completive, DEP_{ib} - dependent intransitive type b, DEP_{t} - dependent transitive, ERG - ergative, etc. - et cetera, EXCL - exclusive, FUT - future tense, INC - incompletive, INCL - inclusive, JUST - ‘just’, NEG - negation, PERF - perfective, PLU_{sap} - plural, speech act participant, PRO - pronoun, PSR - possessor, Spec - specifier of, SUFF - suffix, V - vowel, ‘-’ - morpheme boundary, ‘.’ - morpheme boundary in lexicalised expression, ‘=’ - clitic boundary

b. ta=kuʔt-pa
1INCL.ABS=eat-INC
‘It eats us.’

c. ʔɨch ʔan=kuʔt-pa jeʔm saapnyi
1PRO 1EXCL.ERG=eat-INC that banana
‘I ate this banana.’

d. jemɨ=ʔam ʔan=ʔix-pa
there=ALR 2:1=see-INC
‘There you see me.’ (de Jong Boudreault 2009: 337, 335, 401)

Example (1-a) shows an intransitive verb with its single argument in first person plural inclusive. It is encoded by *ta=*. In (1-b) *ta=* shows up again. This time it is the marker for the internal argument of the transitive verb. The argument is in the first person and outranks the third person external argument on the person hierarchy. Since Sierra Popoluca always realises only one argument, the higher ranked internal argument is encoded. Example (1-c) shows another transitive clause, but this time the external argument outranks the internal argument (1:3). Therefore the external argument is marked on the verb with *ʔan=*.

The full paradigm of agreement markers is shown in (2) (see de Jong Boudreault 2009: 396, Elson 1960: 207):[2]

(2) *Agreement markers in Sierra Popoluca:*

	Set A/Ergative	Set B/Absolutive	Set C/Local
1.EXCL	ʔan=	ʔa=	
1.INCL	tan=	ta=	
2	ʔin=	mi=	
3	ʔi=	∅=	
2:1			ʔan=
1:2			man=

Looking at the subordinate clause (in brackets) in (3-a) with an intransitive verb, one can observe that instead of the expected absolutive marker the ergative marker *ʔan=* emerges. In the examples (3-b) and (3-c) a transitive verb is embedded. In these cases we find the anticipated markings: ergative *ʔin=* for the external argument and absolutive *ʔa=* for the internal argument. This

---

[2] The notation ‘x:y’ means subject person x acts on object person y.

yields an accusative pattern since here the single argument of the intransitive verb and the external argument of the transitive verb are treated in the same way (ergative marking) while the internal argument of the transitive verb receives a different treatment (absolutive marking), cf. de Jong Boudreault (2009: 419, 726, 727).

(3) a. dya ʔa=joʔy-neʔ-W=ʔam
NEG 1EXCL.ABS=be.angry-PERF-CMP=ALR
[ʔan=put-W3]
[1EXCL.ERG=exit-DEP$_{ib}$]
'I wasn't angry when I left.'
b. mich dya=ʔam mi=ʔoy-W [ʔin=ʔaʔm-W$_2$
2PRO NEG=ALR 2ABS=go/return-CMP [2ERG=see-DEP$_t$
ʔin=choomo]
2PSR=grandmother]
'You didn't go see your grandmother.'
c. ʔoy=tyi=ʔam [ʔa=ʔaʔm-taʔm-W$_2$]
go/return$_{aux}$=JUST=ALR [1.EXCL.ABS=see-PLU$_{sap}$-DEP$_t$]
'They just went to see me.' (de Jong Boudreault 2009: 419,726,727)

These irregularities in argument encoding are triggered in certain multi-verb constructions (de Jong Boudreault 2009, Marlett 1986, Elson 1960):

(4)
(i) temporal adverbial clauses, which are not introduced by a Spanish adverbial
(ii) embedded clauses with the subordinators ∅, *mo*, *=mu*
(iii) multi-verb constructions with the progressive auxiliary *sɨʔ*
(iv) multi-verb constructions with Type II auxiliaries[3]
(v) multi-verb constructions with Type I auxiliaries where the embedded verb is in passive voice

However, the embedded verbs in Sierra Popoluca display further special characteristics: in the constructions mentioned above, the verbs lack any kind of mood or aspectual marking, but receives dependent morphology instead. In

---

[3] The distinction between Type I and Type II auxiliaries is based on the pattern that emerges when an auxiliary combines with a verb. With Type II auxiliaries an accusative pattern emerges. With Type I auxiliaries an accusative pattern emerges only if the dependent verb is in the passive (cf. de Jong Boudreault 2009).

(5) I contrast the matrix verb and the embedded verb of (3-a) in order to illustrate this (de Jong Boudreault 2009: 419).

(5) a. ʔa=joʔy-neʔ-W=ʔam
1EXCL.ABS=be.angry-PERF-CMP=ALR
b. ʔan=put-W3
1EXCL.ERG=exit-DEP$_{ib}$

In Sierra Popoluca a (matrix) verb consists at least of the verbal root, a person proclitic and markings for aspect and/or mood. But as we have seen, this is not the case with the embedded verbs of multi-verb constructions, in which split ergativity is triggered. This view is in line with Boudreault saying that these verbs are best described as aspectless. The dependent marking consists of an inaudible consonant represented in the glosses by -*W*.[4,5] Despite its inaudibility this consonant has effects on the assignment of stress (cf. de Jong Boudreault 2009).

2.2. Jacaltec

Jacaltec, a Mayan language spoken in Guatemala, also basically instantiates an ergative system of agreement:

(6) a. ch-ach hin-mak-a'
ASP-2ABS 1ERG-hit-FUT
'I will hit you'
b. xc-ach toyi
ASP-2ABS go
'You went.' (Craig 1977: 119, 333)

In contrast to Sierra Popoluca, Jacaltec realises both arguments of a transitive verb via person marking on the verb. This can be seen in (6-a): ergative *hin-* for

---

[4] According to Boudreault, intransitive verbs embedded under Type I auxiliaries that are not in passive voice also receive dependent marking, which simply differs from the marking used in the other constructions (-*i* instead of -*W*). My impression is that these embedded clauses are actually nominalisations, since -*i* is also a nominaliser and those verbs are inflected for plural with the nominal plural marker. This also explains why in auxiliary-I-constructions with the dependent marker -*i* no split-ergativity arises.

[5] In the glosses, -*W* is further differentiated in -*W*2 for transitive verbs and -*W*3 for intransitive verbs.

the first person external argument and absolutive *-ach* for the second person internal argument. On the intransitive verb in (6-b) we find the same marker encoding the second person single argument as with the internal argument of the transitive verb in (6-a). Absolutive markers can either be unbound or clitics on the aspect words. Ergative markers always precede the verbal stem. A full paradigm of agreement markers is given in (7), cf. Craig (1977).

(7) *Agreement markers in Jacaltec:*

Person/ Number	Absolutive	Ergative	
		C-initial verbal root	V-initial verbal root
1 Sg	hin	hin-	w-
2 Sg	hach	ha-	haw-
3 Sg	∅+CL	s-+CL	y-+CL
1 Pl	hoñ	cu-/co-	y-
2 Pl	hex	he-	hey-
3 Pl	∅+PL+CL	s-+PL+CL	y-+PL+CL

Jacaltec shares with Sierra Popoluca the property that in some subordinate clauses split ergativity arises. (8) is an example of subordination without an overt complementizer where in (8-a) a transitive and in (8-b) an intransitive verb is embedded.

(8) a. x-∅-w-ilwe [hach hin-col-ni]
ASP-3ABS-1ERG-try [2ABS.PRO 1ERG-help-SUFF]
'I tried to help you.'
b. x-∅-w-il [ha-cañalwi]
ASP-3ABS-1ERG-see [2ERG-dance]
'I saw you dance.' (Craig 1977: 115f.)

The transitive context shows no deviations; both arguments are encoded as in the matrix clauses: the external argument receives ergative marking and the internal argument is realised by an absolutive marker. However, in the intransitive case the absolutive marker is expected to encode the sole argument, but the ergative marker emerges. Again, this yields an accusative pattern: the ergative marking shows the typical nominative distribution by encoding the sole argument of an intransitive verb and the external argument of a transitive verb, while the internal argument is realised by absolutive marking. In Jacaltec this phenomenon is triggered in the following constructions:

(9) (i) aspectless complement clauses
(ii) aspectless temporal adverbial clauses (Craig 1977)

It is obvious that the factor of aspectless verbs plays an important role here. Other subordinate clauses with embedded verbs bearing aspectual marking do not show split ergativity: (10) is an example of subordination with the *tato*-complementizer, where the embedded intransitive verb exhibits the expected absolutive marker for the sole argument.

(10)	x-∅-aw-abe	[tato	ch-in	to-j	hecal]
	ASP-3ABS-2ERG-hear	[that	ASP-1ABS	go-FUT	tomorrow]

‘You heard that I will go tomorrow.’ (Craig 1977: 232)

2.3. Generalisations

Two unrelated languages which both exhibit an ergative encoding system and split ergativity in certain subordinate clauses have been presented. Their split ergativity in subordination is never triggered by the subordination itself but is influenced by other factors. This has already been observed by Dixon (1994). In Sierra Popoluca and Jacaltec this factor is aspectlessness.

Crucially, there is no genuine change from an ergative to an accusative encoding system in these embedded clauses; rather, the domain of the ergative marker is extended. The only difference between the split and the normal ergative system occurs in intransitive contexts: instead of the expected absolutive, we find ergative marking. This leads to a pseudo-accusative system where the ergative marker has a distribution like the nominative marker in accusative systems. I will come back to this at the end of section 3.2.

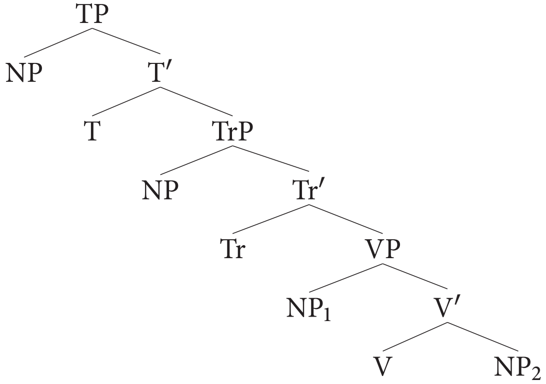
3. Analysis

As a background theory on ergativity I will adopt the analysis of ergative encoding systems by Murasugi (1992) and its reconstruction by Müller (2009). After introducing these theories, I will come to the details of my analysis of split ergativity.

3.1. Syntactic Basis: Ergativity vs. Accusativity in Murasugi (1992) and Müller (2009)

Murasugi (1992) proposes an approach towards deriving the accusative and ergative encoding system in one and the same syntactic structure. In that account, she assumes the following sentence structure:

(11) *Sentence structure in Murasugi (1992)*



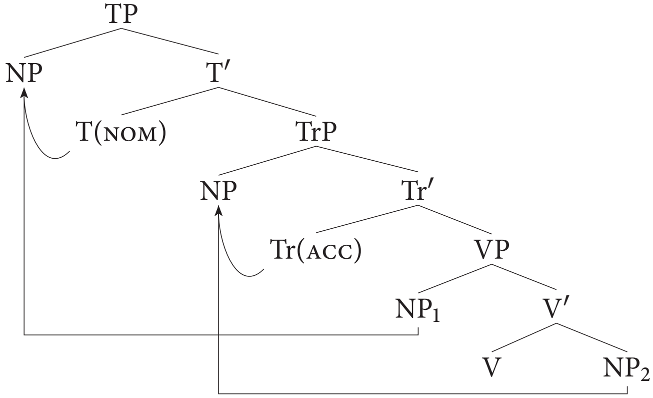
The core idea is that intransitive contexts work similarly in both systems, whereas transitive contexts differ. TP and Tr(ansitivity)P are the functional projections of the heads that, among other things, assign case. In this system T assigns the unmarked case, i.e., absolutive in the ergative system and nominative in the accusative system. Tr assigns the marked case, i.e., accusative and ergative. Case is assigned when the NP moves to the specifier of the functional head.

Moreover, Murasugi (1992) establishes three economy principles (cf. Murasugi 1992: 24):

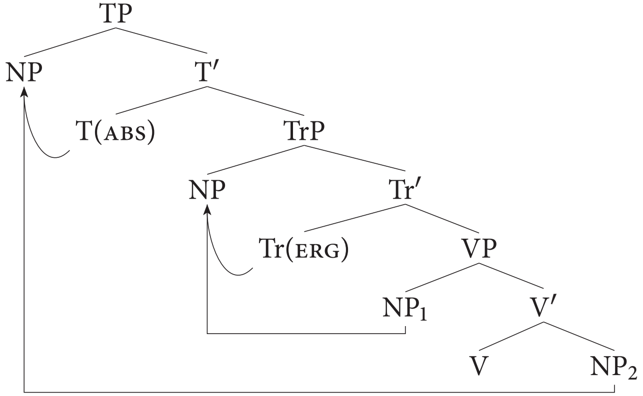
(i) **Closest Available Source:** At each level of a derivation, a target must take the closest available source NP.
(ii) **Closest Featured Target:** At each level of a derivation, a source NP must move to the closest featured target.
(iii) **Procrastinate:** An operation must be done as late as possible.

The parameter that eventually leads to the difference of the two systems is the strength of the features of T and Tr. The strong features need to be checked at S-structure and therefore require overt NP movement. The features of the remaining functional head, which are not strong, will be checked on LF and hence do not require overt movement. In the accusative system T is strong and therefore the closest NP, which is the external NP, moves to Spec-T and the internal NP moves to Spec-Tr. In the ergative system Tr's features are strong causing the closest NP – again the external NP – to move to its specificier position. The internal NP can now only move to Spec-T.[6] This results in crossing paths in the accusative system (see (12)) and nested paths in the ergative system (see (13)).

(12) *Movement in the accusative system:*



(13) *Movement in the ergative system:*



---

[6] In contrast to the ergative system, the movements in the accusative system happen overtly.

Murasugi argues that Tr has no case features in intransitive contexts and hence only movement to Spec-T is possible for case assignment. As already mentioned, T assigns the unmarked case. Hence according to this system nominative or absolutive encode the sole argument.

Müller (2009) develops a reconstruction of Murasugi's system. The background assumption is that syntactic structure is built bottom-up, incrementally, by the operations Merge and Agree. Müller defines Merge and Agree as follows (cf. Müller 2009: 273):

(14) *Merge:*
α can be merged with β, forming a projection of α, if α bears a subcategorization feature [+F+] and F is the label of β.

(15) *Agree:*
α can agree with β with respect to a feature bundle Γ, if a., b., and c. hold:
  a. α bears a probe feature [$*$F$*$] in Γ, β bears a matching goal feature [F] in Γ.
  b. α m-commands β.
  c. There is no γ, such that (i) and (ii) hold:
    (i) γ is closer to α than β.
    (ii) γ bears a feature [F] that has not yet participated in Agree.

Especially (15-c-i) is important in what follows. It is based on a definition of Closeness from which it follows that '[. . .] the specifier of a head is closer to the head than a category that is further embedded in the complement of the head' (Heck and Müller 2007: 174).

With respect to the syntactic structure, Tr is replaced by v. This head also introduces the external argument. Furthermore this system also includes the numeration – a pre-syntactic collection of all lexical items that are to be used in the derivation.

In Müller's (2009) system, Agree is responsible for case assignment by checking the case features of the functional head and the DP under identity. The operation is triggered by features on T and v which act as probes.[7] This

---

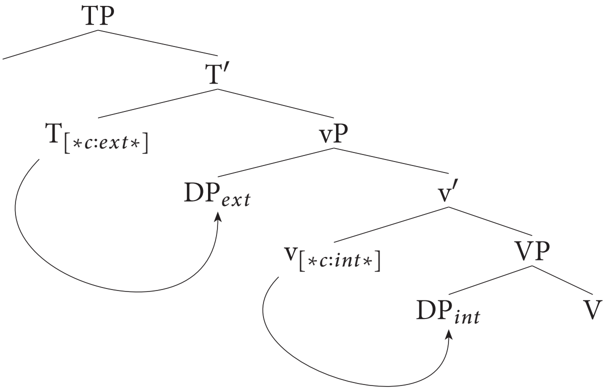
[7] Müller argues that case and agreement are basically the same. The only difference between case and agreement is the locus of the morphological reflex of the Agree relation in [CASE] (head marking vs. dependent marking): if this feature is spelled out on the functional head, it results in agreement; if, however, it is realised on the DP, it results in case marking. Languages

point needs some further explanation. Müller assumes that there is merely one case feature that can bear two values: [CASE:ext] or [CASE:int]. The corresponding probe features are localised on T (external case) and v (internal case). Agree can proceed independently of the feature values of the functional heads and the arguments. But the derivation will crash if the feature values of the functional heads and the DPs do not coincide.

The central part of this theory is v. It has a special role since it does not just assign case but also introduces the external argument. Hence, it participates in both elementary operations (Merge and Agree). Crucially, when v is merged, the context for the application of both operations is created. Assuming that they cannot proceed simultaneously, one needs to be carried out before the other. The idea is that the solution of this conflict is language-specific. It is a language-specific choice whether Agree takes priority over Merge or vice versa. This ordering of the elementary operations replaces Murasugi's feature strength and is thus responsible for the emergence of the accusative vs. ergative system in transitive contexts.

If Agree has priority over Merge, the internal argument will be assigned the internal case from v since the internal argument is the only potential goal at this stage of the derivation. Subsequently, the external argument will be introduced, which then receives the external case from T. This yields the accusative system:

(16) *Agree before Merge – accusative system:*

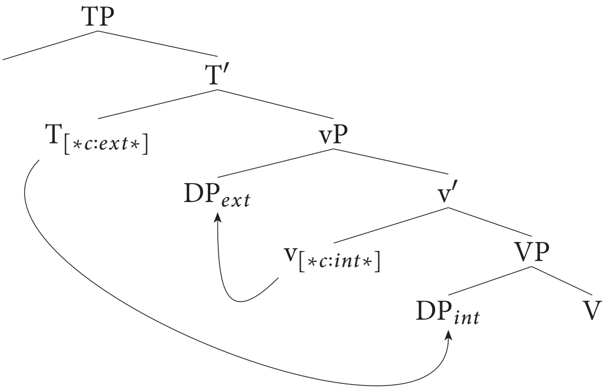


---

that have both case and agreement, and that employ different patterns in the two areas, are addressed by postulating additional Agree operations that do not involve [CASE]. I adopt this general view.

The second possibility is that Merge applies prior to Agree. In that case, the external argument is introduced before v undergoes Agree. Therefore, it is closer to v than the internal argument, according to the definition of Closeness (see (15-c-i)). Hence, Agree takes place between v and the external argument which therefore receives the internal case. When T is merged, it assigns the external case to DP_{int}. This results in the ergative system.

(17) *Merge before Agree – ergative system:*



All of this holds in transitive contexts. For intransitives more needs to be said. First of all it needs to be ensured that there are only as many case features on functional heads as arguments. Otherwise, the derivation would crash because of unchecked features. This is accomplished by Müller's Feature Balance criterion, which applies to the numeration:

(18) *Feature Balance:*
For every feature specification [∗F:α∗], there must be a matching feature specification [F:α]. (Müller 2009: 279)

As a consequence of this criterion, either T or v has to lose its case feature in an intransitive context and with that the ability to assign case. But how is it determined which functional head maintains its case feature in the numeration? Müller suggests that this is decided by means of unmarkedness. He therefore invokes that the external case is the syntactically as well as morphologically unmarked case. Hence, the unmarked [∗case:ext∗] on T remains and [∗case:int∗] will not appear on v. The result is that the sole argument of an intransitive verb bears external case, just like the external argument of a transitive verb, if Agree takes priority over Merge (=accusative system, see

(16)), or the internal argument of a transitive verb, if Merge takes priority over Agree (=ergative system, see (17)).

With the background on ergativity in place, we can now have a closer look at split ergativity in subordination contexts.

3.2. Minimalist Analysis of Split Ergativity

What needs to be done now is to extend the existing analyses, such that they can cope with clause-type based split ergativity in intransitive contexts. This means a mechanism has to be created which leads to the assignment of the ergative case instead of the absolutive case in aspectless intransitive subordinate clauses. At the same time the absolutive case has to be preserved in transitive aspectless subordinate clauses.

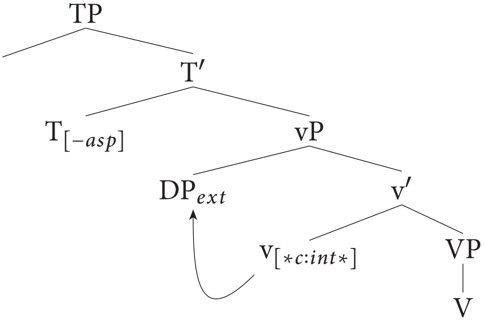
I adopt Müller's approach almost completely. I agree with him regarding transitive contexts, but concerning intransitive contexts I will only adopt the Feature Balance criterion, since this is the critical point for split ergativity in subordination.

Furthermore, I assume that case features are lexical properties of the functional heads v and T that can be deleted in the numeration. As I have already pointed out in section 2.3, the only difference between split ergativity and the ordinary ergative system is that the ergative (internal case, assigned by v) instead of the expected absolutive (external case, assigned by T) shows up in intransitive contexts. Therefore, I suggest a new constraint that determines which functional head keeps its case feature in an intransitive context:

(19) *Constraint on case assignment in intransitive contexts:*
In intransitive contexts the case feature of the structurally highest, non-defective functional head remains.

First of all this is a constraint that just like Feature Balance applies to the numeration. As a consequence, the number of possible numerations is reduced and derivations which would crash anyway are prevented from the very beginning. Now, the numeration is pre-syntactic and therefore contains no structure. Nevertheless, intransitive contexts can be detected at this stage already by counting the number of elements of the category D, which are representatives of future arguments. If the numeration of a clause contains two D elements, a transitive context is given; if it contains only one D element, the clause will be intransitive. Furthermore, since Merge is triggered by selection features [+F+], it is already foreseeable that T (bearing [+v+]) will be merged above v. Hence, T will be the highest functional head in the structure. But there is a second requirement in the constraint that demands that the highest head must not be defective. Recall from the generalisations that no split ergativity in subordination is triggered by subordination itself; rather, it is subject to other factors. For Sierra Popoluca and Jacaltec this feature is aspectlessness. Since aspect is assigned by T to v, the embedded Ts obviously lack these features, for no aspect is assigned. Usually T has to bear these features in these languages. De Jong Boudreault (2009) states that a (matrix) verb in Sierra Popoluca consists at least of a verbal root with person marking as well as aspect (or mood) marking. This seems to hold for Jacaltec, too. We may speculate that this is a general condition for T heads. The embedded aspectless T heads violate this condition. They are thus defective and because of the principle in (19) the case feature cannot remain on T. Now v becomes the highest non-defective functional head. Consequently, [∗case:int∗] remains on v and the sole argument of the intransitive verb is assigned the external case (= ergative) by v.

(20) *Intransitive context with aspectless T:*



It follows that there is no change in encoding systems in split ergativity: the unexpected ergative case in intransitive contexts is assigned by v (= internal case), while in the accusative system nominative (= external case) is assigned in intransitive contexts. Thus, what happens is a change to an accusative pattern, but not a deep change from ergativity to accusativity in the system: to carry out a complete change of system, cases would have to be swapped in transitive contexts.

Moreover, there seems to be no way for v to be defective in a converging derivation, for the single feature it necessarily has to posses is [+V+].[8] If v does not bear this feature, no Merge of v and VP is possible and the derivation crashes anyway. Hence, v will never be defective and remains as an alternative for assigning case.

Note that the constraint in (19) merely expresses a preference of non-defective heads over defective ones. This does not mean that defective heads generally cannot bear case features. In transitive contexts, when no choice is to be made between the two heads, T does assign the external case, although it might be aspectless. In transitive contexts case assignment works as described in section 3.1.

However, a structure without aspectual information cannot survive. This problem is solved by subordination. The appropriate complementizer of the embedded CPs is a reflex of the defective T. It must necessarily be subordinating in order to ensure that the missing aspectual information can be obtained from the matrix verb. It could either be assumed that this complementizer bears a special selection feature [+$T_{[-asp]}$+] or that a defective T receives a diacritic '_ ' and that the selection feature is thus [+T_ +].

In languages which do not exhibit split ergativity in subordination due to the lack of aspectual features, it must be assumed that these features do not belong to the inventory of necessary features. Their absence is of no importance and thus does not create defectiveness. T therefore remains the highest non-defective functional head and is allowed to assign the unmarked case.

3.3. Optimality in the Numeration

Constraint (19) is rather complex and its functioning might be best illustrated within the framework of Optimality Theory. To this end, the constraint will now be split up into three constraints (plus Feature Balance).[9]

---

[8] Note that the notion of defectiveness used here differs from the common notion (cf. Chomsky 2001, 2005).

[9] The constraint *U-CASE/[–ASP,–TRANS] is introduced specifically for the languages discussed in this paper and is too specific to derive split ergativity in general. To achieve this, further research regarding languages with this phenomenon is needed, for there might be other factors besides lack of aspectual features triggering split ergativity in subordination. After finding out the common denominator of the triggers, the context in which the unmarked case is prohibited can be determined. But the constraint presented here might suffice for my purpose.

(21) a. Feature Balance
For every feature specification [∗F:α∗], there must be a matching feature specification [F:α].

b. *No Case
Case has to be assigned.

c. *u-Case/[−asp,−trans]:
In aspectless intransitive contexts, T must not bear [∗case:ext∗].

d. Faith/u-Case:
Preserve the unmarked case ([∗case:ext∗] on T).

The constraint ranking is as follows:

(22) Feature Balance ≫ *No Case ≫ *u-Case/[−asp,−trans] ≫ Faith/u-Case

These constraints apply to the numeration, so the lexical items with their features serve as input for the optimization procedure. I reduce them to a minimum. Since the numeration does not contain complete DPs, one can only infer the number of arguments from the number of D elements. Therefore, elements with the category D serve as representatives of the arguments here. The tableau T_1 in (23) shows an intransitive numeration with an aspectless T.

(23) *T_1: Optimization of numeration with aspectless T; 'split ranking':*

Input: T[−asp],[∗case:ext∗], v[∗case:int∗], D	F-Balance	*No Case	*u-Case	Faith Case
C_1: T[−asp],[∗case:ext∗], v, D[case:ext]			*!	
☞C_2: T[−asp], v[∗case:int∗], D[case:int]				*
C_3: T[−asp],[∗case:ext∗], v[∗case:int∗], D[case:ext]	*!		*	
C_4: T[−asp],[∗case:ext∗], v[∗case:int∗], D[case:int]	*!		*	
C_5: T[−asp], v, D		*!		*

The third and fourth candidates cannot win because both functional heads bear a case feature. That means there is one argument but two cases to be assigned. Therefore, Feature Balance is violated. The last candidate is ruled out because there are no case features at all. This violates *No Case. Finally, the first candidate cannot be the winner, for the defective T bears a case feature. This causes a violation of *u-Case/[−asp,−trans]. Hence, the second candidate with v as the only case assigner in an intransitive aspectless context becomes optimal.

In a structure with aspect on the T head, the expected external case turns out as the winner of the competition, see tableau T_2 in (24). *u-Case/[–asp,–trans] does not rule out the first candidate with the T head as the only case assigner in an intransitive context anymore, for its context is not given.

(24) *T_2: Optimization of numeration with T bearing aspect; 'split ranking':*

Input: T[*case:ext*], v[*case:int*], D	F-Balance	*No Case	*u-Case	Faith Case
☞C_1: T[+asp],[*case:ext*], v, D[case:ext]				
C_2: T[+asp], v[*case:int*], D[case:int]				*!
C_3: T[+asp],[*case:ext*], v[*case:int], D[case:ext]	*!			
C_4: T[+asp],[*case:ext*], v[*case:int], D[case:int]	*!			
C_5: T[+asp], v, D		*!		*

For languages without split ergativity in subordination, it can be assumed that the absence of aspectual features (or another factor) does not create a defective head. As a consequence, the constraints have to be reranked:

(25) Feature Balance ≫ *No Case ≫ Faith/u-Case ≫ *u-Case/[–asp, –trans]

With the same input as in T_1 the competition ends with a different optimal candidate, namely the one in which the T head assigns the unmarked case to the sole argument of an intransitive verb, see T_3 in (26).

(26) *T_3: Optimization of numeration with T bearing aspect; no split:*

Input: T[–asp],[*case:ext*], v[*case:int*], D	F-Balance	*No Case	Faith Case	*u-Case
☞C_1: T[–asp],[*case:ext*], v, D[case:ext]				*
C_2: T[–asp], v[*case:int*], D[case:int]			*!	
C_3: T[–asp],[*case:ext*], v[*case:int*], D[case:ext]	*!			*
C_4: T[–asp],[*case:ext*], v[*case:int*], D[case:int]	*!			*
C_5: T[–asp], v, D		*!	*	

Since Faith/u-Case is ranked above *u-Case/[−asp,−trans], the candidate whose D bears the unmarked case (absolutive) will always win, independently of whether T bears aspect in intransitive contexts or not.

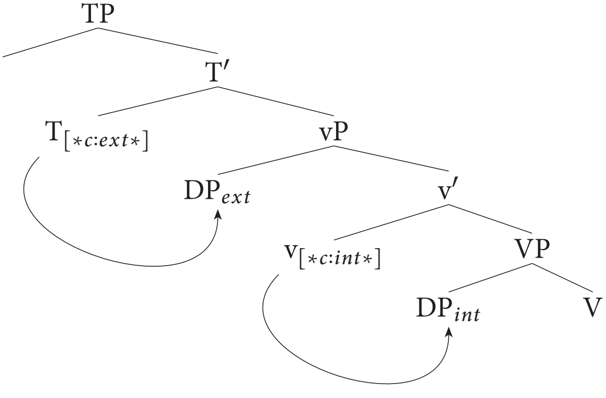
3.4. An Alternative Approach

In what follows I would like to go into the proposal by Bobaljik (1993) regarding the derivation of accusative and ergative systems and why it has problems with the phenomenon of split ergativity. For the purpose of facilitating the comparison, the clause structure in (16) is anachronistically adopted.

Bobaljik (1993) suggests that there are no differences between ergative and accusative systems in transitive contexts; nominative and ergative are practically the same case, likewise accusative and absolutive. This is the exact opposite of Murasugi's (1992) and Müller's (2009) theory, where the ergative and accusative system work the same in intransitive contexts but differ in transitive contexts.

With respect to clause structure, Bobaljik assumes additional AgrPs above VP and TP (v does not exist) and movements of the NPs into their specifier-position. This concept is here replaced by Agree relations between the functional heads T and v and the arguments. Taking the terms 'external case' now for nominative and ergative and 'internal case' for absolutive and accusative, the structure looks as follows:

(27) *Transitive contexts in both systems:*



In order to determine which case is assigned in intransitive contexts, Bobaljik proposes the 'Obligatory Case Parameter' (cf. Bobaljik 1993: 50):

(28) *Obligatory Case*
Case X is obligatorily assigned/checked.

(29) *Obligatory Case Parameter*
 a. In nominative/accusative languages, case X is nominative (= erg).
 b. In ergative/absolutive languages, case X is absolutive (= acc).

As a result, the difference between ergative and accusative systems arises in intransitive contexts. The cases that show up in these contexts originate on different functional heads: absolutive is assigned by v and nominative by T.

This strict principle makes it impossible to explain why we sometimes find the ergative marking in intransitive contexts, for the absolutive has to be assigned obligatorily. Consequently, it would be necessary to alter the 'Obligatory Case Parameter' or replace it by a new principle. The difficulty there is that depending on the system a different functional head is active in intransitive contexts. Hence, a factor must be found that associates T and v, but ensures that depending on the system the right functional head assigns case. This problem might be solved with the help of the idea of unmarkedness of cases.

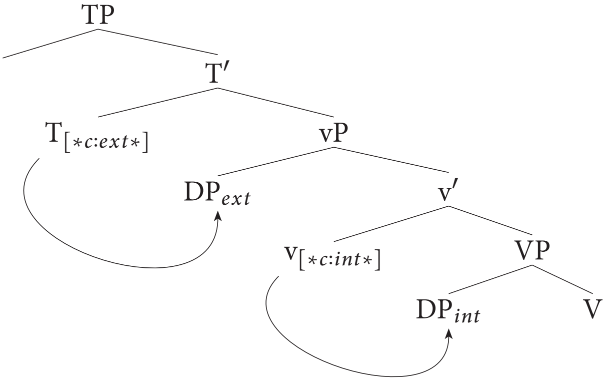
The task is to explain why v loses its case feature in intransitive contexts in the languages presented in section 2. The most obvious solution would be to assume that an aspectless (defective) T attracts case features, which stands in complete contrast to the constraint in (19). This would actually result in a true change of the encoding system: as already mentioned, the difference between ergative and accusative systems in Bobaljik's theory lies in intransitive contexts and would be neutralised in this special case.

4. Conclusion and Outlook

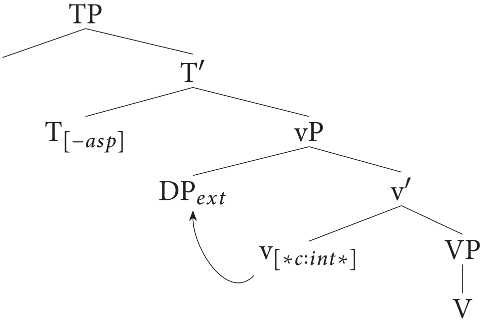
In this paper I have shown how the phenomenon of split ergativity in subordination can be derived in minimalist syntax. Based on the analysis of ergativity in Murasugi (1992) and its reconstruction by Müller (2009), I proposed a 'Constraint on case assignment in intransitive contexts'. This constraint derives split ergativity in intransitive contexts in the languages Sierra Popoluca and Jacaltec. In addition, the language Päri, a western Nilotic language, also exhibits split ergativity in certain subordinate clauses and seems to fit into this analysis, too. But further work is needed to confirm this hypothesis.

Interestingly, the present analysis does not only predict split ergativity but also an analogous split accusativity. This means that there can be languages which basically exhibit an accusative system (for case assignment in transitive cases see (16), repeated in (30)), but show accusative marking instead of nominative marking in certain intransitive (subordination) contexts, where T is defective (see (31)). Hence, the accusative would receive the typical absolutive distribution and create an ergative pattern.

(30) *Accusative system: transitive context*



(31) *Accusative system: intransitive context with defective T*



The next task then is to find either languages which instantiate exactly this prediction or independent parameters that may explain why no defective Ts appear in accusative languages and hence no 'split accusativity' can be triggered.

Another result of this paper is that a strict distinction between deep encoding systems and surface-oriented encoding patterns has to be made. To change from an ergative to an accusative encoding pattern there are two possibilities: either the cases in the transitive contexts are swapped or the case of the external argument of a transitive verb is used for the sole argument of an intransitive verb (ergative). In the same way, the change from an accusative to an ergative system can be achieved. But that is different with encoding systems. According to the present analysis the differences between basic encoding systems are interpreted differently: in Murasugi (1992) both systems behave identically in intransitive contexts, but differ in transitive cases. Bobaljik (1993), on the contrary, assumes that the systems work equivalently in transitive contexts but different functional heads are at work with intransitive verbs. But still there is only one parameter in each of the theories whose setting must be changed in order to achieve a change of system.

References

Bobaljik, Jonathan (1993): Ergativity and Ergative Unergatives. In: *Papers on Case and Agreement II*. Mit Working Papers In Linguistics, MIT:Cambridge, Mass., pp. 45–88.

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): *On Phases*. MIT, Cambridge, Mass.

Craig, Colette Grinevald (1977): *The structure of Jacaltec*. University of Texas Press, Austin and London.

de Jong Boudreault, Lynda J. (2009): A Grammar of Sierra Popoluca (Soteapanec, a Mixe-Zoquean Language). PhD thesis, University of Texas, Austin.

Dixon, Robert M. W. (1994): *Ergativity*. Cambridge University Press, Cambridge.

Elson, Ben (1960): 'Sierra Popoluca Morphology', *International Journal of American Linguistics* **26**(3), 206–223.

Heck, Fabian and Gereon Müller (2007): Extremely local optimization. In: *Proceedings of the 26th WECOL*. California State University, Fresno, pp. 170–183.

Marlett, Stephen A. (1986): 'Syntactic Levels and Multiattachment in Sierra Popoluca', *International Journal of American Linguistics* **52**(4), 359–387.

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. *In:* K. K. Grohmann, ed., *Explorations of Phase Theory: Features and Arguments*. Mouton de Gruyter, Berlin, pp. 269–308.

Murasugi, Kumiko (1992): Crossing and Nested Paths. PhD thesis, MIT, Cambridge.

Possessor Case in Udmurt: Multiple Case Assignment Feeds Postsyntactic Fusion

Anke Assmann, Svetlana Edygarova, Doreen Georgi,
Timo Klein & Philipp Weisser*

Abstract

In this paper we investigate the case split on the possessor in Udmurt: traditionally, the choice between ablative and genitive possessor case is said to be driven by the grammatical function (GF) of the XP containing the possessor. Given this generalization, the case alternation in Udmurt seems to require look-ahead or counter-cyclic operations, which constitutes a problem for a strictly derivational model of syntax. We argue that the case split is not driven by GFs; rather, it is determined by the case value of the XP that contains the possessor. Under the new generalization, a local reanalysis of the case split is possible. We present a case stacking analysis according to which the possessor always bears genitive but may be assigned another structural case by an external head. Stacking of genitive and a semantic case is *bled* due to the different complexity of semantic and structural cases, a restriction on the number of case slots and the different timing of case assignment from DP-internal vs. DP-external heads that is due to cyclicity. If case stacking applies in the syntax, it *feeds* fusion of the structural case values into a single case feature set in the postsyntactic morphological component. If accusative and genitive stack on the possessor, only the default semantic case marker (the ablative marker) can realize the resulting feature set. In any other context the genitive marker is chosen. We thus claim that there is no abstract ablative on the possessor; instead, the morphological ablative marker realizes a combination of two abstract structural cases.

---

* This research is supported by the DFG (Deutsche Forschungsgemeinschaft) within the project "Lokale Modellierung nicht-lokaler Abhängigkeiten in der Syntax" (Local modeling of non-local dependencies in syntax). For valuable comments we would like to thank the participants of the syntax colloquium at the University of Leipzig as well as the audiences at BCGL 7 (Brussels, December 2012), ConSOLE XXI (Potsdam, January 2013), and GLOW 36 (Lund, April 2013). Particularly, we would like to thank Ekaterina Georgieva (University of Szeged) for putting us in contact with Svetlana Edygarova.

Rule Interaction in Grammar, 23–63
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

1. Introduction

In Udmurt, a Uralic language spoken in the Udmurt Republic in Russia, possession can be expressed in a number of ways, with minor semantic differences (see Edygarova 2009 for an overview). The primary possessive construction, which is the focus of this paper, has the following properties: (i) the possessor precedes the possessum, (ii) the possessor bears a case suffix and (iii) the possessum agrees in person and number with the possessor, indicated by a suffix on the possessum (cf. (1)). The interesting fact about this construction is that the case of the possessor varies between genitive and ablative, that is, Udmurt exhibits a case split on the possessor. The possessor cases are in complementary distribution. According to the literature, the grammatical function (GF) of the possessum determines which case value the possessor bears: genitive is the default possessor case; ablative occurs if the DP containing the possessor functions as a direct object (Csúcs 1988, Vilkuna 1997, Kel'makov 1999, Winkler 2001, Nikolaeva 2002, Koptjevskaja-Tamm 2003, Suihkonen 2005, Edygarova 2009). In example (1-a) the DP that contains the possessor is the subject of the verb *ugnï* 'to dress' and the possessor can thus only bear genitive; in (1-b), however, the DP that contains the possessor functions as the direct object of the transitive verb 'to see' and the possessor must bear ablative.

(1) *Possessor cases in Udmurt*:
a. so-len/*leš anaj-ez siče ug diśaśki
he-GEN/ABL mother-3SG such NEG.PRES.3 dress
'His mother does not dress such a way.' (Edygarova 2009: 105)
b. so-leš/*-len eš-s-e ažži-śko
he-ABL/GEN friend-3SG-ACC see-PRES.1SG
'I see his friend.' (Edygarova 2009: 101)

Under the assumption that the GF of a DP is determined by its location in the syntactic structure, the case value of the possessor in Udmurt seems to depend on the nature of the external head that selects the DP containing the possessor.

Similar case splits on the possessor exist in other languages as well, e.g. in Bezhta (Daghestanian): the possessor bears the so-called *direct genitive* if the entire DP is assigned nominative case; if the DP bears any other case, the possessor gets the so-called *oblique genitive* (cf. (2)).[1] Assuming that case is assigned by functional heads to certain positions in the tree, the position of the DP containing the possessor matters for the case of the possessor in Bezhta as well, just like in Udmurt. In what follows, we concentrate on the case split in Udmurt.

(2) *Bezhta (Kibrik 1995: 20)*:

a. abo-s is
father-GEN$_{dir}$ brother.NOM
'father's brother'

b. abo-la is-t'i-l
father-GEN$_{obl}$ brother-OBL-DAT
'to father's brother'

c. is-t'i-la biLo-ʔ
brother-OBL-GEN$_{obl}$ house-INESS
'in the brother's house'

From a theoretical perspective, the case split in Udmurt is puzzling for the following reason: given a strictly derivational model of syntax in which the structure unfolds in a bottom-up fashion (cf. Chomsky 1995 et seq.), the relevant information about the GF of the DP containing the possessor is not available at the point of case assignment to the possessor. The possessor is assigned case within the DP; later the entire DP is merged with an external head. However, the nature of the external head is already relevant at the point of DP-internal case assignment. The decision which case to assign to the possessor thus seems to require look-ahead. On this background, the case split in Udmurt raises two immediate questions:

1. Is there a way to reconcile the apparent non-local nature of the Udmurt case split with a strictly derivational model of syntax that neither allows for look-ahead nor for counter-cyclic operations?
2. Why is the alternative possessor case the ablative and not one of the various other cases in Udmurt?

According to the literature, it is the GF of the DP containing the possessor that is decisive for the case split. However, the term "direct object" is never

---

[1] Other Daghestanian languages with a case split on the possessor that depends on the case of the possessum include Tsez, Khvarh and Hinugh (cf. Kibrik 1995).

precisely defined. By testing several possible interpretations of this term, we argue that the case split in Udmurt does not depend on GFs; we will show that it is driven by the case value that the dominating DP is assigned, just like in Bezhta:

(3) *Empirical generalization*:
The possessor in Udmurt bears ablative if the XP immediately dominating the possessor bears the accusative marker. The possessor bears genitive elsewhere.

This new generalization facilitates a reanalysis of the case split that does neither require look-ahead nor counter-cyclic case assignment. We will propose a case stacking analysis according to which a DP in Udmurt can receive up to two structural cases. Possessor DPs are always assigned genitive, but may receive another structural case in addition (nominative, genitive, accusative). Morphologically, there is only a single case slot. Hence, if case stacking takes place in the syntax, it will feed fusion of the two case values into a single case feature set in the postsyntactic morphological component. This set is realized by the most specific matching case marker. The combination of genitive and accusative on the possessor yields the abstract representation of a semantic case. The most specific matching marker is the ablative exponent because it is the default semantic case marker in Udmurt. We thus claim that the possessor never bears *abstract* ablative case in Udmurt; it sometimes bears an ablative *marker*, however, this marker realizes the combination of two abstract structural cases and not of abstract ablative case. The postsyntactic manipulation of abstract case features results in the illusion that two different abstract cases (genitive or ablative) can be assigned to the possessor, although ablative is never assigned to it in the syntax. The distinction between abstract and morphological case (cf. Legate 2008) is thus crucial for the analysis.

The paper is structured as follows: Section 2 discusses the locality problem that case assignment in Udmurt seems to pose in more detail. In section 3 the exact distribution of the ablative case is determined by testing the correct meaning of the term “direct object”. Section 4 presents the analysis that derives the generalization gained from the preceding section. Afterwards, empirical and theoretical consequences of the analysis are discussed in section 5. Section 6 concludes.

2. A Look-Ahead Problem

In this section, we will have a closer look at why the case split in Udmurt is interesting from a locality perspective. We will see that case assignment in Udmurt is a challenge for standard assumptions about locality. This is not because case assignment applies over a very long distance; rather, it is the timing of operations that causes problems.

We assume the standard minimalist phrase structure in (4): the internal argument of a transitive verb (DP_{int}, the direct object) is merged as the sister of V and the external argument of a transitive verb (DP_{ext}, the subject) is merged in the specifier of the functional head v which itself takes VP as a complement. There are two more functional projections above vP, headed by T and C. Under a strictly derivational model of syntax (e.g. minimalism, cf. Chomsky 1995 et seq.), this clause structure unfolds step by step in a bottom-up fashion by successive Merge operations. Merge takes two syntactic objects α and β and creates a new syntactic object $[_{\gamma}\ \alpha\ \beta\ ]$.

(4) $[_{\text{CP}}\ \text{C}\ [_{\text{TP}}\ \text{T}\ [_{\text{vP}}\ \text{DP}_{ext}\ [_{\text{v}'}\ \text{v}\ [_{\text{VP}}\ \text{V}\ \text{D}P_{int}\ ]\ ]\ ]\ ]\ ]$

Case is assigned by functional heads to DPs under Agree (Chomsky 2000, 2001); in languages like Udmurt, with a nominative-accusative pattern of case assignment, v assigns accusative to DP_{int} and T assigns nominative to DP_{ext}; D assigns abstract case to the possessor. We take Agree to involve valuation of case on a DP. Prior to case Agree, DPs do not bear a case value, which is represented in what follows by an empty box after the case feature: [case:□].

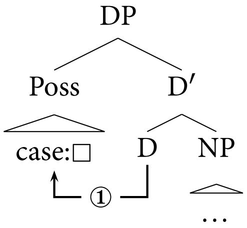
For the moment, we also adopt two main assumptions that are usually (even though not always explicitly) made about case assignment:

1. All syntactic dependencies (including case assignment) are local, that is, they neither involve look-ahead nor counter-cyclic operations.
2. The case values are only manipulated at the point of case assignment.

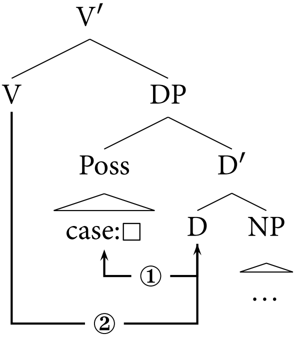
When looking at the case split in Udmurt, a plausible analysis could be that the D head in Udmurt is special in that it cannot only assign genitive but that, in addition, it can also assign ablative to the possessor, depending on the GF on the DP containing the possessor.

Given these assumptions, case assignment to the possessor in Udmurt faces a look-ahead problem: the possessor is assigned a case value (genitive or ablative) by the D head of the DP containing the possessor and the possessum (= step ① in (5)). The choice of the concrete case value seems to depend on the GF of the containing DP. GFs are determined by the position of that DP in the structure; for example, a DP is a direct object if it is the sister of V. However, the information about the structural position of the entire DP is not available at the point of case assignment to the possessor within the DP; the containing DP is merged into the structure only *after* case assignment to the possessor took place (= step ② in (6)).

(5)



(6)



It does not help to reverse the order of operations and to assign case to the possessor *after* the information about the grammatical function of the containing DP is available: if the DP is merged with its selecting head *before* the possessor is assigned case, the relevant information about the category of the selecting head is available. However, case assignment then involves two elements (the DP internal case assigner D and the possessor) that are contained in the DP cycle, but the DP is already dominated by another cycle, i.e. the VP/V′ cycle. This dependency thus violates the Strict Cycle Condition, cf. (7).

(7) *Strict Cycle Condition (based on Chomsky 1973)*:
  a. No operation can apply to a domain dominated by a cyclic node α in such a way as to affect solely a proper subdomain of α dominated by a node β which is also a cyclic node.
  b. Every projection is a cyclic node.

Hence, the dependency of the Udmurt case split on the GF of the DP dominating the possessor inevitably poses a problem for a strictly derivational model of syntax. At the point of case assignment to the possessor, either the relevant information about the GF is not yet available or if it is, it is too late to assign case DP-internally. Theoretically, there are two solutions to that problem. (i) The assumption that all syntactic dependencies apply in accordance with the SCC (cf. (7)) and that they do not require look-ahead (assumption 1) must be wrong: at least possessor case assignment in Udmurt must be non-local in that sense. (ii) The assumption that case is only manipulated at the point when the case value is assigned (assumption 2) is wrong: case values can additionally be manipulated in a postsyntactic component.

In what follows, we pursue solution (ii) since we do not want to give up standard syntactic locality restrictions like the SCC. This decision is supported by the results that will be presented in the following section: the distribution of the ablative does *not* depend on the GF of the DP dominating the possessor, but rather on its case value. And this information can be made locally available on the possessor without look-ahead under a case stacking analysis and given a postsyntactic realizational morphology.

3. Empirical Background

In this section, we will broaden the empirical basis for the analysis of the case split in Udmurt. In section 3.1, it will be shown that the genitive and the ablative possessor occupy the same structural position; in section 3.2, the functions of the ablative outside of the possessor construction are examined in comparison with the other cases. Finally, we will run a number of tests in section 3.3 in order to disambiguate the term "direct object" and to gain a more accurate generalization about the distribution of the ablative on the possessor.

3.1. The Structural Position of the Possessor

A number of other Uralic languages also show a case split on the possessor, although the factors conditioning the split are different from those found in Udmurt. These languages are, e.g., Hungarian, Finnish, Estonian, Komi (cf. König and Haspelmath 1998, Nikolaeva 2002; see also Deal to appear about a similar alternation in Nez Perce). The alternation is best studied in Hungarian in which nominative and dative possessors alternate:

(8) *Two possessor cases in Hungarian (Szabolcsi 1994)*:

a. (a) Mari kalap-ja
(the) Mari.NOM hat-POSS.3SG
'Mari's hat'

b. Mari-nak a kalap-ja
Mari.DAT the hat-POSS.3SG
'Mari's hat'

What is important for the present discussion is that it has been convincingly argued for Hungarian that the nominative and the dative possessor occupy different structural positions in the DP (cf. Szabolcsi 1984, 1994). There are two types of evidence for this conclusion. (i) Linear order: the determiner *a(z)* obligatorily follows the dative possessor but precedes the nominative possessor, cf. (8). (ii) Extraction asymmetries: only the dative possessor can be extracted from the DP, cf. (9).

(9) *Possessor extraction in Hungarian (Szabolcsi 1984)*:

a. Mari-nak nem ismert-em [t′ t növér-é-t]
Mari-DAT not knew-1SG sister-POSS.3SG-ACC
'I never knew any sister of Mari.'

b. *Mari nem ismert-em [t′ t növér-é-t]
Mari.NOM not knew-1SG sister-POSS.3SG-ACC
'I never knew any sister of Mari.'

Wh-possessors must be in the dative and precede the determiner, see (10):

(10) Wh-*possessors in Hungarian (Szabolcsi 1994)*:

a. *ki kalap-ja
who.NOM hat-POSS.3SG
'whose hat?'

b. ki-nek a t kalap-ja
who-DAT the hat-POSS.3SG
'whose hat?'

Szabolcsi (1994) takes this as evidence that the dative possessor is in a derived position, viz. SpecD, the DP-counterpart of SpecC (an operator position that serves as an escape hatch for movement out of DP). It moves to this position from a position lower down in the structure associated with nominative case.

The question that arises is whether there is also evidence for two different positions of genitive and ablative possessors in Udmurt. Linear order is not a viable test for Udmurt: there is no element like the Hungarian determiner relative to which the two possessors align differently. However, the extraction test suggests that the genitive and the ablative possessor occupy the same structural position: in contrast to Hungarian, the possessor can always be extracted, regardless of its case value. This is illustrated in (11) for extraposition. Usually, the possessor precedes the possessum, but the possessor can also be extraposed to the right edge of the clause. That extraposition of the possessor has taken place is evident because Udmurt is an SOV language but in (11) the possessor follows the verb.

(11) *Extraposition of the genitive and ablative possessor in Udmurt*:
a. man’eryz sytše peres’ Mikta-len
manner.3sg such old Mikta-gen
‘Such is old Mikta’s style.’ (Vilkuna 1997: 224)
b. valze jusky so-les’
horse.acc.3sg unharness.imp.2sg s/he.abl
‘Unharness his horse!’ (Vilkuna 1997: 224)

There is also no evidence from other factors such as word order and/or agreement that genitive and ablative possessors pattern differently and thus that they occupy different structural positions in Udmurt. We therefore assume that the case split does not arise from a difference in positions of the possessor in the DP.

3.2. Functions of the Ablative

Udmurt has altogether 15 cases, as shown in (12). We divide the cases into two groups: structural and semantic cases. Structural case is assigned to a certain position in the syntactic structure; it is neither tied to specific theta-roles of the elements that receive them (like inherent case) nor to a specific class of verbs that exceptionally assign them to their arguments (like lexical case). We use the term “semantic case” for any non-structural case. This distinction will play an important role in the analysis and we will come back to it in more detail in section 4. For the moment, it suffices to keep in mind the term “semantic case”.

(12) *Case system in Udmurt*:

Structural cases	Semantic cases	
nominative	dative	inessive
accusative	ablative	elative
genitive	abessive	illative
	adverbial	egressive
	instrumental	transitive
	approximative	terminative

Let us concentrate first on the semantic case ablative. In addition to encoding possessors, the ablative marker has a number of other functions in Udmurt: according to Edygarova (2009) and Winkler (2001), it is used to express comparison (cf. (13-a)), the cause of an action (cf. (13-b)), the material something is made of (cf. (13-c)), and it marks adjuncts expressing origin and source (cf. (13-d)). Furthermore, the ablative is governed by some postpositions (cf. (13-e)) and by verbs of asking.

(13) *Uses of the ablative (Edygarova 2009: 108, Winkler 2001: 22-23)*:

a. vit'ton-leš uno
fifty-ABL more
'more than fifty' *comparison*

b. so-leš žad'-em
he/she-ABL be tired-2.PST/3SG
'(he) got tired with him' *cause*

c. basma-leš leśt-em arberi-os
cloth-ABL make-PART thing-PL
'things which are made from cloth' *material*

d. mon so-leš gožtet bašt-i
I he-ABL letter.ACC get-1.PST
'I got a letter from him.' *source*

e. ta-leš ažlo
DEM-ABL before
'before this' (temporal) *postposition*

Thus, the ablative appears in a variety of contexts. What is remarkable is that these contexts do not seem to form a natural class. This becomes even more obvious when looking at the distribution of the other semantic cases: they are much more restricted in their applicability (see Winkler 2001: 16ff. for an overview). We conclude from these facts that the ablative suffix is the default semantic case marker in Udmurt. This means that it could potentially encode any DP that does not bear a structural case, but it is often blocked by the other semantic case markers with a more specific meaning. The ablative as a default semantic case marker will be important for the analysis of the ablative case on the possessor and will be formally implemented in section 4.

3.3. Distribution of the Ablative: Finding the Correct Generalization

We concluded in section 3.1 that there is no asymmetry between the ablative and the genitive possessor with respect to their structural position. Under these circumstances, the question arises as to what the correct generalization is that predicts which of the two cases is used in which context. Traditionally, the generalization is that a possessor gets ablative if the DP which contains the possessor has the GF direct object; the genitive is used elsewhere. However, the term "direct object" is ambiguous and it is never precisely defined in any of the descriptive works on the possessor cases in Udmurt. There are basically three possible interpretations of "direct object":

1. *Thematic role:*
   The possessor gets ablative if the XP immediately dominating the possessor DP has the macro-role patient.[2]
2. *Position in the tree:*
   The possessor gets ablative if the XP immediately dominating the possessor DP is selected by V.
3. *Case:*
   The possessor gets ablative if the XP immediately dominating the possessor DP is assigned accusative.

The occurrence of the ablative in example (1-b) is compatible with each of these interpretations. Therefore, it is necessary to find contexts in which the interpretations make different predictions in order to find out which of them is the correct one. The question is whether in a given context the possessor must get genitive or ablative.[3]

---

[2] Dominance is to be understood non-reflexively in these definitions.

[3] Unless references are indicated, the following Udmurt data are due to Svetlana Edygarova, a native speaker of Udmurt.

We start with interpretation 1 according to which the possessor of a DP with the macro-role patient gets ablative; in a DP with any other theta-role it should get genitive. A relevant test case for that hypothesis is the active-passive alternation. Take as a starting point a transitive verb that has a DP with a possessor as its internal argument and assigns the patient role to that DP, cf. (14-a). If this verb is passivized, the DP with the possessor becomes the sole argument of the passivized verb. Since passivization only changes GFs but does not change theta-roles, the DP containing the possessor still bears the patient role and should thus get ablative according to hypothesis 1. However, as (14-b) shows, this prediction is not borne out. The possessor in the sole argument DP of a passivized verb receives genitive even if this DP bears the patient role. Interpretation 1 is thus falsified.

(14) *Possessor case in active-passive alternation*:

a. Petyr Masha-leš puny-z-e zhug-i-z
Peter Masha-ABL dog-3SG-ACC beat-1PST-3SG
'Peter beat Masha's dog.'

b. Masha-len/*-leš puny-jez zhug-em-yn val
Masha-GEN/-ABL dog-3SG beat-PART-INES AUX.1PST
'Masha's dog was beaten.'

Next, we test whether interpretation 2 makes the correct predictions: only if the DP containing (i.e., immediately dominating) the possessor remains in its VP-internal base position can the possessor receive ablative. Since Udmurt is a head-final language with predominant SOV order, it cannot be read off of the surface position of the internal argument relative to the verb whether the argument DP is still in the VP or whether it is moved out of the VP; it will precede the verb in any case. We therefore need other means to make sure that the DP with the possessor is indeed in the VP. First, we can control for the VP-internal position of the DP by means of adverbs that mark the VP-boundary. Example (15) is based on the passivized sentence in (14-b). In addition, it contains a temporal adverb that marks the VP-boundary. Since the DP with the possessor is to the right of that adverb, it must be in the VP. As is shown in (15), the possessor still gets genitive case. If, however, hypothesis 2 was on the right track, the possessor should get ablative, contrary to fact. This falsifies interpretation 2.[4]

(15) *Adverb 'yesterday'*:
tolon Masha-len puny-jez zhug-em-yn val
yesterday Masha-GEN dog-3SG beat-PART-INES AUX.1PST
'Yesterday Masha's dog was beaten.'

Thus, only interpretation 3 remains: the possessor gets ablative if the DP containing it receives accusative case. This hypothesis predicts that a possessor contained in a DP that bears a case different from accusative receives genitive case. This was shown in (1-a) for nominative case. (16) shows this for dative case (a semantic case). There are a number of transitive verbs in Udmurt that lexically govern a case different from accusative. The verb *akyltnï* 'to bother' is one that assigns dative to its internal argument. If the dative marked argument contains a possessor, this possessor must not bear ablative but genitive instead, cf. (16). This is correctly predicted by hypothesis 3. In addition, this example also falsifies interpretations 1 and 2: the DP containing the possessor is probably in a VP internal position and has the patient role, but the possessor bears the genitive. Another example that supports interpretation 3 will be presented in (21) for a genitive DP containing a genitive possessor.

(16) *Dative assigning verb*:
Petyr [Masha-len suzer-ez-ly] akylt-e
Peter Masha-GEN sister-3SG-DAT bother-PRES.3SG
'Peter is bothering Masha's sister.'

So far, the conclusion is that only interpretation 3 is compatible with the empirical facts. Thus, the following generalization about the distribution of the ablative and genitive case on the possessor arises:

(17) *Empirical generalization*:
The possessor in Udmurt bears ablative if the XP immediately dominating the possessor bears the accusative marker. The possessor bears genitive elsewhere.

---

[4] We are currently running a number of other tests to show that the DP containing the possessor can remain inside the VP and nevertheless bears genitive (VP ellipsis, VP topicalization, VP coordination). Unfortunately, the results are not yet available to us.

This generalization is also in accordance with a number of other facts. If the relevant factor that triggers ablative marking on the possessor is accusative marking of the containing DP, it is predicted that the source of the accusative should not play a role, i.e., that the possessor bears ablative also if the accusative is not assigned by v to the internal argument of V as in (1-b). That this prediction is borne out can be shown with ECM constructions.[5] There are two types of ECM constructions in Udmurt. One is very similar to the English ECM in that the matrix ECM verb embeds a clause and assigns accusative to the subject (the external argument) of that clause. If the embedded subject contains a possessor, this possessor must bear ablative and the possessum bears accusative (cf. (18)) as predicted by the generalization in (17).

(18) *ECM construction in Udmurt*:
Petyr Masha-leš puny-z-e tyloburdo-os-ty kutyl-e
Peter Masha-ABL dog-3SG-ACC bird-PL-ACC.PL catch-PRES.3SG
malpa
think.PRES.3SG
'Peter believes Masha's dog to catch birds.'

Note that this example also falsifies interpretation 2: the DP containing the possessor (the subject of the embedded clause) is not the sister of the matrix V (it is included in the sister of V). If one assumes that ECM subjects only receive a thematic role from the embedded verb, this example also falsifies interpretation 1: the subject DP containing the possessor bears the agent role of the verb *kutylnï* 'to catch'.

In the second and more widespread ECM construction in Udmurt, the embedded clause is nominalized. The subject of the embedded transitive verb is in a sense the possessor of that derived noun and it is consequently marked genitive instead of nominative. This can be seen if a nominalized clause is not embedded under an ECM verb but is rather used as the sentential subject of a verb, cf. (19). Here, the subject of the nominalized clause, *Peter*, bears genitive:

---

[5] Another potential test case for the generalization in (17) would be one in which a DP with a possessor is selected by an accusative assigning postposition (Udmurt is head-final outside of the DP). The majority of postpositions governs nominative in Udmurt; there are also some postpositions that assign semantic cases such as ablative or dative. There is, however, not a single postposition that governs accusative (Winkler 2001: 19, S. Edygarova, p.c.). Hence, the prediction that there should be ablative on the possessor of the DP selected by an accusative assigning postposition cannot be tested.

(19) Petyr-len Masha-leš pyny-z-e vi-em-ez myn-ym
Peter-GEN Masha-ABL dog-3SG-ACC kill-PART-3SG 1SG-DAT
ug jara
NEG.PRES.3 please.SG
'[Peter's killing Masha's dog] does not please me.'

If, however, the nominalized clause is embedded under an accusative assigning ECM verb, the subject of the nominalized verb must bear ablative instead of genitive, cf. (20). This is expected given the generalization in (17): the subject of a nominalized clause that behaves like a possessor in that it is usually assigned genitive shows ablative marking if the XP containing it (probably vP or TP) is assigned accusative by the external head (the ECM verb assigning accusative to its sister node). Example (20) also shows that the case split not only concerns prototypical possessors as in *Peter's head* but that possessor is to be understood more abstractly; in Udmurt the relevant factor seems to be genitive marking (outside of the special accusative environments where ablative replaces genitive). In any case, the two ECM constructions behave alike with respect to the case split: accusative assignment into the embedded clause triggers ablative marking on usually genitive marked DPs in the embedded subject, as predicted by the generalization in (17).

(20) mon Petyr-leš Masha-leš puny-z-e vi-em-z-e
1SG Peter-ABL Masha-ABL dog-3SG-ACC kill-PART-3SG-ACC
adzdz-i
see-1PST.1SG
Lit. 'I saw Peter's killing Masha's dog.'

Finally, consider what happens if there are recursive possessors: if a DP that contains multiple possessors is not assigned accusative, as e.g. in (21) where it is the sole argument of an intransitive verb that bears nominative zero marker, all possessors in that DP bear genitive, as expected. If, however, a DP that contains multiple possessors is assigned accusative, only the structurally highest possessor gets ablative; the possessors that are embedded more deeply in the structure still get genitive marking, cf. (22) (see also Edygarova 2010: 177).

(21) Masha-len apaj-ez-len puny-jez iz'-e
Masha-GEN sister-3SG-GEN dog-3SG sleep-PRES.3SG
'Masha's sister's dog is sleeping.'

(22) Petyr Masha-len apaj-ez-leš puny-z-e zhug-i-z
Peter Masha-GEN sister-3SG-ABL dog-3SG-ACC beat-1PST-3SG
'Peter has beaten Masha's sister's dog.'

This pattern is predicted by the generalization in (17): the lower possessor *Masha* is immediately dominated by the DP *Masha's sister* that bears ablative. But since it is accusative marking of the dominating DP that is a prerequisite for ablative marking on the possessor, the lower possessor must bear genitive. The highest possessor receives ablative marking because the possessum gets hier accusative case.

To conclude, hypothesis 1 is falsified by the data from the active-passive alternation in (14) and hypothesis 2 can be rejected on the basis of the placement of low adverbs before a DP with a genitive possessor in (15). In addition, both hypotheses can be refuted by the ECM constructions in (18) and (20).

In this section, we have provided evidence that (i) the ablative and the genitive possessor occupy the same structural position (which we take to be SpecD), (ii) the ablative is the default semantic case marker in Udmurt, and (iii) the decisive factor that governs the distribution of the ablative on the possessor is not the GF but rather the case value of the immediately dominating XP. In the next section, we present a local and cyclic derivation of (iii) that crucially relies on case stacking and (ii), the default nature of the ablative marker.

4. Deriving the New Generalization

The generalization in (17), with the case marker of the DP immediately dominating the possessor as the decisive factor for the Udmurt case split, facilitates a reanalysis of the phenomenon that is in accordance with the Strict Cycle Condition and does not require look-ahead. We propose that the possessor is assigned the prototypical possessor case (abstract genitive) in the DP *plus* the case value of the external head – an instance of syntactic case stacking.[6] In this way, all the relevant information necessary to decide between the genitive and the ablative marker is locally available on the possessor in the postsyntactic morphological component. The derivation of the generalization in (17) is

---

[6] Diachronic facts of the Finno-Ugric language family corroborate this analysis. For Finnish, e.g., it has been argued, that the local cases, among them the ablative, is a combination of two case markers. See Abondolo (1998: 167) and Suhonen (1988: 302) for details. Thanks to András Bárány for pointing these facts out to us.

divided into three subparts that contain assumptions about the lexicon, the syntax of case assignment and the morphological realization of abstract case.

4.1. The Lexicon

As has already been introduced in section 3.2, Udmurt exhibits a rich case system (cf. Winkler 2001). We divide the cases into groups of structural and semantic cases, cf. (23); the defining property of the former group is that these cases are assigned to specific positions in the syntactic structure, independently of the thematic role assigned to those positions.

(23) *Case system in Udmurt*:

Structural cases	Semantic cases	
nominative	dative	inessive
accusative	ablative	elative
genitive	abessive	illative
	adverbial	egressive
	instrumental	transitive
	approximative	terminative

We assume that abstract cases are decomposed into the binary features [±obl(ique)] and [±obj(ect)] (cf. Bierwisch 1967). A case is [+obj] if it is a case that a verb assigns to its complement (maybe lexically). Prototypically, this is the accusative, but we have already seen that verbs in Udmurt may also assign other cases to their complement, e.g. ablative or dative. [+obl] characterizes those cases that are not standardly assigned to the core arguments of a verb in a given argument encoding pattern. In a language like Udmurt with an accusative alignment pattern, the cases standardly assigned to the core arguments, i.e., the [−obl] cases, are nominative and accusative; all remaining cases are [+obl].[7] As a consequence, all semantic cases are represented as [+obl, +obj]. The structural cases are negatively valued for at least one of the two binary features. The decomposition for the three structural cases nominative, accusative and genitive is shown in (24-a); the decomposition for the semantic cases is given in (24-b).

---

[7] In this system, genitive is classified as a structural case but at the same time it is represented as an oblique case. See also Halle (1997) for a case decomposition that characterizes genitive as oblique and structural.

(24) a. *Structural cases:*

NOM	[−obl,−obj]
ACC	[−obl,+obj]
GEN	[+obl,−obj]

b. *Semantic cases:*

ABL	[+obl,+obj], [−f,−g,. . .]
DAT	[+obl,+obj], [+f,−g,. . .]
INSTR	[+obl,+obj], [−f,+g,. . .]
. . .	

Since the semantic cases are all specified as [+obl, +obj], further features are necessary to distinguish between them. We assume that this is done by a set of semantic features which are abstractly represented as [±f], [±g], . . . in (24); their exact meaning does not matter for our purposes. As a consequence, the semantic cases are more complex than the structural cases (cf. Béjar and Massam 1999). They consist of two case feature sets: like the structural cases they have a feature set that contains values for [±obl] and [±obj], but unlike the structural cases they have a second set of semantic features.

The assumption that semantic cases are more complex than structural cases is independently motivated by the fact that in various non-related languages, the markers of the semantic cases are built upon a structural case marker, exemplified by the paradigms in (25) (Arkadiev 2006).[8] In Romani for example, the semantic cases are built on the basis of the structural case accusative.

(25) a. *Case in Romani*:

	'pigeon'
Nom	golumbo
Acc	golumbo-s
Loc	golumbo-s-te
Dat	golumbo-s-ke
Abl	golumbo-s-tyr
Ins	golumbo-s-a

b. *Case in Naukan Eskimo*:

	'dog'
Nom	aɣna-q
Erg/Gen	aɣna-m
Ins	aɣna-m-ɨŋ
All	aɣna-m-un
Loc1	aɣna-m-i
Loc2	aɣna-kun

With this background, we turn to our assumptions about the lexical properties of relevant items. First, we assume that all case bearing elements in Udmurt have exactly two case slots; more precisely, they have a lexical property allowing them to receive up to two case feature sets; see (26-a) for D and (26-b) for N.

---

[8] In Nanosyntax, semantic cases are also more complex than structural cases: privative case features are represented in a hierarchy in which the semantic cases dominate the structural cases, cf. Caha (2008, 2009).

(26) a. D {[case:□], [case:□]} b. N {[case:□], [case:□]}

Each of these unvalued "slots" can be valued by a syntactic or a semantic case feature set, each of which is a bundle of binary case features. Consequently, D and N heads (as any other case bearing elements) can be assigned up to two abstract case feature sets. This is to say that Udmurt exhibits syntactic case stacking (see section 5 and footnote 6 for further discussion). A number of languages can overtly stack cases on the possessor: it carries the genitive marker plus the case marker realizing the case assigned to the DP containing it. An example from Huallaga Quechua is given in (27). What we assume is that this stacking happens in Udmurt syntax as well (stacking of abstract cases), although the cases cannot be stacked overtly in Udmurt (stacking of case exponents). We will come back to that issue in section 4.3.

(27) *Case stacking in Huallaga Quechua (Pylkkänen 2002)*:

Hipash-nin-ta	kuya-ː	Hwan-pa-ta
daughter-3POSS-ACC	love-1	Juan-GEN-ACC

'I love Juan's daughter.

Since there are two case slots on D and N heads in Udmurt, a comment on the Case filter is necessary: in order to fulfill the Case filter, every DP, i.e., its D head and the head of the NP selected by D, must receive at least one abstract case, i.e., one case feature set; it is not necessary for the derivation to converge that both case slots on D and N are valued by an abstract case. An unvalued case slot does not violate Full Interpretation as long as there is a valued case slot on the same head.[9] Only if none of the case slots of a head is valued, a fatal violation of the Case filter obtains.

The second assumption concerns the valuation potential of structural and semantic cases: we assume that semantic cases fill two case slots on a head rather than one like structural cases (similar ideas have been put forward by Béjar and Massam 1999, Richards 2008). The reason for this is that semantic cases are more complex than structural cases: they consist of a syntactic and a semantic case feature set. If a semantic case is valued on D (or N), one of D's case slots is filled by the syntactic case feature set [±obl, ±obj]; the second

---

[9] Alternatively, one might assume that unvalued case features can be deleted by default at the end of the derivation if there is a valued case slot on the same head. The decision between the two options does not have any crucial impact on the analysis of the case split in Udmurt and we continue to adopt the solution without deletion of unvalued case slots in what follows.

slot is filled by the semantic case feature set. It is not possible to value only one of the two sets of the semantic case on D; both of them must be copied under Agree with D, they form a unity. An important consequence of this assumption is that only two structural cases can stack. A structural and a semantic or two semantic cases cannot stack because one semantic case alone already fills the two case slots on D (or N) and no further valuation is then possible. In what follows, the relevant configuration will be one in which the head of the possessor DP has been assigned genitive case and is to receive a semantic case from an external head: the prior assignment of the genitive blocks valuation of the semantic case because there are only two case slots on the D head and genitive+semantic case would need three slots.[10]

4.2. The Syntax

As indicated before, we assume that abstract case, represented by decomposed case features, is assigned by functional heads to arguments under Agree. D assigns abstract genitive case to a possessor in SpecD; v assigns abstract accusative case to the internal argument, and T assigns abstract nominative case to the external argument under c-command.[11]

In order to model case concord, we assume that abstract cases are assigned to all case-bearing elements in the DP via *Multiple Agree* (Hiraiwa 2001), i.e. to D and N heads (and other DP-internal heads such as heads of number and adjectival projections if they are present). That means that the case of a func-

---

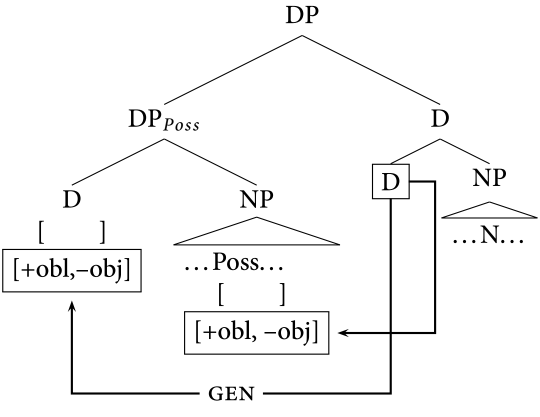
[10] The fact that the possessum can bear a semantic case motivates that N and D heads must have at least two case slots because a semantic case needs to value two case slots. That these heads have *exactly* two case slots and not more is motivated by the distribution of genitive and ablative in multiple possessor constructions. See section 5.2 on cross-linguistic variation concerning the number of cases that can stack and the derivation of multiple possessor constructions in (32) below for motivation on this upper boundary of the number of case slots on D and N: if D and N heads had more than two case slots, the analysis in section 4 would wrongly predict that not only the structurally highest possessor but also more deeply embedded possessors should get ablative marking in recursive possessor constructions.

[11] Under the assumption that the case assigner must c-command the DP, a problem with genitive case assignment arises: we assume that possessor DPs are specifiers within DP. Consequently, they are not in the c-command domain of the D head which assigns genitive case. In order to solve this problem one could adopt the search algorithm in Řezáč (2004). In this approach, the search space of Agree can be expanded "upward" if and only if there is no suitable goal in the c-command domain of the probe. Assuming that the complement NP of D does not constitute a suitable goal, the search space of D is expanded to include the specifier.

tional head can be assigned to more than one element.[12] Only those elements that have an unvalued case feature can receive a value. In this system, case assignment for abstract structural and semantic cases is identical, i.e., both are assigned in the syntax. For the sake of concreteness, we assume that semantic cases are assigned to their complements by zero adpositions (cf. Pylkkänen 2002, Hole 2008, among others).

Given these assumptions, case assignment in clauses with a possessor proceeds as follows. We start with genitive assignment in the DP. As shown in (28), the possessor is merged as the specifier of the D head of the possessum. D assigns genitive [+obl, −obj] to its specifier. More precisely, it assigns it to the N head of the possessor and thereby values one of its two case slots; in addition, given the possibility of Multiple Agree, the D head in the box assigns genitive to the D head that selects the possessor NP, filling one of its two case slots, too. The second case slot on the D and the N head in the possessor DP remains unvalued.

(28) *Genitive case assignment in the DP*:



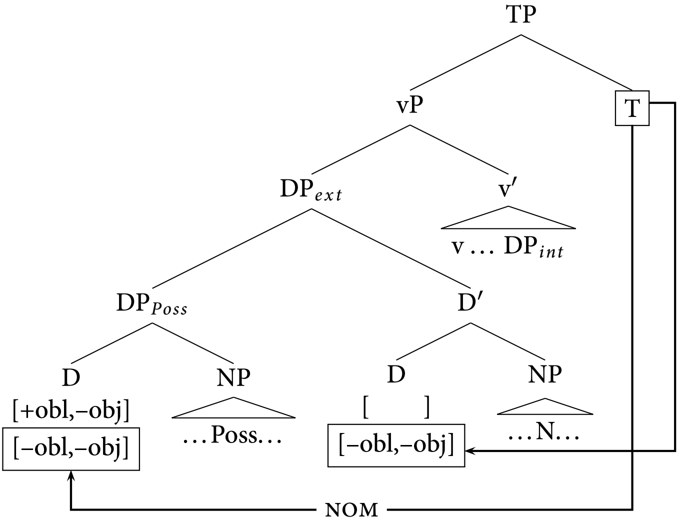
---

[12] There are two related proposals on concord that would also be compatible with our analysis. (i) Case is assigned to the sister node δ of a case assigner and, unless blocked by independent principles, the case value on δ spreads downward in the domain dominated by δ (see Matushansky 2008, Bjorkman to appear, Erlewine 2012). This mechanism is, however, countercyclic (cf. the SCC in (7)). (ii) Concord can also be modeled as feature-sharing (cf. Frampton and Gutman 2006, Schoorlemmer 2009). But then two different mechanisms are applied to model agreement and concord. Multiple Agree handles both phenomena in a uniform way.

In the following derivations, the entire DP in (28) is merged into different positions in the clausal spine. The initial step of DP-internal genitive assignment to the D and N head in the possessor DP always takes place, but for the sake of clarity we do not indicate it anymore in the following trees. For the same reason, we will only indicate case assignment to D heads in what follows; we omit case assignment to N heads (the possessum), but the reader may verify that no complications arise for case assignment to them; the N head will always show the same case as the D head that selects the NP projected by N.

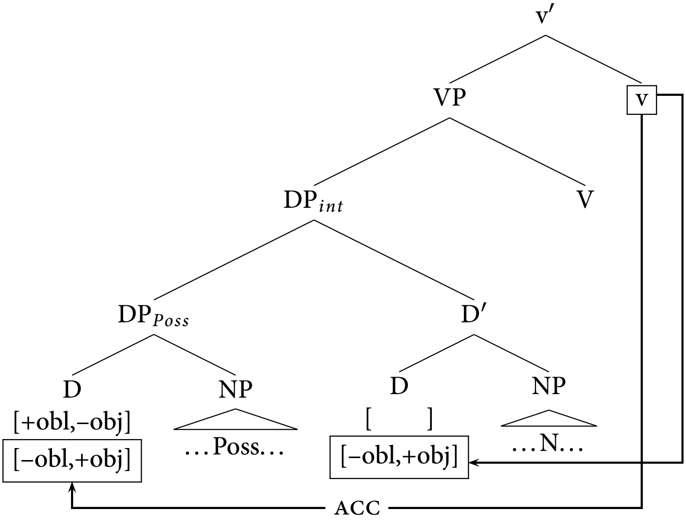
First, the DP containing a possessor is merged as the external argument of a transitive verb in Specv. This DP is assigned nominative by the c-commanding T head, cf. (29). T values nominative [−obj, −obl] on the D head of the possessor DP contained in the external argument DP. The head of the possessor already bears genitive, but the second slot is still available for nominative from T; genitive and nominative thus stack on this D head. In addition, T assigns nominative to the D head of the external argument, filling one of the two case slots on this D head.

(29) *Nominative case assignment*:



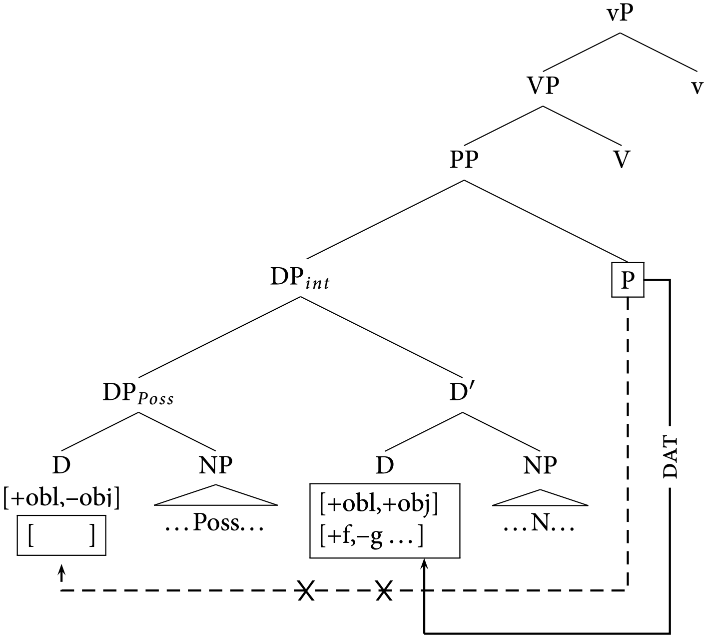
Virtually the same happens if the DP in (28) is merged as the internal argument of a transitive verb. In that position, it is assigned accusative [−obl, +obj] by v, cf. (30). Just as T, v assigns accusative to the D head of the internal argument, filling one of its two case slots, and to the D head of the possessor contained in the internal argument, filling its remaining case slot. The result is stacking of genitive and accusative on the possessor's D head.

(30) *Accusative case assignment*:



The DP in (31) is assigned a semantic case. This happens if the DP is the sister of a (phonologically empty) zero postposition. This PP may, for example, be merged as the internal argument of V, cf. (31). For concreteness, we illustrate semantic case assignment with dative, but the derivation would be the same with any other semantic case value. The only thing that varies between them is the semantic features [±f, ±g, …]. In (31), P assigns dative [+obl, +obj], [+f, −g, …] to the D head of the internal argument. Since semantic cases are complex, they value both case slots of this D head: one with the syntactic features [+obl, +obj] and the other one with the semantic case feature set [+f, −g, …]. Crucially, however, P cannot assign dative to the D head of the possessor (see the crossed out arrow in (31)). The reason is that this head already bears genitive which, due to cyclicity, has been valued on D *before* the entire DP is merged as the complement of P. But since semantic cases need to fill two case slots, there is not enough "space" left on the D head of the possessor for the dative case: prior assignment of the genitive (or any other case) thus *bleeds* assignment of a semantic case in the syntax. As a consequence, there is no case stacking on the possessor if the case assigned to the dominating DP is a semantic case. The possessor D head thus ends up only bearing genitive in these contexts.[13]

(31) *Dative case assignment*:



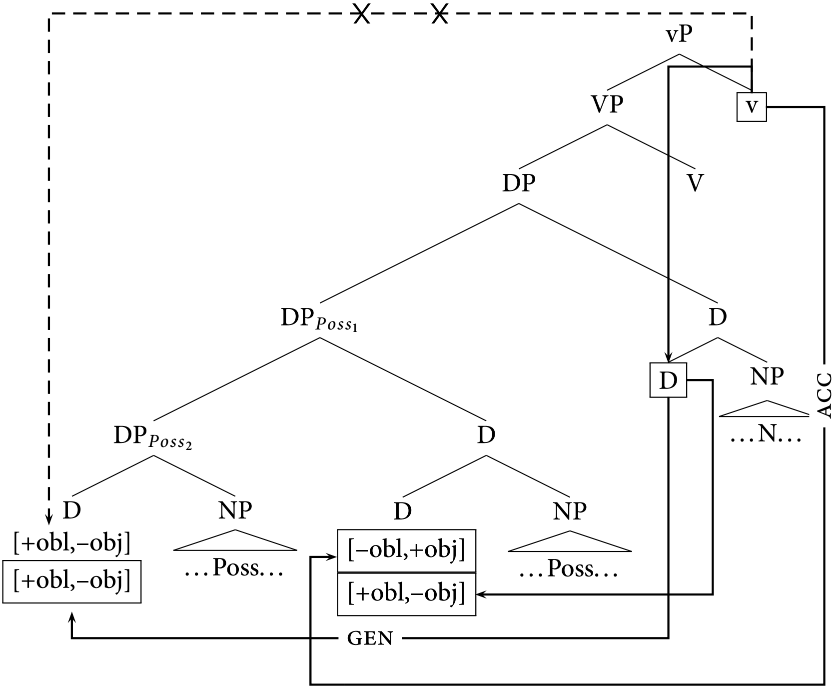
Finally, consider a case of double possessors with double genitive assignment to the most deeply embedded possessor. In (32), the structurally highest DP contains a possessor DP (DP$_{Poss_1}$) in its specifier which in turn contains a possessor (DP$_{Poss_2}$). The lowest possessor (DP$_{Poss_2}$) gets genitive [+obl, –obj] from the D head that selects it, as illustrated in (28). This first step is not indicated in (32). Next, DP$_{Poss_1}$ is merged in the specifier of the D head in the box (the head of the structurally higher possessum). This D head also assigns genitive to its specifier (DP$_{Poss_1}$). More precisely, it assigns it to all elements in its specifier with an empty case slot (see the arrows starting from the boxed D in (32)). This is the case for the D head of DP$_{Poss_2}$ and the D head of DP$_{Poss_1}$. As a result, the D head of DP$_{Poss_2}$ bears genitive twice and the D head of DP$_{Poss_1}$ bears it once, its second case slot is yet unvalued.

---

[13] The second case slot of D can remain unvalued. Recall that it is sufficient for the Case filter if one of the two case slots is valued, see the discussion in section 4.1.

The complex DP containing the two possessors can now be merged with another head, e.g. in the specifier of T, v or as the complement of P. The interesting case is the one in which this DP is merged as the sister of a transitive verb such that it gets accusative case from the closest c-commanding case assigner v. Recall that, empirically, it is the case that in such a configuration only the structurally highest possessor ends up with ablative, which depends on having received accusative (cf. (22)); the most deeply embedded possessor must bear genitive, which means that it does not receive accusative. The tree in (32) with the complex DP merged as the sister of V illustrates why this is the case: v assigns accusative [–obl, +obj] to all elements in its c-command domain that have unvalued case slots. Accusative can thus be assigned to the D head of the structurally higher possessor (DP_{Poss_1}) and to the D head selecting DP_{Poss_1} (see the arrows starting from v in (32)). The former received genitive case before but still has an empty case slot for another structural case; the latter still has two empty case slots of which the accusative values one. Since the D head of the structurally higher possessor is assigned accusative in addition to genitive, we will find ablative marking on this possessor. Now consider the D head of the lower possessor DP_{Poss_2}. It cannot receive accusative from v because its two case slots are already filled, each of them is valued by the genitive. Hence, there is no unvalued case feature left for the accusative (see the crossed out arrow in (32)). And since accusative marking of the possessor's D head is a prerequisite for its occurrence with ablative marking, DP_{Poss_2} will bear the genitive suffix. The different morphological marking of the two possessors in a multiple possessor construction thus follows from the restriction of the number of case slots in Udmurt (an assumption that will be further discussed in section 5) and their structural position: due to its deeply embedded position in the tree, the lower possessor gets genitive twice and is then not able to get accusative; the higher possessor does not get the first genitive value because it is not in the domain of case assignment of the relevant head, and it can thus receive accusative from the external head. This mechanism also works if more than two possessors occur. Only the highest will be able to carry the ablative marker.

(32) *Double possessors: genitive case assignment*:



In the next subsection, we will show how exactly a D head with the accusative value stacked on the genitive value ends up with the ablative case morpheme.

4.3. The Morphology

We assume a postsyntactic realizational morphology. The syntax only operates with abstract feature bundles that are realized by exponents in the morphological component. For concreteness, we adopt the framework of Distributed Morphology (DM, Halle and Marantz 1993, 1994, Harley and Noyer 1999): vocabulary items (VIs) that are pairings of morphosyntactic features with phonological information are inserted into terminal nodes in the syntactic structure. VIs can be underspecified with respect to their morphosyntactic features which leads to competition for insertion between them. This competition is resolved by the Subset Principle and Specificity: only the most specific matching VI can be inserted into a terminal node, i.e., the VI that has the largest subset of the morphosyntactic features of the terminal.

In the present discussion, the question is how the abstract binary case features are realized by case exponents. We assume that case VIs in Udmurt are exclusively specified for positive features, cf. (33) (see Zwicky 1977, Wunderlich 1996, Harley and Ritter 2002, Nevins 2003, among others).

(33) *Case vocabulary items in Udmurt*:
[+obl,+obj,+f] ↔ */li/* (DAT)
[+obl,+obj] ↔ */leš/* (ABL)
[+obl] ↔ */len/* (GEN)
[+obj] ↔ */e/* (ACC)
[ ] ↔ */Ø/* (NOM)

The nominative exponent is the completely underspecified elsewhere marker. The genitive and the accusative exponent are underspecified: they are only specified for one of the two syntactic case features [+obl] and [+obj], respectively. All semantic case exponents are fully specified for the structural case features. In order to distinguish between the various semantic cases, the relevant VIs are specified for some of the semantic features [±f, ±g, . . .]. In (33), the dative VI, for example, is specified for the semantic feature [+f] in addition to being specified as [+obl, +obj]. Another semantic case VI may instead be specified as [+g] or [+f, +g]. How exactly the semantic case features are distributed over the semantic case VIs is of no importance and we thus only indicated the specification for the dative in (33). What is of central importance, however, is the specification of the ablative exponent: crucially, this semantic VI is fully specified for [+obl, +obj], as all the semantic case VIs; however, if it is completely underspecified for the semantic case features, in contrast to all remaining semantic case VIs. This encodes the default character of the ablative; recall the conclusion from section 3.2 that the ablative marker is the default semantic case exponent in Udmurt. It occurs in a variety of contexts that do not form a natural class. The representation in (33) reflects this fact: the underspecified ablative exponent is in principle compatible with *every* syntactic context that is specified as [+obl, +obj]; it is, however, often blocked by a more specific semantic case VI.

We can now have a look at the operations that happen in the morphological component. In the last section we assumed that Udmurt has case stacking in the syntax. However, Udmurt obviously does not exhibit overt case stacking, i.e., stacking of case exponents on the possessor as, for example, Huallaga Quechua in (27) does. To implement this, we assume that Udmurt has a filter which excludes the co-occurrence of two case exponents. As a consequence, a problem arises if case stacking took place in the syntax because only a single case exponent can be realized but two abstract case values are present on a terminal. In order to obey the filter, a repair operation applies in the morphological component prior to vocabulary insertion: the two case features sets on the terminal node (D or N) have to fuse into a single feature set (see Noyer 1992, Halle and Marantz 1993 for the concept of fusion). Fusion is a set-building operation which unifies the features of the two case slots into one.[14] Several instances of identical values are reduced to a single instance of that value. The results of fusion are shown in (34). To the left of the fusion arrow, it is shown which abstract cases (are supposed to) stack; to the right, the resulting feature set is indicated. The VI from (33) that realizes this fused feature structure is given in brackets.

(34) a. *genitive + dative:*
[+obl,–obj] + [ ] $\xrightarrow{\textit{fusion}}$ [+obl,–obj] (=gen. VI)

b. *genitive + nominative:*
[+obl,–obj] + [–obl,–obj] $\xrightarrow{\textit{fusion}}$ [+obl,–obl,–obj] (=gen. VI)

c. *genitive + genitive:*
[+obl,–obj] + [+obl,–obj] $\xrightarrow{\textit{fusion}}$ [+obl,–obj] (=gen. VI)

d. *genitive + accusative:*
[+obl,–obj] + [–obl,+obj] $\xrightarrow{\textit{fusion}}$ [+obl,–obl,+obj,–obj] (=abl. VI)

e. *nominative + accusative:*
[–obl,–obj] + [–obl,+obj] $\xrightarrow{\textit{fusion}}$ [–obl,–obj,+obj] (=acc. VI)

---

[14] Note that fusion as we use it here is a bit different from the concept in the literature. Fusion is standardly applied to two terminal nodes (that are sister nodes) and fuses them into a single node which contains all the features of the two original nodes. As a consequence, only a single VI can be inserted into the newly created terminal. In the present analysis, two feature sets on a *single* terminal node fuse into one feature set. In contrast to the standard definition, we take fusion to be a set-building operation. Therefore, not all the features of the two sets are part of the fused set; identical features are reduced to a single instance of that feature.

We begin with the trivial case in which the D head of a possessor that has been assigned genitive in the DP is to be assigned an abstract semantic case from an external head in addition; this semantic case is the dative in (34-a), but the result would be the same for any other semantic case. Recall that only structural cases can stack in the syntax and thus fuse in the morphological component. Any stacking of an abstract semantic case and the genitive is excluded already in the syntax (cf. the derivation in (31)). The D head of the possessor has only one valued case slot in such a context. Thus, fusion of the genitive feature set with an empty set trivially results in the feature structure of the genitive [+obl, –obj]. The most specific matching VI for this feature set is the genitive VI.

Next, consider the fusion of abstract nominative and abstract genitive in (34-b). This happens, for example, if a DP that contains a possessor is merged as the external argument of a transitive verb and is assigned nominative from T (cf. the derivation in (29)). Since the nominative contributes only negative feature values to the fused feature set, and since VIs only spell out positive feature values, the nominative will never have an impact on the realization of the case features. The most specific matching VI for the fused set [+obl,–obj,–obj] is thus determined by the positive feature from the abstract genitive alone; it is again the genitive VI.

In the case of multiple possessors, all possessors except for the structurally highest one end up with genitive-genitive stacking (cf. (32)). Since fusion is a set-building operation, fusion of the representation of the genitive with an identical feature set results in a single instance of the representation of the genitive, which is of course also realized by the genitive VI, cf. (34-c).

So far, the possessor always bears the genitive case marker. The interesting context is the one shown in (34-d): abstract accusative and abstract genitive stack. This happens for example if a DP containing a possessor is merged as the internal argument of a transitive verb and thus receives accusative from v (cf. the derivations in (30) and in (32)). If the features of the accusative and the genitive case fuse, they create a new feature set which contains both [+obl] and [+obj]. Crucially, these features must be realized by a *semantic* case VI, since these VIs are specified for [+obl, +obj] and are thus the most specific matching VIs. The structural case VIs, including the genitive VI, would also match these features but they are less specific than the semantic case VIs. The question is now which of the semantic VIs realizes the fused feature set. The only matching semantic VI is the ablative VI, which is the most underspecified semantic case marker. The reason is that since two abstract structural cases have been fused, there are no semantic features ([±f, ±g, ...]) in the newly created feature structure; structural cases simply do not bear these semantic features in the first place, they are less complex than the abstract semantic cases. All semantic case VIs except for the ablative are specified for semantic features and are thus not a subset of the fused feature set in (34-d). It is because of the underspecification of the ablative VI for these semantic features (which leads to its default nature) that it shows up in this context.

For the sake of completeness, there is another possible combination of abstract case features that does, however, not involve case stacking on a possessor: nominative and accusative, cf. (34-e). These could potentially stack on the subject DP of a clause embedded under an ECM verb (without nominalization) or on the head of an internal argument of a transitive verb. In the latter case, the internal argument receives accusative from v, filling one of its case slots, and nominative from T, filling the second case slot. As before, the nominative with its negative values does not have any influence on the realization of the accusative. The most specific matching item is the accusative VI, in accordance with the empirical facts.

To conclude, the possessor always bears a genitive marker unless it is assigned accusative in addition to genitive in the syntax; in this case, it bears the ablative marker. Crucially, this ablative marker does not realize the abstract semantic ablative case ([+obl, +obj], [−f, −g, ...]) on the possessor; rather, it realizes the combination [+obl, +obj] that arises due to fusion of abstract genitive and accusative.

Finally, note that the analysis presented above naturally accounts for the case pattern with multiple possessors where only the highest possessor can receive ablative case (cf. the derivation in (32)). Lower possessors will receive genitive twice which leaves no slot left for the accusative case to be assigned, which in turn does not create a context where the ablative marker can be inserted. This context only arises on the structurally highest possessor and thus only this possessor will bear the ablative marker. This analysis crucially relies on the restriction that D heads (and all other case bearing items) cannot have more than two case slots: if they had more, the abstract accusative could spread to more deeply embedded possessors in recursive possessor constructions and these possessors would then be wrongly predicted to bear the ablative marker (see section 5.2 for further discussion of this restriction).

4.4. Interim Conclusion

Under the assumption that D and N heads in Udmurt bear two case slots, Udmurt (sometimes) allows for case stacking in the syntax. Postsyntactic morphological fusion unifies the two case slots (see (34)). If abstract genitive and accusative features are combined, the fused case feature set can only be realized by the ablative marker. In all other combinations of the genitive and another structural case, fusion results in a feature structure that must be realized by the genitive marker. The ablative case on possessors is not an abstract case assigned in the syntax (there are no semantic case features on the possessor that are part of the abstract ablative case). The analysis thus crucially relies on the distinction between abstract and morphological case.

In the present account, case assignment in the syntax is local. The look-ahead problem does not arise because under a case stacking analysis the relevant information about the cases assigned to the possessor DP-internally and DP-externally is present on the possessor; there, it is manipulated by fusion in the postsyntactic morphological component.

5. Discussion of the Consequences

This section provides a discussion of technical and empirical consequences. We start with a discussion of some theoretical issues, especially the cyclicity and locality of syntactic operations. Afterwards we will turn to a discussion of the typology of case stacking predicted by the present account.

5.1. Theoretical Issues

5.1.1. Look-Ahead and Counter-Cyclicity Revisited

In section 2 we have argued that under the two assumptions that (i) case of DPs is determined in syntax only and (ii) all operations apply locally, an analysis of the case split with possessors in Udmurt runs into the problem of look-ahead and counter-cyclicity.

The analysis developed in section 4 overcomes both problems. The problems were avoided by adopting the possibility of case stacking and the postsyntactic morphological operation fusion. Due to case stacking, the information which case the DP containing the possessor receives is available on the possessor as well, in addition to the genitive case. Postsyntactic fusion combined with specific vocabulary insertion rules ensured that the ablative case marker is used when the DP containing the possessor has been assigned accusative.

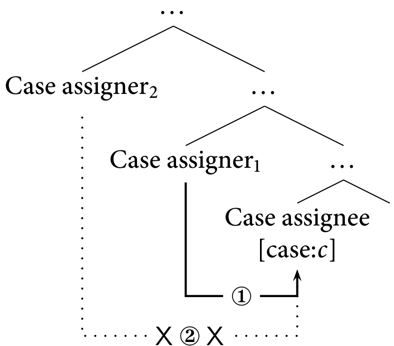
Since in the present approach all the relevant information is locally available on the possessor due to case stacking and since the final determination of the case marker is postponed to the morphological component, the analysis in section 4 does not encounter either look-ahead or counter-cyclicity.

5.1.2. *Locality of Case Assignment without Locality Restrictions*

One further interesting outcome of the analysis in section 4 is that the locality of case assignment is not the result of absolute locality domains, like e.g. phases (Chomsky 2001). Rather, the number of case slots restricts the locality of case assignment.

More concretely, the locality between a case assigner and a case assignee comes about as follows: the case assignee enters the derivation and has a limited number of case slots. Case assigners that enter the derivation shortly after the case assignee – and are consequently representationally close to the case assignee – may enter into a case relation with the case assignee. Case assigners that enter the derivation much later than the case assignee – and are consequently not close to the case assignee – are likely *not* to be able to establish a relation with the case assignee since the case assignee's capacities of entering into case relations are exhausted as soon as all the case slots are filled, see (35).

(35)



Assuming that the case assignee in (35) bears only one case slot, only case assigner$_1$ can establish a case assigning relation with it, since it is the first case assigner to enter the derivation. Case assigner$_2$ comes to late and can therefore not establish a long-distance case assignment relation with the case assignee (see the crossed out arrow in (35)).[15]

Finally, note that even though the present account derives the locality of case assignment without the concept of phases (Chomsky 2001), it is in principle compatible with it because the possessor being in SpecD is visible to the functional heads v and T at all time.[16]

5.1.3. *Morphology as an Autonomous Component of Grammar*

Finally, it should be noted that if the present approach is on the right track, it suggests that morphology has to be a component different from syntax. In the analysis in section 4, the syntax and the morphology have conflicting constraints concerning the number of case slots. While the syntax tolerates the two case slots coming from the lexicon, the morphology has a constraint that prohibits the occurrence of two case slots on one head. Fusion has to apply as a repair mechanism. Thus, it is plausible that the conflict of constraints arises because the syntax and the morphology are different components.

5.2. Cross-Linguistic Variation

In the analysis in section 4, it is assumed that Udmurt exhibits syntactic case stacking even though it does not exhibit overt case stacking like Huallaga Quechua in (27) does. Thus, we take syntactic case stacking to be a very widespread phenomenon that is, however, sometimes disguised by the way morphology realizes stacked abstract cases. In fact, a number of other phenomena have been treated as case stacking as well under the term *Suffixaufnahme*. Suffixaufnahme is the traditional label for case stacking on possessors (cf. Plank 1995 for an overview on Suffixaufnahme). The present analysis predicts a certain range of variation between languages along the following three

---

[15] This is a derivational re-interpretation of Relativized Minimality (Rizzi 1990).

[16] Note that in the present approach, DP cannot be a phase, at least not in Udmurt. If DP was a phase, the NP complement of the D head would already be transfered at the point when the functional heads v or T assign case to the DP. Assuming Multiple Agree, the case slots on the case bearing elements inside the NP could not be filled and the derivation would crash due to the Case Filter. Thus, the assumption of DP being a phase is incompatible with the assumption of case assignment as Multiple Agree. Note further that this is not only a problem of Multiple Agree, but a problem of case concord in general: assuming that DPs are phases and that case is assigned by a DP-external head, the elements in the complement of D should be able to receive case after they have been transfered, which is not possible.

parameters: (i) a language has syntactic case stacking or not; (ii) a language has overt (morphological) case stacking or not; (iii) there are (no) restrictions on the number of cases that can stack (syntactically or overtly). We will show that examples for all these language types exist and that, consequently, the pattern in Udmurt presents just one of various possible realization strategies. We take this state of affairs as indirect evidence for the case stacking analysis of the Udmurt case split.

Parameter (i) is trivial: a language may or may not have syntactic case stacking. If it does not, then there can of course be no overt case stacking. We simply expect the realization of the single case value on a terminal. What is of interest are languages with syntactic case stacking. The question is whether such languages can stack cases overtly, this is parameter (ii) on the morphological realization of case stacking:

(36) *Parameter (ii) on the realization of abstract case values:*[17]

a. Realization of all cases: overt *case stacking*; e.g. Huallaga Quechua.

b. Realization of only one case:
  (i) *Case attraction*: the case value that is assigned last is realized; e.g. Rithangu.
  (ii) *Allomorphy*: a portmanteau morpheme realizes all abstract cases at once; e.g. Udmurt, Beztha (and other Daghestanian languages).
  (iii) *First case*: the case that is assigned first to an element is realized; e.g. languages without case stacking that do neither apply the allomorphy nor the case attraction strategy, e.g. German.
  (iv) *Phonological repair*:
  Phonologically identical case markers are not tolerated; e.g. Jiwarli, Old Georgian, Dyirbal.

If all of the abstract case values are realized by an overt case marker, a language is said to have case stacking or Suffixaufnahme (cf. the strategy in (36-a)). Huallaga Quechua (see (27)) is such a language. If, however, a language does not allow for the realization of more than one case marker, it can choose from

---

[17] See Corbett (1995), Moravcsik (1995) for a similar though not identical typology of case stacking in the DP.

among four different strategies to fulfill the morphological restriction to a single case marker.

The first strategy (cf. (36-b-i)) is reminiscent of case attraction: it is only the case value that is assigned last to an element that is morphologically realized. This pattern is found e.g. in Rithangu (Pama-Nyungan, Schweiger 1995: 354f.). The genitive case morpheme of the possessor is replaced by the case of the head noun if the latter is ablative, locative, allative or pergressive. If the head noun bears a different case, only the genitive is realized on the possessor, cf. the data in (37).

(37) *Rithangu, case attration in the DP*:

a. n̠u-ŋu d̠awal
2SG-GEN country.NOM
‘your country’

b. waːn-i+nu+ra n̠uː-kala-liʔ d̠awal-liʔ
go-FUT+NOW+1SG 2SG-LIG-ALL country-ALL
‘I will now go to your country.’ *head=allative*

c. yaka-nʔ-gu+ña̅+ra l̥a-na mad̠al̥uŋgu-y
this=AUG=GEN=3SG=ACC=1SG spear-PAST hook=spear-INS
‘I speared him with this [man’s] hook spear.’ *head=instrumental*

The second strategy is the allomorphy strategy in (36-b-ii): the stacked cases are realized by a marker M that does not correspond to any of the morphemes that would realize each of the abstract cases alone. Rather, it seems to be the case that the marker M is a kind of portmanteau morpheme that realizes all stacked cases at once (cf. Moravcsik 1995: 462 for the term portmanteau in this context). This is the situation we find in Udmurt: if abstract genitive and accusative stack, we find a morpheme that does neither correspond to the genitive nor to the accusative VI, but another marker (which, in Udmurt, is identical to the marker of the abstract ablative). Indeed, the Udmurt pattern has been described as a special case of case stacking in the literature (cf. Corbett 1995, Kibrik 1995, Moravcsik 1995).

Another possible strategy of languages that have syntactic but not overt case stacking is given in (36-b-iii): the case that is assigned first to an element is realized; the cases that are assigned later are ignored for morphological realization which is the reverse of the case attraction strategy. Any language without overt stacking that neither applies the case attraction nor the allomorphy strategy can be described in this way, e.g. German. Note that an ambiguity arises: it cannot be detected on the surface that these languages have syntactic case stacking. All of these languages could also be described by saying that they do not have syntactic case stacking in the first place, i.e., that only a single case value can be assigned to an element.

Finally, there are languages that do not have a morphological restriction on the number of cases that can be realized but a phonological restriction (cf. strategy (36-b-iv)): in some languages, case stacking is possible, but if two stacked case morphemes on the possessor are phonologically identical, one of them is deleted. This is the case in Jiwarli (Pama-Nyungan, Austin 1995), Old Georgian (Kartvelian, Boeder 1995: 182) and Dyirbal (Pama-Nyungan, Schweiger 1995); see Dench and Evans 1988 for further examples.

The last parameter concerns the number of cases that can stack:

(38) *Parameter (iii) on the number of cases that can stack*:
   a. *Number of cases limited*:
      (i) *limited to one*:
          no case stacking or allomorphy / case attraction strategy
      (ii) *limited to two*:
          in Kanyara and Mantharta languages (West Australia) only two cases can stack.
          ...
   b. *Number of cases unlimited*: e.g. Martuthunira.

Recall that we assumed that in Udmurt only two cases can stack in the syntax (see footnote 10 for discussion.) This is a stipulation, but apart from the fact that it makes correct predictions about the distribution of the ablative marker in structures with recursive possessors, it can be justified by the following fact: languages with overt case stacking also have restrictions on the number of cases that can stack. In Kanyara and Mantharta languages (West Australia, Austin 1995), for example, the number of case markers that can stack overtly is limited to two. Hence, language-specific restrictions on the number of case slots seem to be unavoidable anyway. In Martuthunira (Pama-Nyungan, Corbett 2006: 135), however, the number of cases that can stack is unlimited, cf. (39).

(39) *Case stacking in Martuthunira*:
Ngayu nhawu-lha [ngurnu tharnta-a [mirtily-marta-a
1SG.NOM see-PST that.ACC euro-ACC joey-PROP-ACC
[thara-ngka-marta-a]]]
pouch-LOC-PROP-ACC
'I saw that euro (hill kangaroo) with a joey (young kangaroo) in (its) pouch.'

To summarize, linguistic variation reduces to (a) variation in the morphological realization of syntactic case stacking and (b) a lexical restriction on the number of cases that can stack (overtly or syntactically). Given these parameters, Udmurt exhibits just one of the expected repair strategies that apply when a language has syntactic case stacking but only a single morphological case slot. The present analysis is a formal implementation of the intuition found in the typological literature that the Udmurt pattern (the allomorphy strategy) is indeed a special case of case stacking in which the two cases are expressed by a single lexical item. This view seems to be on the right track given that in all the languages with the Udmurt pattern that we know of (in particular the Daghestanian languages like Bezhta in (2)) the case split depends on the case of the possessor (genitive) plus the case it is assigned by an external head.

6. Conclusion

Udmurt exhibits a case split: possessors bear either a genitive or an ablative case suffix. These cases are in complementary distribution. Traditionally, the case split in Udmurt is described as being driven by the GF of the XP containing the possessor, defined via its position in the syntactic structure in minimalism. The choice of the possessor case in the DP thus seems to require look-ahead: at the point of case assignment to the possessor in the DP this DP is not yet merged with an external head and therefore, its GF is not yet determinable. We have argued that the case split does not depend on GFs; rather, it is determined by the case value that the DP containing the possessor is assigned. This new generalization facilitates a local reanalysis in terms of case stacking: the possessor is always assigned genitive in the DP and it may in addition be assigned another structural case from the external head which selects the DP (assignment of a semantic case in addition to the genitive is *bled* by a restriction on the number of case slots on D and N heads and the different timing of case assignment from DP-internal and DP-external heads that follows from cyclicity). Since there is only a single slot for a case marker in Udmurt, the two case features fuse into a single feature structure in the postsyntactic morphological component. Syntactic case stacking thus *feeds* postsyntactic fusion. Only in case of a combination of genitive and accusative does a feature structure arise which is realized by the default semantic VI, the ablative exponent. There is never abstract syntactic ablative case on the possessor. This analysis does neither require look-ahead nor counter-cyclic case assignment at any point of the derivation. Independent motivation for the case stacking analysis comes from cross-linguistic variation: Udmurt simply uses one of the various expected strategies to resolve the conflict that arises when several abstract cases 'compete' for a single morphological case slot. Some languages realize both cases (overt case stacking), some only one of them and others, like Udmurt, fuse the abstract cases. The resulting case feature is realized by an exponent that may be different from the exponents that would have realized each of the two original case values, thus creating the illusion that the possessor is sometimes assigned genitive and sometimes ablative in the syntax, although it is never assigned abstract ablative.

References

Abondolo, Daniel (1998): Finnish. *In:* D. Abondolo, ed., *The Uralic Languages*. Routledge, pp. 149–183.

Arkadiev, Peter M. (2006): Two-Term Case Systems in Cross-Linguistic Perspective. *In:* B. Gyuris, ed., *Proceedings of The First Central European Student Conference In Linguistics*. Hungarian Academy of Sciences, Budapest.

Austin, Peter (1995): Double Case Marking in Kanyara and Mantharta Languages, Western Australia. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 363–379.

Béjar, Susana and Diane Massam (1999): 'Multiple Case Checking', *Syntax* 2(2), 65–79.

Bierwisch, Manfred (1967): Syntactic Features in Morphology: General Problems of So-Called Pronominal Inflection in German. In: *To Honour Roman Jakobson*. Mouton, The Hague, pp. 239–270.

Bjorkman, Bronwyn (to appear): The Syntax of Syncretism. Proceedings of NELS 40.

Boeder, Winfried (1995): Suffixaufnahme in Kartvelian. *In:* F. Plank, ed., *Double*

Case: Agreement by Suffixaufnahme. Oxford University Press, New York, pp. 151–215.

Caha, Pavel (2008): The Case Hierarchy as Functional Sequence. *In:* M. Richards and A. Malchukov, eds, *Scales*. Universität Leipzig, Institut für Linguistik, pp. 247–276. Vol. 86 of Linguistische Arbeitsberichte.

Caha, Pavel (2009): The Nanosyntax of Case. PhD thesis, CASTL, Tromsø.

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

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and 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.

Corbett, Greville G. (1995): Slavonic's Closest Approach to Suffixaufnahme: The Possessive Adjective. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 265–282.

Corbett, Greville G. (2006): *Agreement*. Cambridge University Press, Cambridge.

Csúcs, Sándor (1988): Udmurt. *In:* D. Abondolo, ed., *The Uralic Languages*. Routledge, London, pp. 276–304.

Deal, Amy Rose (to appear): Possessor Raising. Linguistic Inquiry.

Dench, Alan and Nicholas Evans (1988): 'Multiple Case Marking in Australian Languages', *Australian Journal of Linguistics* 8, 1–47.

Edygarova, Svetlana (2009): 'Attributive Possession in Udmurt Language', *Linguistica Uralica* XLV, 101–118.

Edygarova, Svetlana (2010): Kategoria Possessivnosti v Udmurtskom Jazyke. PhD thesis, University of Tartu, Estonia. Dissertationes Philologiae Uralicae Universitatis Tartuensis 7.

Erlewine, Michael Yoshitaka (2012): Dissociating the Syntax and Morphological Realization of Kaqchikel Agent Focus. Ms., MIT.

Frampton, John and Sam Gutman (2006): How Sentences Grow in the Mind: Agreement and Selection in Efficient Minimalist Syntax. *In:* C. Boeckx, ed., *Agreement Systems*. Benjamins, Amsterdam, pp. 121–157. Linguistik aktuell 92.

Halle, Morris (1997): 'Distributed Morphology: Impoverishment and Fission', *MIT Working Papers in Linguistics* 20, 425–449.

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

Halle, Morris and Alec Marantz (1994): Some Key Features of Distributed Morphology. *In:* A. Carnie, H. Harley and T. Bures, eds, *Papers on Phonology and Morphology*. Vol. 21 of *MIT Working Papers in Linguistics*, MITWPL, Cambridge, Mass., pp. 275–288.

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

Harley, Heidi and Rolf Noyer (1999): 'Distributed Morphology. State-of-the-Article', *Glot International* 4(4), 3–9.

Hiraiwa, Ken (2001): Multiple Agree and the Defective Intervention Constraint in Japanese. *In:* O. Matushansky, A. Costa, J. Martin-Gonzalez, L. Nathan and A. Szczegielniak, eds, *Proceedings of the HUMIT 2000*. University of California, Los Angeles, pp. 67–80. Vol. 40 of *UCLA Working Papers in Linguistics*.

Hole, Daniel (2008): 'Dativ, Bindung und Diathese', Habilitationsschrift, Humboldt-Universität zu Berlin.

Kel'makov, Valentin K. (1999): *Udmurtin kielioppia ja harjoituksia*. Suomalais-Ugrilainen Seura, Helsinki.

Kibrik, Aleksandr E. (1995): Direct-Oblique Agreement of Attributes in Daghetanian. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 216–229.

König, Ekkehard and Martin Haspelmath (1998): Les constructions à possesseur externe dans les langues d'Europe. *In:* J. Feuillet, ed., *Actance et valence dans les langues de l'Europe*. Mouton de Gruyter, Berlin, pp. 525–606.

Koptjevskaja-Tamm, Maria (2003): Possessive Noun Phrases in the Languages of Europe. *In:* F. Plank, ed., *Empirical Approaches to Language Typology*. de Gruyter, Berlin/New York, pp. 621–722.

Legate, Julie Anne (2008): 'Morphological and Abstract Case', *Linguistic Inquiry* 39(1), 55–101.

Matushansky, Ora (2008): Predication: A Case Study. *In:* F. Marušic and R. Žaucer, eds, *Studies in Formal Slavic Linguistics. Contributions from Formal Description of Slavic Languages 6.5*. Peter Lang, Frankfurt am Main, pp. 213–239.

Moravcsik, Edith A. (1995): Summing up Suffixaufnahme. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 451–484.

Nevins, Andrew (2003): Do Person/Number Syncretisms Refer to Negative Values?. Handout of a talk given at the LSA meeting 2003, lingbuzz/000104.

Nikolaeva, Irina (2002): The Hungarian External Possessor in a European Perspective. *In:* C. Hasselblatt and R. Blokland, eds, *Finno-Ugrians and Indo-Europeans: Linguistic and literary contacts*. Shaker, Maastricht, pp. 272–285.

Noyer, Rolf (1992): Features, Positions and Affixes in Autonomous Morphological structure. PhD thesis, Massachusetts Institute of Technology, Cambridge, Mass.

Plank, Frans (1995): (Re-)Introducing Suffixaufnahme. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 3–110.

Pylkkänen, Liina (2002): Introducing Arguments. PhD thesis, MIT, Cambridge, Mass.

Řezáč, Milan (2004): Elements of Cyclic Syntax: Agree and Merge. PhD thesis, University of Toronto.

Richards, Marc (2008): Quirky Expletives. *In:* R. d'Alessandro, G. Hrafnbjargarson and S. S.Fischer, eds, *Agreement Restrictions*. Mouton de Gruyter, Berlin, pp. 181–213.

Rizzi, Luigi (1990): *Relativized Minimality*. Vol. 16 of *Linguistic Inquiry Monographs*, MIT Press, Cambridge, Massachusetts.

Schoorlemmer, Erik (2009): *Agreement, Dominance and Doubling. The Morphosyntax of DP*. LOT, Utrecht.

Schweiger, Fritz (1995): Suffixaufnahme and Related Case Marking Patterns in Australian Languages. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 339–362.

Suhonen, Seppo (1988): Geschichte der Ostseefinnischen Sprachen. *In:* D. Sinor, ed., *The Uralic Languages*. Brill, Leiden/New York/København/Köln, pp. 288–313.

Suihkonen, Pirkko (2005): On the Categories and Functions Developed from the Possessive and Deictic Suffixes in Udmurt. *In:* C. Hasselblatt, E. Koponen and A. Widmer, eds, *Lihkun lehkos! Beiträge zur Finnougristik aus Anlaß des sechzigsten Geburtstages von Hans-Hermann Bartens*. Harrassowitz Verlag, pp. 401–431.

Szabolcsi, Anna (1984): 'The Possessor that Ran Away from Home', *The Linguistic Review* 3, 89–102.

Szabolcsi, Anna (1994): The Noun Phrase. *In:* F. Kiefer, ed., *Syntax and Semantics, vol 27: The Structure of Hungarian*. Academic Press, San Diego, pp. 179–274.

Vilkuna, Maria (1997): Word Order in European Uralic. *In:* A. Siewierska, ed., *Constituent Order in the Languages of Europe*. Mouton de Gruyter, Berlin, pp. 173–233.

Winkler, Eberhard (2001): *Udmurt*. Languages of the world: Materials, Vol. 212, LINCOM Europa, München.

Wunderlich, Dieter (1996): Minimalist Morphology: The Role of Paradigms. *In:* G. Booij and J. van Marle, eds, *Yearbook of Morphology*. Kluwer, Dordrecht, pp. 93–114.

Zwicky, Arnold (1977): Hierarchies of Person. *In:* W. Beach and S. Fox, eds, *Papers from the 13th Chicago Linguistic Society*. CLS, pp. 714–733.

Rule Ordering in Verb Cluster Formation: On the Extraposition Paradox and the Placement of the Infinitival Particle *te/zu*

Martin Salzmann[*]

Abstract

This paper addresses two puzzles in the domain of verb cluster formation and proposes a solution in terms of rule ordering. The first puzzle is the so-called extraposition paradox where extraposition can target a VP that is part of a verb cluster only if the VP is topicalized but not when the VP remains clause-final. I propose that verb cluster formation takes place at PF under adjacency and thus after extraposition and topicalization. Extraposition and topicalization can therefore bleed cluster formation, leading to a crash of the derivation if the VP remains in-situ. The second puzzle involves the placement of the infinitival marker *te/zu* in Dutch and German. I will show that the cross-linguistic differences in placement follow from the fact that the rule that associates the particle with the verb takes place at different points of the PF-derivation in the two languages. While it is an early operation in Dutch and is still sensitive to hierarchical structure, it is a late process in German and is therefore subject to linear order and adjacency. Both operations interact with other PF rules, and I will demonstrate that it is possible to determine a strict and non-contradictory (and predominantly intrinsic) ordering of the rules which as a side-effect provides evidence for the articulation of the PF-component. Finally, I will show that the *zu*-placement facts do not provide decisive evidence in favor of either a right-branching or a left-branching VP-structure; rather, the advantages and disadvantages of the two views turn out to largely balance each other out.

1. Introduction

While the PF-branch of grammar was for a long time kind of the syntacticians waste basket that hosted syntactic phenomena (e.g. stylistic rules) that could

---

[*] An earlier version of this work was presented at the workshop Interaction of syntactic primitives at the Annual Meeting of the DGFS 2013 in Potsdam. I thank the audience at that occasion for helpful discussion. This research has been supported by grant PA00P1_136379/1 from the Swiss National Science Foundation.

Rule Interaction in Grammar, 65–121
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

not so easily be accommodated within the standard assumptions of syntactic theory (even though many of them seemed rather syntactic, e.g. locative inversion), the last two decades have seen a shift towards serious investigations into the structure of the PF-component. On the one hand, this was triggered by developments within syntactic theory that attempted to circumscribe more narrowly the operations that syntax proper is supposed to perform. On the other hand, the introduction of Distributed Morphology in Halle and Marantz (1993) radically changed the view on the morphology-syntax interface and opened new prospects for post-syntactic operations. An important point in that development was the proposal by Embick and Noyer (2001) who provided convincing evidence for a division of the PF-component into at least two subcomponents. The division was empirically motivated on the basis of post-syntactic movement operations. Since the PF-component gradually transforms hierarchical syntactic structure into a linear structure that can be interpreted at the interface, movement operations that apply early in the PF-branch will tend to be sensitive to hierarchical structure while later movement operations will be sensitive to linear properties of the structure such as adjacency. Operations of the first type are termed Lowering, which basically amounts to downward head-movement; operations of the second-type are called Local Dislocation. Both operations can be responsible for the placement of clitics and affixes that surface in a position different from the phrase marker whose terminal nodes they realize. A very recent contribution to the research into the PF-component is Arregi and Nevins (2012) who provide further evidence for a highly articulated post-syntactic component and pervasive interaction of post-syntactic operations instantiating classic feeding, bleeding and opaque relationships.

The goal of this paper is a modest attempt to contribute to this discussion by examining two phenomena from West-Germanic syntax that lend themselves to a post-syntactic treatment, viz. cluster formation and the placement of the infinitival particle. As I will argue, a full account requires the postulation of a number of post-syntactic operations that need to apply in a certain order. This rule interaction is then used as a diagnostic: If it can be demonstrated that a larger number of such rules can be ordered in a non-contradictory way, we have made progress towards an understanding of the structure of the PF-component.

The paper is organized as follows: In section 2, I will discuss an extraposition paradox and argue that verb cluster formation takes place post-syntactically. Based on these findings, in section 3, I will investigate the placement of the infinitival particle in German and Dutch and propose solutions to account for the cross-linguistic variation as well as for the intricate placement pattern in German and its dialects. Section 4 concludes.

2. The Extraposition Paradox

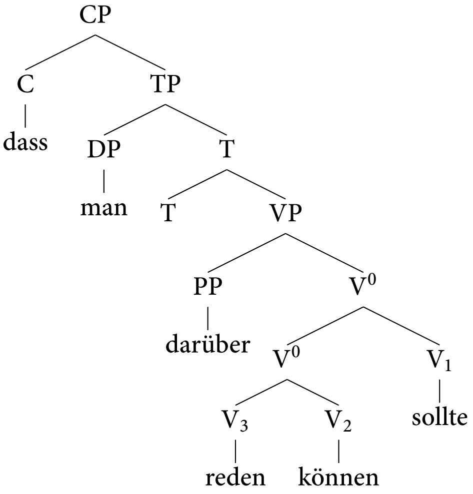
2.1. The Problem

One prominent feature of West-Germanic OV-languages like Dutch and German is the clustering of verbal elements at the end of the clause in V-final structures, as in the following example (under verb second, where the finite verb moves to C, only the non-finite verbs occur together):[1]

(1) 321 *Standard German*
dass man darüber [reden$_3$ können$_2$ sollte$_1$]
that one about.it talk.INF can.INF should.3SG
'that one should be able to talk about it'

Such sequences are usually referred to as verb clusters (for a detailed overview, cf. Wurmbrand 2005); furthermore, there is a long tradition (starting with Evers 1975) that analyzes verb clusters as complex syntactic heads, e.g. as follows:

(2)



---

[1] Numbers on verbs indicate the embedding relations, i.e. 1 stands for the highest, i.e. the embedding verb, 2 for the immediately embedded verb etc.

The major empirical evidence for a complex head comes from the following observation: Extraposition, an operation which targets VP in German, cannot target VP3 or VP2 in a V-final structure involving a verb cluster (cf. van Riemsdijk 1998: 640ff., Haider 2003: 92ff., Bayer et al. 2005: 91):[2]

(3) *Standard German*

dass man $[_{\text{VP}_1}$ $[_{\text{VP}_1}$ $[_{\text{VP}_2}$ $[_{\text{VP}_2}$ $[_{\text{VP}_3}$ $[_{\text{VP}_3}$ $\text{t}_{darüber}$ reden$_3$] ***darüber**]
that one talk.INF about.it
können$_2$] ***darüber**] sollte$_1$] **darüber**]
can.INF about.it should about.it
'that one should be able to talk about it'

This restriction follows directly under a complex head analysis since XP-movement cannot target segments of V, but only the maximal projection. The real challenge obtains under topicalization. Suddenly, extraposition *can* target VP3:

(4) *Standard German*

$[_{\text{VP}_3}$ $[_{\text{VP}_3}$ $\text{t}_{darüber}$ reden$_3$] **darüber**] sollte$_1$ man schon $[_{\text{VP}_1}$ $[_{\text{VP}_2}$
talk.INF about.it should.3SG one indeed
t_{VP_3} können$_2$] t_{sollte}]
can.INF
'that one should be able to talk about it'

In other words, (4) seems to be derived from an ungrammatical structure. There are two possibilities to resolve this tension: Either cluster formation is taken to be optional or excorporation is allowed.

In the first case, cluster formation (in the sense of forming a complex head) would not take place in the V-final structure so that extraposition can target VP3, thus providing a base for (4). However, once cluster formation is taken to be optional, it is no longer clear how the ungrammatical versions of (3) can be

---

[2] I adopt a left-branching VP-structure in what follows; the exact relationship between syntax and linear order is discussed in more detail in section 3.4.1; the consequences that arise once a right-branching VP-structure is adopted are discussed in section 3.7

For the purpose of this paper, I will label all verbal elements involved in verb clusters as lexical verbs, i.e. including modals, auxiliaries and others even though there may be reasons to classify some of them as functional elements (containing more structure), cf. Wurmbrand (2004*a*); as far as I can tell, nothing in what follows hinges on this.

derived. We are thus faced with a paradox: cluster formation would have to be optional and obligatory at the same time. The alternative, excorporation, does not fare much better: To derive (4), both V_1 and and V_2 would have to excorporate before topicalization takes place. While excorporation of V_1 may in principle be plausible as it moves to C, this is not the case with V_2, as it is generally assumed that the verbs stay in their base position in V-final structures.

One can thus conclude that neither excorporation nor cluster formation is sufficient to account for the pattern in (3) and (4). Interestingly, apart from Wurmbrand (2007), the issue has not been addressed in much detail. Haider (2003) assumes that the topicalization structure and the V-final structure are not transformationally related; rather, both the topicalized VP and the (partial) clause-final complex head are independently base-generated as such. Unfortunately, he does not spell-out how the two are related to each other: if the VP undergoes long-distance movement or occurs with a 4-verb-cluster, there will be at least part of a complex head clause-finally:

(5) *Standard German*
[$_{VP_4}$ [$_{VP_4}$ $t_{darüber}$ reden$_4$] darüber] sollte$_1$ man schon [$_{VP}$ können$_3$
talk.INF about.it should one indeed can.INF
wollen$_2$ t_{sollte}].
want.INF
'One should want to be able to talk about it.'

The interpretation of the topicalized constituent as a complement of only a part of that complex head (i.e. V_3 *können* 'can') is certainly non-trivial (like other interpretive aspects such as adverbial modification, see Wurmbrand 2007 for critical discussion). A tentative solution is sketched in Haider (2010: 307).[3] Bader and Schmid (2009: 202, fn.11), who assume that verb clusters are base-generated complex heads, admit that for cases like (4), it may be necessary that V_1 actually selects a VP and not a V^0 as they assume elsewhere. But once this possibility is granted, the cluster property in (3) cannot be derived anymore.

---

[3] Note that such examples cannot be reanalyzed as cases of left-dislocation with deletion of the fronted proform (that would anaphorically refer to the fronted VP) because topicalization and left-dislocation do not always pattern the same. As pointed out in Haider (2003: 95f.), once lexical specifications of the verb (oblique cases, *wh*-complements) are involved, left-dislocation becomes impossible while topicalization does not.

Haegeman and van Riemsdijk (1986: 451), who assume that verb clusters result from reanalysis, simply mention that verb-clusters permit extraction and thus are not "lexical" in the sense of being impenetrable; the cluster paradox thus seems to remain unaccounted for under their approach as well. It is therefore fair to conclude that approaches adopting a complex syntactic head (either base-generated or derived in syntax) cannot resolve the contradiction between (3) and (4) in a straightforward way.[4]

2.2. The Solution: Timing

I would like to propose a solution to the extraposition paradox that makes crucial use of timing. I adopt the standard assumptions that extraposition, V-to-C-movement and topicalization take place in syntax. Where I differ from much of the literature is that I assume that verb cluster formation in the sense of forming a complex head takes place at PF under adjacency and thus arguably represents an instance of what Embick and Noyer (2001) have referred to as Local Dislocation. Importantly, what is inverted are not syntactic sisters (e.g. V_1 and VP_2 as e.g. in Haegeman and van Riemsdijk 1986, Wurmbrand 2004*b*). Since the empirical evidence for this position is discussed in much detail in Salzmann (to appear), I will not reproduce the arguments here but will simply take this as given.[5] Under these assumptions, the extraposition paradox can be accounted for as follows:

I will start with the cluster property of the V-final structure in (3). There are no complex heads in the syntax but stacked VPs instead. Extraposition can therefore in principle target either VP_3, VP_2 or VP_1, as indicated in the following structure:

---

[4] See Wurmbrand (2007) for an explanation of the paradox in (3) and (4) without the postulation of a complex head. She argues instead that extraposition is subject to prosodic restrictions which force extraposed constituents to appear at the edges of prosodic constituents. In a V-final structure, this is only the case if the extraposee occurs adjoined to VP_1 while under topicalization, VP_3 constitutes a separate prosodic domain and thus constitutes a legitimate attachment site.

[5] Adjacency constraints on verb cluster formation have, of course, been proposed before, cf. e.g. van Riemsdijk (1998: 639-645) where cluster formation involves syntactic head-movement. However, given that head-movement is normally not subject to such a constraint, an adjacency condition seems stipulative. Placing cluster formation as an adjacency-sensitive operation in the post-syntactic component seems more in line with current conceptions about the architecture of grammar.

(6) *Standard German*

dass man [$_{\text{VP}_1}$ [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ [$_{\text{VP}_3}$ t$_{\textit{darüber}}$ reden$_3$] **darüber**]
that one talk.INF about.it
können$_2$] **darüber**] sollte$_1$] **darüber**]
can.INF about.it should about.it
'that one should be able to talk about it'

I will assume that cluster formation is obligatory if the verbal elements are adjacent at PF, at least in descending order, cf. fn. 17 for details. Under a left-branching VP-structure, cluster formation will be string-vacuous in descending orders (which amounts to re-bracketing); under a right-branching VP-structure as discussed in section 3.7 (and also in 312 clusters based on left-branching structures, cf. section 3.4.2), cluster formation will additionally involve reordering/inversion. Technically, I assume that clustering is enforced by a surface constraint. In (6), cluster formation will be blocked if extraposition targets VP$_3$ or VP$_2$ as the extraposed PP will destroy adjacency between the verbal elements. Put differently, extraposition to VP$_3$ or VP$_2$ *bleeds* cluster formation. In this case, the derivation crashes because the surface constraint requiring the formation of a complex head is violated. Only extraposition to VP$_1$ is an option. This derives (3).

In the topicalization structure in (4), extraposition targets VP$_3$. This is licit since cluster formation does not take place at the point where VP$_3$ is still in its base-position. Rather, topicalization of VP$_3$ destroys the context for cluster formation: there is only one verbal element in the prefield so that no cluster formation needs to take place (the two remaining verbal elements at the end of the clause, however, do undergo cluster formation). In this case, topicalization also bleeds cluster formation, but the result is grammatical because no surface/PF-constraint is violated.

This approach based on timing of operations has an additional advantage. It directly explains why verb-second movement never involves a complex head: the element fronted to C is always just a single verb (the finite verb), complex verbs are ruled out:

(7) *Standard German*

a. *Man [reden$_3$ können$_2$ sollte$_1$] darüber schon.
one talk.INF can.INF should about.it indeed
'One should be able to talk about it.'

b. Man sollte$_1$ darüber schon [reden$_3$ können$_2$ t$_{sollte}$].
one should about.it indeed talk.INF can.INF
'One should be able to talk about it.'

Under the present approach, this follows automatically since at the point where verb second movement takes place, the verbs have not yet formed a cluster (i.e. like topicalization, verb second movement bleeds cluster formation). Consequently, only V$_1$ will move. Approaches that assume the formation of a complex head in syntax and relate V-final and V-second structures via movement have to make extra assumptions to rule out movement of the complex head (such as e.g. the complexity constraint in Neeleman and Weerman 1993: 460ff.).

To summarize: in this section, I have discussed the following rules/operations: extraposition, topicalization, and verb-cluster formation. They apply in the following order:

(8) extraposition > topicalization > verb-cluster formation

The ordering between the first two operations is intrinsic as it follows from cyclicity, e.g. as formulated in the *Strict Cycle Condition* (Chomsky 1973) or the Extension Condition (Chomsky 1995). The ordering between topicalization and cluster formation is also intrinsic as it results from the fact that the two operations take place in different components of the grammar that are sequentially ordered, viz. syntax vs. PF (which can thus also be regarded as two separate cycles). In two configurations, application of R$_1$ can bleed application of R$_2$: In (3), extraposition bleeds cluster formation, in (4) topicalization bleeds cluster formation.

3. The Placement of the Infinitival Particle *te/zu*

In this section, I will discuss the placement of the infinitival particle in Dutch/German. As we will see, the placement is not fully straightforward, at least in German, where the particle does not always occur in the position expected on the basis of its morphosyntactic properties. The major focus of attention is to pin down the point where the placement occurs. As it is dependent on a number of operations involving the verbal complex, the exact ordering of the various operations will be crucial.

3.1. (Standard) Dutch vs. (Standard) German

Dutch and German differ from each other with respect to the placement of the infinitival particle: While it occurs at the end of the verb cluster in German, it surfaces at the beginning of the cluster in Dutch. In the following example, the matrix verb 'think' takes a non-finite complement clause where the hierarchically highest verb appears with the particle *te/zu*.[6]

(9) a. 321 *Standard German*
Er dachte, das Buch [lesen$_3$ können$_2$ **zu** müssen$_1$].
he thought the book read.INF can.INF to must.INF
'He thought he had to be able to read the book.'

b. 123 *Standard Dutch*
Hij dacht het boek [**te** moeten$_1$ kunnen$_2$ lezen$_3$].
he thought the book to must.INF can.INF read.INF
'He thought he had to be able to read the book.'

Based on these data, the most straightforward analysis seems to be to treat the particle as a prefix, as e.g. proposed in Haider (1993). The particle simply surfaces on the hierarchically highest verb, viz. V$_1$. The surface difference would be the result of independent differences in verb cluster formation; for instance, one might argue (as in the classical analysis by Evers 1975) that cluster formation involves adjunction to the left in German but adjunction to the right in Dutch, leading to reversed, i.e. ascending surface structure in Dutch while the descending order of the base is retained in German. Further evidence for a prefix analysis comes from the following examples where there are two verbs (*versprechen* 'promise' and *versuchen* 'try') that select a *te-/zu*-infinitive. Here, the particle appears twice, on each verb that is dependent on a predicate selecting a *zu*-infinitive:

(10) a. 321 *Standard German*
dass er [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ das Buch **zu** lesen$_3$] **zu** versuchen$_2$]
that he the book to read.INF to try.INF
versprach$_1$]
promised
'that he promised to try to read the book'

---

[6] Lexical verbs selecting a non-finite complement usually take a so-called *te/zu*-infinitive where the infinitival verb is accompanied by the particle while modals and a few other verbs like perception verbs require a so-called bare infinitive, i.e. an infinitival verb without particle.

b. 123 *Standard Dutch*
dat hij [$_{\text{VP}_1}$ beloofde$_1$ [$_{\text{VP}_2}$ **te** proberen$_2$ [$_{\text{VP}_3}$ het boek **te**
that he promised to try.INF the book to
lezen$_3$]]]
read.INF
'that he promised to try to read the book'

However, as we will see in the following subsection, a prefix analysis turns out to be too simple for certain configurations in German.

3.2. Against Prefix Status: Misplaced *zu* in Standard German

In most cases – like those discussed in the previous subsection –, the infinitival particle indeed surfaces before the verb where one expects it to surface, viz. on the hierarchically highest verb of the selected *zu*-infinitive complement. However, there are two instances in the standard language where *zu* seems to occur in the wrong position, that is, to be misplaced. Both involve verb clusters that are partially ascending, i.e. involve 132 or 312 order. The placement of *zu* in non-finite 132 orders was first discussed in Bech (1983). In the following example, the conjunction *ohne* 'without' selects a *zu*-infinitive. The infinitival complement contains a verb cluster with 132 order. Consequently, one would expect the hierarchically highest verb of the cluster, viz. V$_1$, to be preceded by *zu*. However, this structure is ungrammatical. Instead, *zu* surfaces before the last element of the cluster, viz. V$_2$. In the following triple from Standard German, the first example involves *ohne* with a finite complement. The second one is the non-finite version of it with *zu* in the position expected on the basis of its morphosyntactic properties. The third example shows misplaced *zu* before V$_2$ (data from Haider 2011: 227):

(11) a. 132 V$_1$ = finite *Standard German*
ohne dass er es mich [hat$_1$ prüfen$_3$ lassen$_2$]
without that he it me has verify.INF let.INF
lit.: 'without that he let me verify it'

b. 132 V$_1$ = non-finite *Standard German*
*ohne es mich [**zu** haben$_1$ prüfen$_3$ lassen$_2$]
without it me to have.INF verify.INF let.INF
'without having let me verify it'

c. 132 V_1 = non-finite *Standard German*
?ohne es mich [haben$_1$ prüfen$_3$ **zu** lassen$_2$]
without it me have.INF verify.INF to let.INF
'without having let me verify it'

The second case of misplaced *zu* involves what Vogel (2009: 308) referred to as the 'scandal construction'. It involves a non-finite verb cluster with 312-order where *zu* again appears before V_2 instead of V_1. In the following example, the matrix verb 'regret' selects a complement with a *zu*-infinitive:[7]

(12) a. 312 *Standard German*
?Er bedauert, es nicht [verhindert$_3$ haben$_1$ **zu** können$_2$].
He regrets it not prevent.PRT have.INF to can.INF
'He regrets not having been able to prevent it.'

b. 312 *Standard German*
*Er bedauert, es nicht [verhindert$_3$ **zu** haben$_1$ können$_2$].
He regrets it not prevent.PRT to have.INF can.INF
'He regrets not having been able to prevent it.'

Clearly, if *zu* is treated as a prefix, the two cases where it is misplaced cannot be derived. There is no agreement in the literature as to the grammatical status of misplaced *zu*. It is generally acknowledged that the constructions are somewhat marked, but beyond this one can find diametrically opposed views. While Vogel (2009) considers the constructions to be the result of rules of grammar, Bech (1983) treats them as not fully grammatical compromises: they cannot be fully grammatical because they are in conflict with independent principles of German grammar: *zu* should be placed on the hierarchically highest verb; furthermore, the particle has to occur before the last element of the cluster. However, in Aux-Mod-Inf clusters ('has want INF'), the structurally highest verb has to undergo inversion; as a consequence, either *zu* is not prefinal or the hierarchially highest verb fails to bear *zu*. According to Bech, there is no way to resolve this conflict without violating some constraint of German grammar.

---

[7] Another peculiarity of this construction is the participial morphology on V_3. Since V_2 is a modal verb, one would expect V_3 to occur in the bare infinitive (which is, in fact, a grammatical alternative). What seems to have happened is the following: the participial morphology required by V_1 is not realized on V_2 (it actually never is in these clusters; instead, the so-called *Infinitives Pro Participio* occurs); instead, it is displaced to V_3. Similarly, the non-finite *zu* expected to occur on V_1 is displaced to V_2; we are thus dealing with two instances of misplaced morphology.

This view may explain the markedness of (11-c) and (12-a), but it does not explain why (11-b) and (12-b) are completely unacceptable. Haider (2011) goes a step further and regards the constructions as grammatical illusions, i.e. the constructions are essentially ungrammatical but appear to be acceptable. I will not take a definitive stand on this issue with respect to Standard German because it cannot be decided so easily on theoretical or empirical grounds. But as we will see in the next section, once non-standard varieties are taken into account, there is reason to believe that the possibility of misplaced *zu* is part of German grammar even if it only surfaces very residually in the standard language.[8]

3.3. Misplaced *z* in Alemannic Varieties of German

While misplaced *zu* in Standard German seems to be a somewhat marked phenomenon of unclear grammatical status, the empirical situation is different in Alemannic varieties of German: Even though infinitives are less common than in the standard language (and prepositional, finite or non-subordinate structures being used instead), misplaced *zu* (whose form is *z* in these dialects) is nevertheless more visible. This is related to the fact that ascending orders in verb clusters are much more common in Alemannic, and especially Swiss German varieties. As a consequence, the conflict between marking V_1 or the last verb of the verb cluster obtains much more frequently, especially in relatively simple 2-verb clusters. In all ascending orders the particle *z* systematically appears (misplaced) before the last verb of the verb cluster. Examples can be found on the internet but also in traditional grammatical descriptions (suggesting that the phenomenon is definitely not just an invention of the formal grammarian). [9] The last example of the following triple was tested in an informal survey with native speakers of various Swiss German dialects: [10]

---

[8] It should be stressed that the existence of misplaced and displaced morphology in West Germanic languages is beyond doubt given the observations in Höhle (2006) and den Dikken and Hoekstra (1997).

[9] The phenomenon is also discussed in Hodler (1969: 560), Bader (1995: 22), and Cooper (1995: 188f.), who provides a number of examples recorded from Swiss German radio.

[10] Because of the general preference for alternative constructions, the acceptability of misplaced *z* examples will invariably be degraded; furthermore, unlike in finite clusters, non-finite ascending clusters with only two verbs are usually judged much more acceptable than more complex ones. Why this should be the case is a question I have to leave for further research.

(13) a. 1…23 *Zurich German*, cf. Weber (1987: 244,fn.1)
Er schiint$_1$ nüüt [wele$_2$ **z** wüsse$_3$] dervoo.
He seems nothing want.INF to know.INF about.it
'He does not seem to be interested in it.'

b. 12 *Swiss German*[11]
Ich liebe d freiheit, selber de tag [chöne$_1$ **z** bestimme$_2$].
I love the freedom self the day can.INF to determine.INF
'I love the freedom to determine my schedule.'

c. 132 *Swiss German*
er behauptet, s Buech bis am Mëëntig [müse$_1$ gläse$_3$
he claims the book till on.the Monday must.INF read.PRT
z ha$_2$]
to have.INF
'he claims having to have read the book until Monday'

Additionally, a fact that to my knowledge has gone unnoticed so far, misplaced *z* is not limited to constructions where the verbal elements are adjacent, as in the examples discussed up to now (such constructions are also referred to as instances of Verb Raising, VR). It also occurs with so-called Verb Projection Raising (VPR), where the verbal elements are separated by non-verbal material:

(14) 1X2 *Swiss German*[12]
ohni mi [welle$_1$ [uf d bullesite **z** stelle$_2$]], im gegeteil,
without me want.INF on the cops.side to put.INF on.the contrary
aber …
but
'without wanting to side with the cops, on the contrary, but …'

Misplaced *z* is thus arguably less marked than in the standard language and I will regard it as a phenomenon that is part of the grammar of these varieties. For the purpose of this paper, I will assume that misplaced *zu* in Standard German is a grammatical phenomenon as well.

---

[11] http://badoo.com/de-ch/0279246484/, found on March 11, 2013

[12] http://www.fcbforum.ch/forum/showthread.php?4328-usschritige-nachem-spiel-!/page4; found on March 11, 2013

Taking the misplaced *zu* facts into account, the descriptive generalization for the placement of *zu/z* is thus as follows: *zu/z* occurs before the last verbal element of a non-finite complement required to be marked with *zu/z*.[13]

3.4. Proposal

I will pursue two goals in this section: First, I will derive the cross-linguistic variation between German and Dutch in a systematic way. Second, I will provide a coherent account of *zu-/z-*placement in German and its varieties.

3.4.1. Dutch vs. German

The goal of this subsection is to show that the operation that associates the particle with the verb is a similar operation in both languages. The cross-linguistic differences result from the fact that the operation takes place at different stages of the PF-component so that different concepts of headedness play a role: while it is structural headedness in the case of Dutch, it is peripherality within a constituent in the case of German (cf. Embick and Noyer 2001: 562 for these notions of headedness).

3.4.1.1. German

I will make the following assumptions for German: In line with recent work on the role of linear order in syntax, I assume that linearization takes place post-syntactically. More concretely (and more generally), complements of V, both verbal and non-verbal, are linearized to the left of the selecting head.[14] If no further operations apply, the left-branching structure reaches the surface. In verb cluster constructions, PF-operations may alter the syntactic structure, sometimes leading to an ascending order. Second, *zu* occupies a functional head above the VP which for reasons of simplicity I will label F.[15]

As a first step, I will tackle the placement of *zu* in a simple example like (9-a), repated here for convenience:

---

[13] For two rare exceptions, cf. Schallert (2012: 252).

[14] I deviate from Kayne (1994) in that the order of head and complement is established by means of linearization parameters, as e.g. in Richards (2008).

[15] I refrain from glossing it as I/T as in older work because the evidence for a separate IP/TP in German is rather scarce; furthermore, *zu* is also present in structures that are arguably smaller than IP/TP, cf. section 3.4.2 below.

(15) 321 *Standard German*
Er dachte, das Buch [lesen$_3$ können$_2$ **zu** müssen$_1$].
he thought the book read.INF can.INF to must.INF
'He thought he had to be able to read the book.'

The starting point will be a left-branching VP-structure with *zu* occupying a functional head above the highest verbal projection; for expository purposes, I will assume that all material that occurs outside the verb cluster has scrambled to Specv; in the left-branching VR-cases, this is optional (the surface string is usually ambiguous between a scrambled and a non-scrambled structure unless there is additional material like adverbials); in the ascending structures to be dicussed below it is obligatory to evade VP-inversion:

(16) 321 *Standard German*
Er dachte, [$_{vP}$ das Buch [$_{FP}$ [$_{VP_1}$ [$_{VP_2}$ [$_{VP_3}$ t$_{das\ Buch}$ lesen$_3$]
he thought the book read.INF
können$_2$] müssen$_1$] **zu**]].
can.INF must.INF to
'He thought he had to be able to read the book.'

In a next step, *zu* is associated with the final verb of the verb cluster *müssen* 'must'. At this point, there are two possibilities: Either *zu* undergoes Lowering in the sense of Embick and Noyer (2001), i.e. it undergoes downward head-movement. Alternatively, the reordering takes place at some later stage and involves linear reordering as in Local Dislocation. Before we can decide between these two options, we need to look at ascending structures as in the examples with misplaced *zu* in (11-c), (13-b) and (14), repeated here for convenience:

(17) a. 132 V_1 = non-finite *Standard German*
?ohne es mich [haben$_1$ prüfen$_3$ **zu** lassen$_2$]
without it me have.INF verify.INF to let.INF
'without having let me verify it'
b. 12 *Swiss German*
Ich liebe d freiheit, selber de tag [chöne$_1$ **z** bestimme$_2$].
I love the freedom self the day can.INF to determine.INF
'I love the freedom to determine my schedule.'

c. 1X2 *Swiss German*
ohni mi [welle$_1$ [uf d bullesite **z** stelle$_2$]], im
without me want.INF on the cops.side to put.INF on.the
gegeteil, aber ...
contrary but
'without wanting to side with the cops, on the contrary, but ...'

The input for the reordering operations will look as follows:

(18) a. 132, ex. (11-c) *Standard German*
ohne [$_{\text{vP}}$ es mich [$_{\text{FP}}$ [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ t$_{es}$ t$_{mich}$ prüfen$_3$]
without it me verify.INF
lassen$_2$] haben$_1$] zu]]
let.INF have.INF to
b. 12, ex. (13-b) *Swiss German*
... [$_{\text{vP}}$ de Taag [$_{\text{FP}}$ [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ t$_{de\ Taag}$ bestimme$_2$] chöne$_1$] z]]
the day determine.INF can.INF to
c. 12, ex. (14) *Swiss German*
ohni [$_{\text{vP}}$ mi [$_{\text{FP}}$ [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ t$_{mi}$ uf d Bullesiite stelle$_2$]
without me on the cops.side put.INF
wele$_1$] z]]
want.INF to

I will assume that ascending structures come about by means of VP-inversion at PF (as e.g. in Haegeman and van Riemsdijk 1986, Williams 2004), i.e. V$_1$ inverts with VP$_2$. If all lexial material is scrambled out of the lexical VP, ascending Verb Raising structures obtain (cf. Broekhuis 1993). If the non-verbal material does not move, it is affected by VP-inversion so that Verb Projection Raising structures obtain. The three examples above are thus transformed into the following structures:

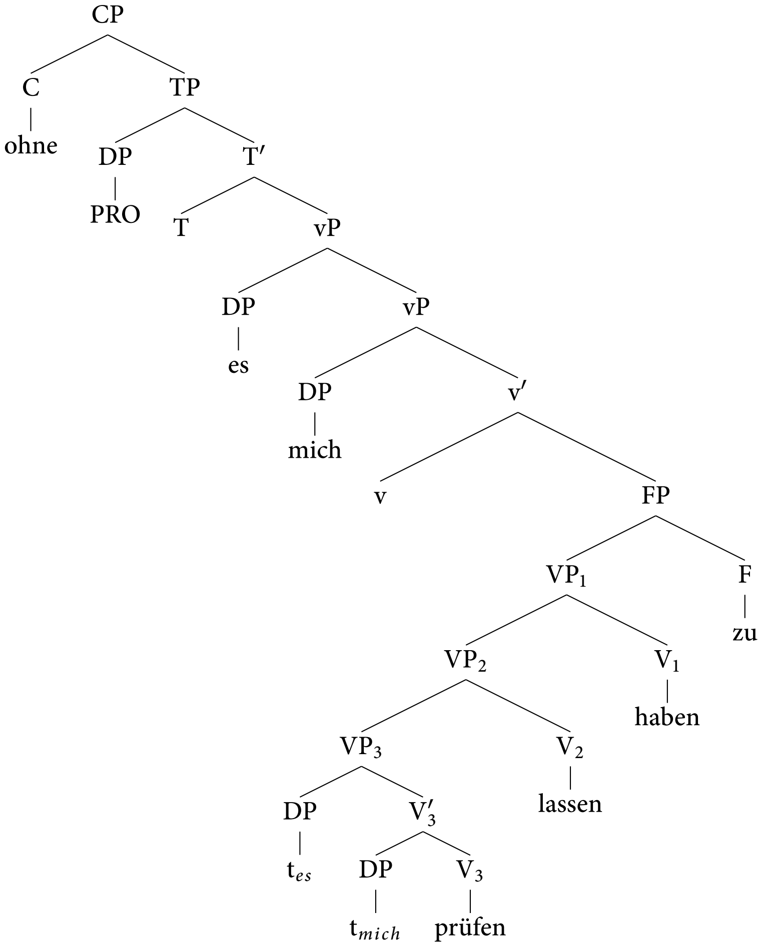
(19) a. 132, ex. (11-c) *Standard German*
ohne [$_{\text{vP}}$ es mich [$_{\text{FP}}$ [$_{\text{VP}_1}$ haben$_1$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ t$_{es}$ t$_{mich}$
without it me have.INF
prüfen$_3$] lassen$_2$]] zu]]
verify.INF let.INF to

b. 12, ex. (13-b) *Swiss German*
… [$_{vP}$ de Taag [$_{FP}$ [$_{VP_1}$ chöne$_1$ [$_{VP_2}$ t$_{de\ Taag}$ bestimme$_2$]] z]]
the day can.INF determine.INF to

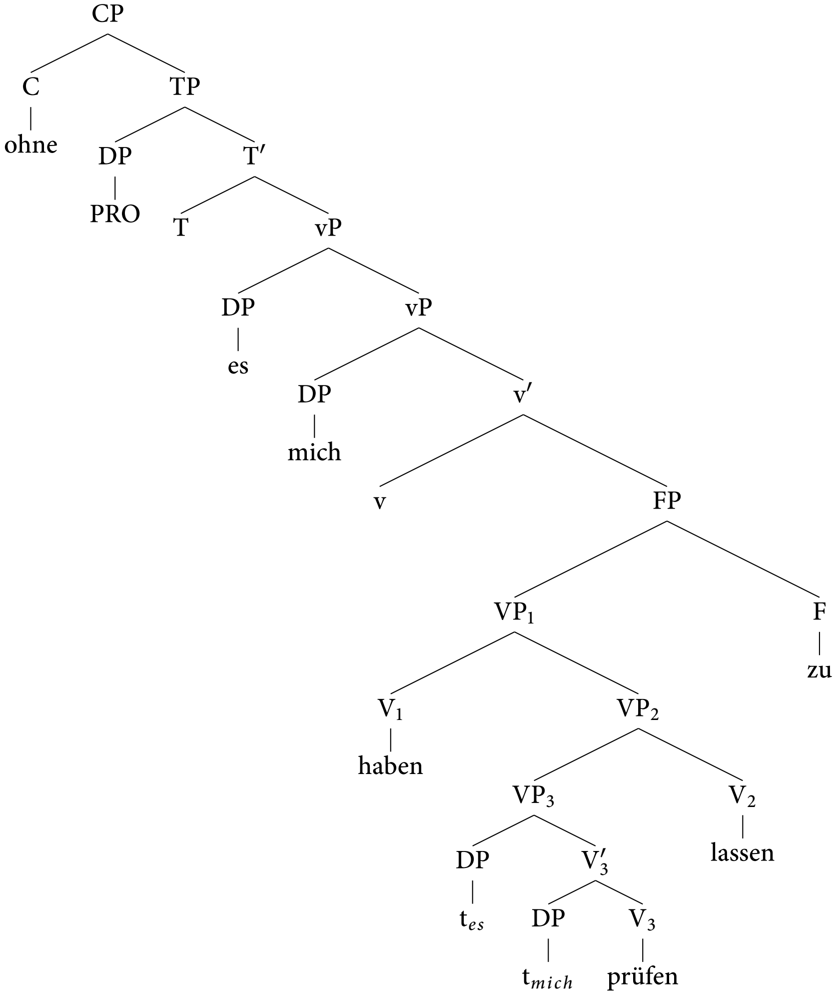
c. 12, ex. (14) *Swiss German*
ohni [$_{vP}$ mi [$_{FP}$ [$_{VP_1}$ wele$_1$ [$_{VP_2}$ t$_{mi}$ uf d Bullesiite
without me want.INF on the cops.side
stelle$_2$]] z]]
put.INF to

The following tree diagrams illustrate the inversion operation for (11-c):

(20) Structure of (11-c) before VP-inversion:



(21) Structure of (11-c) after VP-inversion:



In the next step, *zu*-placement applies. Since *zu* does not end up on the head of VP$_1$, viz. *haben* 'have', but on the rightmost element of the VP, viz. V$_2$ *lassen* 'let', it must be an operation that is sensitive to linear order and adjacency, viz. a late PF-operation like Local Dislocation that affixes a head onto another one and inverts the two.

The operation can be sketched as follows for the three examples under discussion (I retain the VP-brackets for purposes of illustration, but it should be noted that there is no [full] hierarchical structure anymore at this point of the derivation; furthermore, the trace of *zu* is only present for purposes of illustration):

(22) a. $[_{\text{FP}}\ [_{\text{VP}_1}\ \text{V1}\ [_{\text{VP}_2}\ [_{\text{VP}_3}\ \text{V3}]\ \text{V2}]]\ \text{zu}]$
$\Rightarrow [_{\text{FP}}\ [_{\text{VP}_1}\ \text{V1}\ [_{\text{VP}_2}\ [_{\text{VP}_3}\ \text{V3}]\ \textbf{zu}+\text{V2}]]\ t_{zu}]$
b. $[_{\text{FP}}\ [_{\text{VP}_1}\ \text{V1}\ [_{\text{VP}_2}\ \text{V2}]]\ \text{z}]$
$\Rightarrow [_{\text{FP}}\ [_{\text{VP}_1}\ \text{V1}\ [_{\text{VP}_2}\ \textbf{z}+\text{V2}]]\ t_{zu}]$
c. $[_{\text{FP}}\ [_{\text{VP}_1}\ \text{V1}\ [_{\text{VP}_2}\ \text{PP V2}]]\ \text{z}]$
$\Rightarrow [_{\text{FP}}\ [_{\text{VP}_1}\ \text{V1}\ [_{\text{VP}_2}\ \text{PP}\ \textbf{z}+\text{V2}]]\ t_{zu}]$

Zu-cliticization will proceed in the same fashion in (9-a). Since both constructions involve (at least partially) descending orders, there will also be string-vacuous cluster formation. The relative ordering between cluster formation and *zu*-cliticization is discussed in section 3.4.2.

Zu thus shows the behavior of a clitic. It must be stressed, however, that *zu* should be classified as an affix, viz. a phrasal affix (the terminology is somewhat confusing in this area): *zu* has selectional restrictions: it can only occur before verbs in the bare infinitive. This was guaranteed in all the examples discussed so far. There are cases, however, where the last verb of the cluster is a participle, as e.g. in Aux-Part clusters that show ascending order in Western dialects of Switzerland. Interestingly, while the ascending order is unproblematic in finite contexts, it is unacceptable in non-finite contexts – irrespective of the position of *zu*. Instead, only the descending order is acceptable in non-finite contexts (data from Raffaela Baechler, p.c.):

(23) a. 12/21; *Swiss German, Western dialects*
das er s Buech hät$_1$ gläse$_2$/ gläse$_2$ hät$_1$
that he the book has read.PRT read.PRT has
'that he read the book'
b. 12; *Swiss German, Western dialects*
*ohni s Buech ha$_1$ **z** gläse$_2$
without the book have.INF to read.PRT
'without having read the book'
c. 12; *Swiss German, Western dialects*
*ohni s Buech **z** ha$_1$ gläse$_2$
without the book to have.INF read.PRT
'without having read the book'

d. 21; *Swiss German, Western dialects*
ohni s Buech gläse$_2$ **z** ha$_1$
without the book read.PRT to have.INF
'without having read the book'

The ungrammaticality of example (23-c) is unsurprising since *zu* does not occur before the last element of the verb cluster. What is unexpected, though, is the ungrammaticality of example (23-b). It makes perfect sense, though, if *zu* is treated as an affix: since V$_2$ already bears a prefix, there is no space for another affix. Furthermore, the selectional restrictions of *zu* would be violated. As a consequence, only the descending order is grammatical in non-finite clusters.

3.4.1.2. Dutch

I now turn to the derivation of the Dutch example in (9-b), repeated here for convenience:

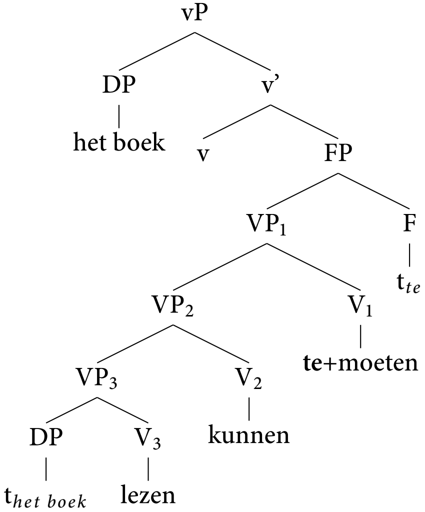
(24) 123 *Standard Dutch*
Hij dacht het boek [**te** moeten$_1$ kunnen$_2$ lezen$_3$].
he thought the book to must.INF can.INF read.INF
'He thought he had to be able to read the book.'

The starting point will be the same as in German: The VP is left-branching with verb clusters starting out as stacked VPs. The particle *te* occupies a functional head above the highest VP and non-verbal material has scrambled from the lexical VP to Specv:

(25) [$_{vP}$ het boek [$_{FP}$ [$_{VP_1}$ [$_{VP_2}$ [$_{VP_3}$ t$_{\textit{het boek}}$ lezen$_3$] kunnen$_2$] moeten$_1$]
the book read.INF can.INF must.INF
te]]
to

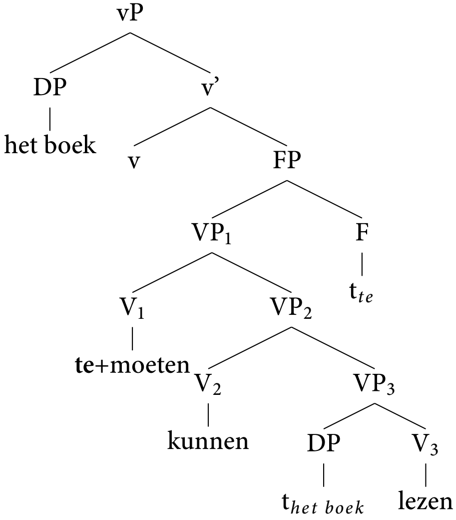
Then, inversion and *te*-placement come into play. Since in Dutch the particle is inverted together with V$_1$, *te*-placement has to precede inversion. Since inversion involves sister nodes and thus hierarchical structure, *te*-lowering will also apply at a point where the hierarchical structure is still available; therefore, *te* undergoes downward head-movement and targets the head of the highest VP, viz. V$_1$ *moeten*. This is illustrated in the following tree diagram:

(26) *te*-Lowering in (9-b):



Thereafter, VP-inversion takes place: V$_1$ inverts with VP$_2$ and V$_2$ inverts with VP$_3$, resulting in an ascending structure:[16]

(27) VP-inversion in (9-b):



---

[16] The example in (10-b) will be derived similarly, with *te*-inversion applying twice and with inversion of V$_1$ with FP$_1$ and V$_2$ with FP$_2$.

The crucial difference between *te*-lowering and *zu*-cliticization is thus the point at which they apply: While *te*-lowering applies at an early stage of the PF-derivation where there is still full hierarchical structure, *zu*-cliticization applies at a later stage where the hierarchical structure has already been converted into a linear one. In the terms of Embick and Noyer (2001), *te*-lowering corresponds to Lowering while *zu*-cliticization corresponds to Local Dislocation. The cross-linguistic variation thus results from the fact that two operations – whose function is a similar one, viz. associate a verb with a marker of non-finiteness – apply at different points in the PF-component.

One may ask at this point why Dutch *te* is treated as an independent element in syntax at all and not as a prefix. As far as I can tell, such an analysis would certainly work for Standard Dutch since *te* is always adjacent to the verb it is supposed to mark. However, an analysis in terms of an independent syntactic element that undergoes lowering is more interesting if the contrast with German is to be described in a systematic way: Instead of treating the two particles as two completely different morphological objects (prefix vs. clitic), it seems more attractive to derive the difference by having the placement rules apply at different points of the derivation. Thereby, a common core can be captured: Both elements need to be associated with a non-finite verb. Furthermore, a look at other Dutch varieties and Afrikaans lends some support to this approach: In both languages, *te* is not lowered onto V_1; instead, it seems to remain an independent element and is inverted like the VPs. This is shown by the following examples involving a 231 order in the verbal complex: *te* does not associate with V_1 but occurs before the entire verbal complex (which in West Flemish contains non-verbal material):

(28) a. *West Flemish*, cf. Haegeman (1998: 635)
mee Valere **te** [[willen$_2$ [dienen boek kuopen$_3$]] een$_1$]
with Valere to want.INF that book buy.INF have.INF
'with Valere having wanted to buy that book'

b. *Afrikaans*, cf. Donaldson (1993)
Die banke moes oop gewees het, om dit gister **te** [[kan$_2$
the bank should open been have to it yesterday to can.INF
betaal$_3$] het$_1$].
buy.INF have.INF
'The bank should have been open to have been able to buy it yesterday.'

If *te* is treated as an independent element in syntax, its varied distribution can be accounted for in a straightforward manner: It undergoes lowering in some (Standard Dutch) but not in all varieties (West Flemish). Furthermore, lowering may apply at different points of the PF-derivation, which accounts for the contrast between (Standard) German and Standard Dutch.

3.4.2. Other Configurations in German

In this subsection, I will discuss the placement of *zu* in Standard German in more detail. I will first determine the relative ordering of *zu*-placement and cluster formation before analyzing the placement of *zu* in the so-called 3rd construction.

3.4.2.1. The Ordering of Cluster Formation

Based on the data in (3) I have been assuming that there is cluster formation in German if the verbal elements end up adjacent to each other in descending order. Importantly, this holds for all verbal elements in descending order in German and its dialects, i.e. also for (9-a), (10-a), for V_3 and V_2 in (11-c) and for V_3 and V_2 in (13-c). In all these constructions, extraposition cannot target the dependent VP, only the clause-final VP is a possible attachment site. [17] Cluster formation was argued to apply after syntax, viz. at some point of the PF-derivation. Since on my assumptions extraposition can block cluster formation, it must apply under adjacency and thus at a late stage of the PF-derivation. Since *zu*-cliticization was shown to apply at a late stage as well, questions arise with respect to the relative ordering of the two operations. The testing ground to decide this issue are examples like (10-a), repeated here, which contain clusters with two *zu*-infinitives:

---

[17] With verbs in ascending orders, things are somewhat more complicated: In ascending Swiss German structures, there can in principle be non-verbal material between the verbs (i.e. VPR is always a possibility); consequently, there cannot be a general surface constraint requiring string-vacuous cluster formation in that case. The same goes for partially ascending structures in German like (11-c) which also permit residual VPR (i.e. non-verbal material beween V_1 and V_3, cf. Bader and Schmid 2009: 224f.). In Dutch, however, ascending structures are almost completely impenetrable; most interveners can be considered incorporated X^0-elements. Consequently, a surface constraint requiring string-vacuous cluster formation would lead to the correct result and may in fact act as a trigger for the evacuation of the verb cluster.

(29) 321, (10-a) *Standard German*
dass er [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ das Buch **zu** lesen$_3$] **zu** versuchen$_2$]
that he the book to read.INF to try.INF
versprach$_1$]
promised
'that he promised to try to read the book'

The starting point, before the PF-operations apply, will be the following left-branching structure (again, I assume for the sake of concreteness that non-verbal material has scrambled to Specv, even though this is optional in descending structures):

(30) 321, (10-a) *Standard German*
dass er [$_{\text{vP}}$ das Buch [$_{\text{VP}_1}$ [$_{\text{FP}_1}$ [$_{\text{VP}_2}$ [$_{\text{FP}_2}$ [$_{\text{VP}_3}$ t$_{\textit{das Buch}}$ lesen$_3$] **zu**]
that he the book read.INF to
versuchen$_2$] **zu**] versprach$_1$]]
try.INF to promised
'that he promised to try to read the book'

Since verb cluster formation is based on adjacency, it seems that *zu*-cliticization has to take place *before* cluster formation:

(31) a. *zu*-cliticization
dass er das Buch [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ **zu**+lesen$_3$] **zu**+versuchen$_2$]
that he the book to+read.INF to+try.INF
versprach$_1$]
promised
b. cluster formation
dass er das Buch [$_{\text{V}}$ [$_{\text{V}}$ **zu**+lesen$_3$]+ [$_{\text{V}}$ **zu**+versuchen$_2$]+ [$_{\text{V}}$
that he the book to+read.INF to+try.INF
versprach$_1$]]
promised

Unfortunately, this ordering seems to lead to a contradiction once the 'scandal construction' (12-a) is taken into account. I repeat the relevant example:

(32) 312, (12-a) *Standard German*
Er bedauert, es nicht [verhindert$_3$ haben$_1$ **zu** können$_2$].
He regrets it not prevent.PRT have.INF to can.INF
'He regrets not having been able to prevent it.'

The starting point will be the following structure:

(33) ... es nicht [$_{FP}$ [$_{VP_1}$ [$_{VP_2}$ [$_{VP_3}$ t$_{es}$ verhindert$_3$] können$_2$] haben$_1$] **zu**]
it not prevent.PRT have.INF can.INF to

Since *zu*-cliticization was assumed to precede cluster formation to derive the correct result for (10-a) = (29), one would expect *zu* to surface on V$_1$, contrary to fact. Rather, the facts suggest that reordering between V$_1$ and V$_2$ must precede *zu*-cliticization. However, this reordering cannot be an instance of VP-inversion because the two verbs are not sisters – VP-inversion would incorrectly place V$_1$ at the beginning of the cluster (string vacuous movement of VP$_3$ to some higher position seems ill-motivated). Consequently, the reordering between V$_1$ and V$_2$ must be the result of a different process. The obvious choice is cluster formation under adjacency. While the cases of cluster formation we have looked at so far, cf. (3), (9-a) and (10-a) were all string-vacuous, this one involves reordering. In fact, it is arguably the same kind of process as *zu*-cliticization. To derive the correct result for the scandal construction, reordering cluster formation has to take place *before zu*-cliticization:

(34) a. reordering/cluster formation (12-a)
... es nicht [[verhindert$_3$] haben$_1$+können$_2$] **zu**
it not prevent.PRT have.INF+can.INF to
b. *zu*-cliticization (12-a)
... es nicht [[verhindert$_3$] haben$_1$+**zu**+können$_2$]
it not prevent.PRT have.INF+to+can.INF

It seems that we have arrived at an impasse: To derive (10-a) = (29), we need to assume that cluster formation follows *zu*-cliticization while the reverse ordering is necessary to derive (12-a) = (32). There are two possibilities to resolve the contradiction.

First, since the two cases of cluster formation differ in that one is string-vacuous while one involves reordering, one could classify them as two independent operations that can be ordered differently with respect to other operations.[18] Concretely, one has to assume that reordering cluster formation can precede *zu*-cliticization, which in turn precedes string-vacuous cluster formation. The resulting order thus looks as follows:

(35) reordering cluster formation > *zu*-cliticization > string-vacuous cluster formation

Cluster formation is an instance of Local Dislocation in both cases as both are sensitive to linear order. As pointed out above, *zu*-cliticization is also an instance of this type of PF-operation.

Second, there is an alternative to resolve the contradiciton that avoids extrinsic ordering: If we assume that the bracketed structure in (29) (= (10-a)) and (32) (= (12-a)) is still available at the point where the PF-operations apply, the placement facts follow under cyclicity if the PF-derivation unfolds bottom-up: In (32), verb cluster reordering will first produce [3[1+2]*zu*], then, *zu* is inverted with V_2, correctly deriving [3[1+*zu*+2]]. In (29), the lower *zu* (F_2) is first cliticized onto V_3: [zu_2+V_3]. Then, [zu_2+V_3] undergoes string-vacuous cluster formation with V_2, resulting in [zu_2+V_3+V_2]. Then, the higher *zu* (F_1) cliticizes onto V_2, resulting in [zu_2+V_3+zu_1+V_2]. Finally, the entire complex undergoes string vacuous cluster formation with V_1: [zu_2+V_3+zu_1+V_2+V_1]. The second solution has the advantage that it avoids extrinsic ordering. Furthermore, the full derivation of (34) will additionally involve string-vacuous cluster formation between V_3 and the complex [V_1+V_2] because 312 clusters behave like complex heads (nothing may intervene between V_3 and V_1). If the ordering is as in (35), cyclicity would be violated. A fully cyclic derivation that avoids rule ordering, however, can derive the correct result. Finally, the cyclic derivation treats both instances of cluster formation as the same operation. Given these advantages, I opt for the second solution. [19]

---

[18] This is a somewhat unsatisfactory move since it seems straightforward to treat string-vacuous cluster formation as a subcase of reordering cluster formation, the latter simply being more complex in involving an additional operation.

[19] An alternative to both solutions would be to derive the impenetrability of descending clusters without cluster formation as e.g. in Wurmbrand (2007) where extraposition is sensitive to prosodic principles. In that case, *zu*-cliticization can be ordered after cluster formation without having any detrimental consequences.

3.4.2.2. *zu*-Placement in the Third Construction/Extraposition

I now turn to *zu*-infinitival complements with ascending order in (Standard) German. There are generally two types of *zu*-infinitives irrespective of whether they occur intraposed, i.e. in descending order, or whether they occur in postverbal (= extraposed) position, i.e. in ascending order. The two types differ with respect to their integration into the matrix clause: Some *zu*-complements are so small in size that they form a monoclausal unit together with the matrix clause for clause-bound processes like scrambling or weak pronoun fronting, i.e. they show restructuring/transparency effects. They are usually treated as VPs, as has been presupposed for (10-a) (= (29)), and are usually referred to as restructuring infinitives. Other complements do not show any of the transparency effects and are therefore usually treated as CPs; they are termed non-restructuring infinitives. Whether a complement is restructuring or non-restructuring usually depends on the selecting predicate. Some select only restructuring infinitives (e.g. *scheinen* 'seem' in Standard German), others select only non-restructuring infinitives (e.g. *bedauern* 'regret'), and a third class can select either type of complement (e.g. *versuchen* 'try'). This classification is also found when the *zu*-infinitive occurs in post-verbal or extraposed position. When a restructuring infinitive occurs in postverbal (= extraposed) position, the construction is referred to as the *3rd construction*. Both constructions are illustrated by the following examples (scrambling of the object pronoun in example (36-b) indicates transparency; the same operation would lead to ungrammaticality in example (36-a)):

(36) a. CP-complement, *Standard German*
ohne $[_{VP_1}$ zu bedauern$_1$ $[_{CP}$ mich zu mögen$_2$]]
without to regret.INF me to like.INF
'without regretting to like me'
b. VP-complement/3rd construction, *Standard German*
ohne mich $[_{VP_1}$ zu versuchen$_1$ $[_{VP_2}$ t_{mich} zu mögen$_2$]]
without me to try.INF to like.INF
'without trying to like me'

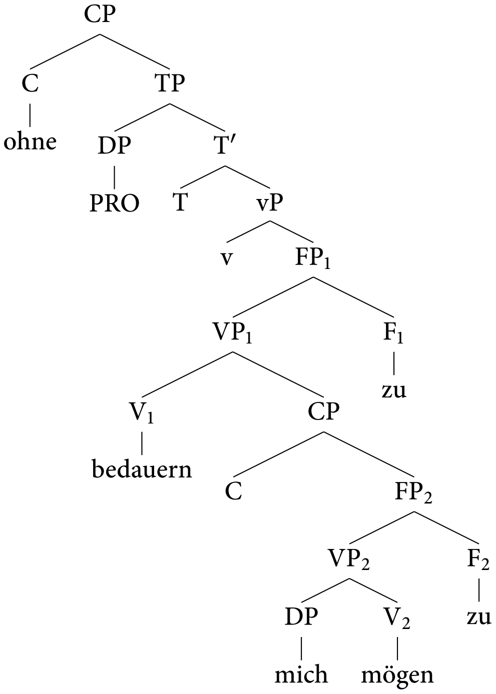
An obvious question is how the ascending orders are derived. There are two possibilities: Either they are the result of PF-inversion – like VPR-structures – or they are derived by extraposition, i.e. movement to the right. In what follows, I will use the placement of *zu* as a diagnostic to determine which of the two options is correct. The starting point will again be a descending order:

(37) a. CP-complement
ohne [$_{\text{FP}_1}$ [$_{\text{VP}_1}$ [$_{\text{CP}}$ [$_{\text{FP}_2}$ [$_{\text{VP}_2}$ mich mögen$_2$] zu]] bedauern$_1$]
without me like.INF to regret.INF
zu]
to

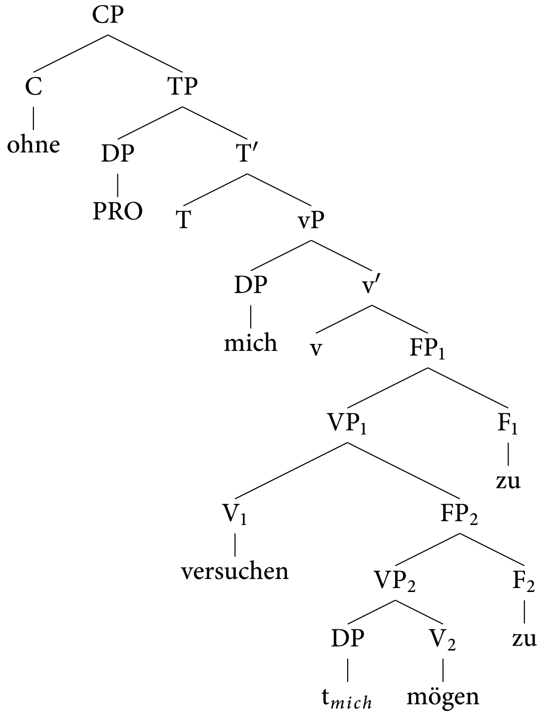
b. VP-complement/3rd construction
ohne [$_{\text{vP}}$ mich [$_{\text{FP}_1}$ [$_{\text{VP}_1}$ [$_{\text{FP}_2}$ [$_{\text{VP}_2}$ mögen$_2$] zu] versuchen$_1$]
without me like.INF to try.INF
zu]]
to

Suppose ascending structures are derived by means of PF-inversion of V_1 with CP/FP$_2$; the result is indicated in the following tree diagrams:

(38) a. non-restructuring predicate, ex. (36-a)



b. restructuring predicate, ex. (36-b)

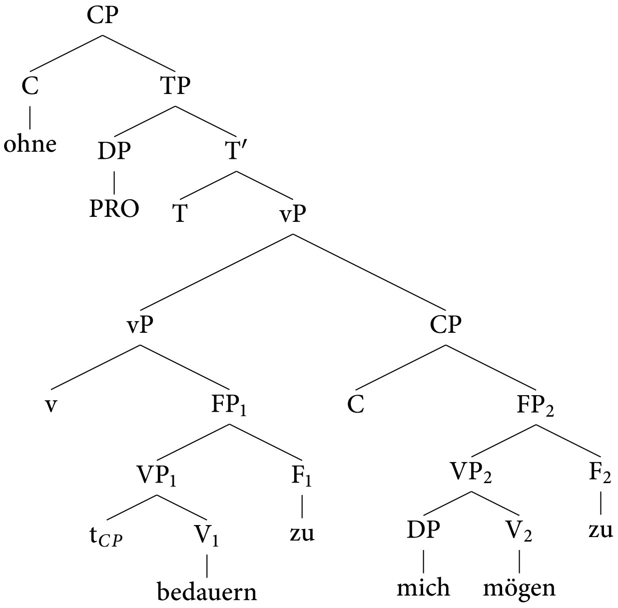


If *zu* then cliticizes onto the right-most verb, an ungrammatical result obtains: both *zu*s are cliticized onto the the last verb and V$_1$ fails to bear *zu* (or, alternatively, the two *zu*s are reduced to one by haplology and there is only one *zu*, which occurs on the right-most verb of the complement):

(39) a. CP-complement
*ohne bedauern$_1$ mich [**zu**+**zu**+mögen$_2$]
without regret.INF me to+to+like.INF
b. VP-complement/3rd construction
*ohne mich versuchen$_1$ [**zu**+**zu**+mögen$_2$]
without me try.INF to+to+like.INF

Consequently, ascending structures must be derived by means of movement to the right (= extraposition) of CP/FP$_2$, as indicated in the following tree structures (For the sake of concreteness, I am assuming adjunction to matrix vP, but a different/higher position would also be conceivable):[20]

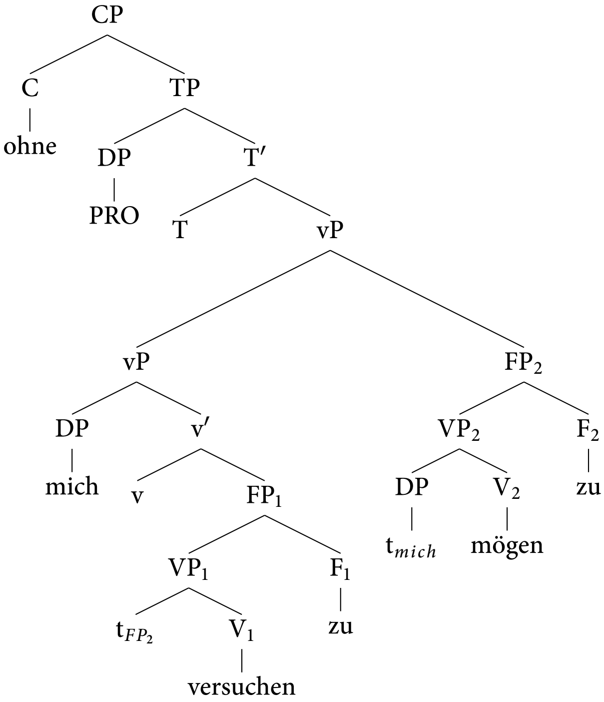
(40) a. non-restructuring predicate, ex. (36-a)



---

[20] The present account presupposes that extraposition takes place in syntax. However, a PF-movement account of extraposition would also be possible as long as it precedes *zu*-cliticization.

b. restructuring predicate, ex. (36-b)



This provides the correct input for the *zu*-cliticization rule: After linearization, both *zu*s are adjacent to the verb which they are supposed to be associated with and cliticization is successful (again, the traces of *zu* are only employed for expository purposes):

(41) a. CP-complement
ohne [$_{vP}$ [$_{vP}$ [$_{FP_1}$ [$_{VP_1}$ t$_{CP}$ **zu**$_1$+bedauern$_1$] t$_{zu1}$]] [$_{CP}$ [$_{FP_2}$
without to+bedauern.INF
mich **zu**$_2$+mögen$_2$ t$_{zu_2}$]]]
me to+like.INF

b. VP-complement/3rd construction
ohne mich [$_{vP}$ [$_{vP}$ [$_{FP_1}$ [$_{VP_1}$ t$_{FP_2}$ **zu**$_1$+versuchen$_1$] t$_{zu_1}$]] [$_{FP_2}$
without me to+try.INF
[$_{VP_2}$ **zu**$_2$+mögen$_2$] t$_{zu_2}$]]
to+like.INF

Zu-placement thus provides new evidence that the 3rd construction in Standard German involves (remnant) extraposition and therefore has to be derived differently than VPR-structures.

Note that extraposition is also necessary for finite complement clauses as e.g. in the following example:

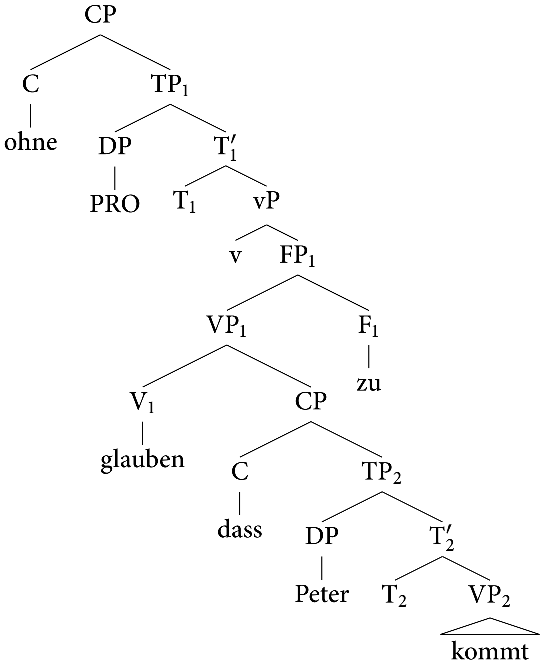
(42) *Standard German*

ohne	zu	glauben	[$_{CP}$ dass	Peter	kommt]
without	to	believe	that	Peter	comes

‘without believing that Peter will come’

If the complement CP starts as a left-hand complement of V and reaches its post-verbal position by means of inversion with V (or is directly linearized as a right-hand-complement, as is sometimes proposed even under left-branching approaches), one would expect *zu* to end up on the rightmost element of the complement clause, contrary to fact. Consider the following tree structure:

(43) CP-complement without extraposition, ex. (42)

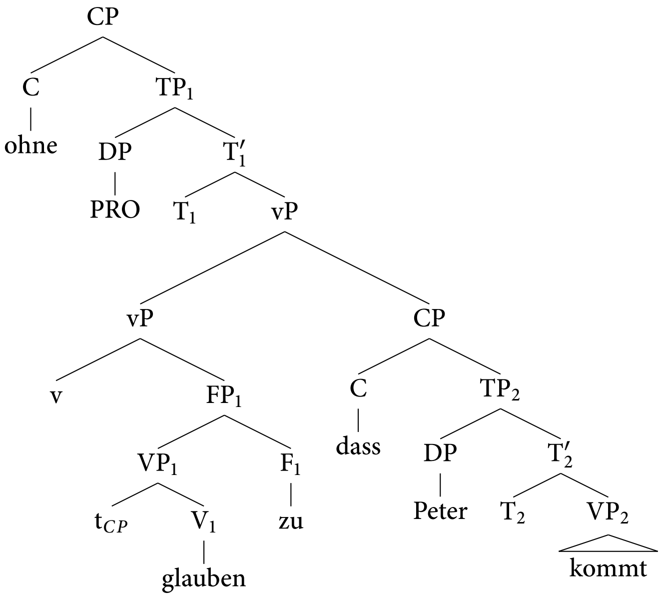


If *zu*-cliticization applied to this structure, the result would be ungrammatical:

(44) *Standard German*
*ohne glauben, [$_{CP}$ dass Peter **zu** kommt]
without believe that Peter to comes
'without believing that Peter will come'

Instead, the CP has to be extraposed (adjoined to matrix vP):

(45)



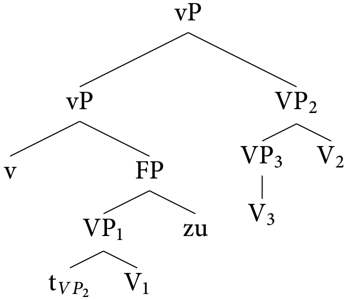
This provides the correct context for *zu*-cliticization: it now targets V_1.

Note incidentally that the *zu*-placement facts tend to argue against extraposition analyses of ascending/VPR-structures as e.g. proposed in Haegeman (1992): The basic idea is that the dependent VP is right-adjoined to the higher VP. Applied to a 3-verb cluster with 132 order, the structure would look as follows (since extraposition involves VP_2, there can be non-verbal material between V_1 and V_3, thereby deriving VPR-structures):

(46) $[_{VP_1} [_{VP_2} [_{VP_3} V_3] V_2] V_1] \Rightarrow [_{VP_1} [_{VP_1} t_{VP_2} V_1] [_{VP_2} [_{VP_3} V_3] V_2]]$

Suppose that this is a cluster within non-finite VP es e.g. in the misplaced *zu* examples in Standard German, cf. (11-c). If extraposition of VP$_2$ targets matrix vP as in the previous cases of extraposition, we obtain the following structure:

(47)



If we then apply *zu*-cliticization, *zu* ends up on V$_1$, as in the ungrammatical (11-b). This can only be avoided if extraposition targets a position below *zu*, i.e. involves adjunction to VP$_1$. Consequently, if all ascending/VPR-structures arise by means of extraposition, one has to assume that extraposition may target different nodes depending on whether a VP (adjunction to VP), or an FP/CP (adjunction to vP) is extraposed. In the approach pursued here, there are two mechanisms that derive ascending structures, VP-inversion and extraposition. It is not so easy to choose between these two options, but there is one piece of empirical evidence that argues that the latter approach is superior.

I will concentrate in what follows on the difference between the 3rd construction and VPR. They are similar in that both constitute monoclausal domains for processes like scrambling and pronoun fronting (vor VPR cf. e.g. Haegeman 1992: 110). Additionally, when an argument of the lexical verb is scrambled to a higher verbal projection, the movement does not show the hallmarks of regular scrambling (like inducing freezing effects and blocking focus projection), cf. Salzmann (2011), Geilfuss-Wolfgang (1991: 25f.). This may suggest that they should be derived in the same way. However, there is a striking asymmetry in the domain of scope: While scrambled elements in VPR-constructions can reconstruct, this does not seem to be possible in the 3rd construction (cf. Salzmann 2011: 454 for VPR and Bobaljik and Wurmbrand 2005: 810,831):

(48) a. dass er [$_{\text{VP}_1}$ 2 Manager wett$_1$ [$_{\text{VP}_2}$ t$_{2\ Manager}$ vo siine Idee
that he 2 managers wants of his ideas
überzüüge$_2$]]
convince.INF
'that he wants to convince two managers of is ideas'
(2 > want; want > 2)

b. weil er [$_{\text{VP}_1}$ alle Fenster vergass$_1$ [$_{\text{VP}_2}$ t$_{alleFenster}$ zu
because he all windows forgot to
schliessen$_2$]]
close.INF
'because he forgot to close all the windows' (all > forget; *forget > all)

Under an extraposition analysis of the 3rd construction, there is a straightforward explanation for the absence of reconstruction: What is extraposed is a remnant VP. Importantly, remnant VPs have been shown to induce scope freezing effects (Barss 1986: 517-542), as expressed in the following generalization:

(49) reconstruction of α to its trace β is blocked if α does not c-command β at S-structure.

This is exactly the configuration that obtains in remnant (VP-) movement: α is A-moved out of a VP; VP is then A$'$-moved to a position above α so that α no longer c-commands its trace β. An example pair illustrating the effect in the domain of verb clusters is the following (slightly adpated from Haider 2003: 101):

(50) a. dass ihr [$_{\text{VP}_1}$ niemand [$_{\text{VP}_2}$ t$_{niemand}$ zu beleidigen$_2$]
that her.DAT no.one to insult.INF
gelang$_1$]
succeeded
'that she managed to insult no.one' ($\neg\exists$ > succeed; succeed > $\neg\exists$)

b. [$_{\text{VP}_3}$ t$_{niemand}$ zu beleidigen$_3$] ist$_1$ ihr [$_{\text{VP}_1}$ niemand [$_{\text{VP}_2}$
to insult.INF is her.DAT no.one
t$_{VP_3}$ gelungen$_2$] t$_{ist}$].
succeeded
'that she managed to insult no.one' ($\neg\exists$ > succeed; *succeed > $\neg\exists$)

It is immaterial at this point how the scope freezing effect is derived (cf. e.g. Sauerland and Elbourne 2002 for a recent proposal); descriptively, it seems that A-movement can in principle reconstruct, leading to ambiguity in intraposed verb cluster constructions like (50-a) and VPR (48-a), but not if remnant movement is involved as in remnant topicalization of part of a verb cluster (50-b). The lack of ambiguity in the third construction (48-b) then finds a straightforward explanation if it is analyzed as an instance of remnant extraposition as well.

To conclude this digression: The facts from *zu*-placement suggest that ascending structures in VPR and the 3rd construction are derived differently. This correlates with a scope asymmetry. While an approach solely based on extraposition cannot capture the scope differences in a natural way, similarities and differences fall out directly if it is assumed that the two constructions are the result of different mechanism, viz. VP-inversion vs. extraposition.[21]

3.5. Ordering of Operations: Synopsis

I have discussed a number of operations that interact in intriguing ways. It turned out to be possible to determine a strict order between these operations without encountering any contradictions. They are summarized in the following table:

(51) Ordering of operations (left-branching): synopsis

extraposition (RC, CP/FP$_2$) > topicalization (VP) >	syntax
te-lowering > VP-inversion > cluster formation (+/-inversion) > < *zu*-cliticization	PF

The ordering in syntax is intrinsic in that it follows from the Strict Cycle Condition/the Extension Condition. The ordering between *te*-lowering and VP-inversion is probably extrinsic. The ordering between cluster formation and *zu*-cliticization again follows from cyclicity (if that concept is adopted for PF). Their ordering with respect to VP-inversion is also intrinsic if it is assumed that

---

[21] If as discussed in section 3.7, a right-branching VP-structure is adopted, the two structures can be distinguished in that the 3rd construction involves extraposition while in VPR the basic linearization is retained.

hierarchy-sensitive operations take place in a different subcomponent of PF than adjacency-sensitive operations.

3.6. Previous Accounts

The previous sections have shown that an analysis in terms of rule ordering is successful in accounting for the placement of *zu*. But before concluding that this is the optimal solution, I would like to briefly discuss previous analyses of misplaced *zu*. They can be divided into syntactic/derivational accounts where *zu* is an independent syntactic element and morphological/realizational approaches where *zu* is just a feature of a non-finite complement that receives morphological expression according to specific rules. I will discuss the two types of approaches in turn.[22]

3.6.1. Syntactic/Derivational Accounts of Misplaced zu

The idea that cluster formation has to precede placement of *zu* can be found in a number of analyses. For instance, von Stechow (1990: 159) argues that *zu* is generated in INFL and incorporated into the verbal complex after reanalysis (which is taken to be the mechanism that generates complex heads and ascending orders, cf. Haegeman and van Riemsdijk 1986). The account seems to involve lowering/affix hopping of *zu*. It is explicitly assumed that this takes place before PF, which means it should be subject to syntactic locality conditions; as a consequence, one would expect *zu* to end up on V_1 and not on V_2. It seems that lowering is sensitive to adjacency in this account, but that seems implausible for a syntactic operation (quite apart from the fact that it violates the c-command constraint on incorporation). Arguably, the underlying intuition was the same as that for the rule of *zu*-cliticization proposed here, but given the framework of that time, a solution by means of a PF-operation was apparently not obvious. Whether this account can be extended to other cases of *zu*-placement and how it would deal with Dutch remains open. A somewhat different proposal is found in Sternefeld (1990: 251) who first argues that it is the rightmost verb that moves to INFL, where *zu* is base-generated. Since movement to INFL follows cluster formation, *zu* ends up on the correct verb

---

[22] A hybrid solution is proposed in Sternefeld (2006) who treats *zu* as a lexical feature of infinitives, which, however, can undergo migration to a different part of the cluster to derive cases of misplaced *zu* like (11-c).

(viz. V_2). But it remains mysterious why it is not the head of the verb cluster V_1 that moves to INFL as would be expected under a syntactic account where locality constraints apply (e.g. minimality). Furthermore, the account requires excorporation of V_2 from the cluster. Sternefeld also considers an incorporation solution as in von Stechow (1990) but points out that this raises problems for *te*-placement in Dutch as in (9-b) where incorporation would have to precede cluster formation/reordering. The issue is eventually left open, and it remains unclear to what extent the placement of the infinitival particle can be handled in a systematic way both within German and cross-linguistically. The solution that comes closest to the current proposal is the one by Hinterhölzl (2009: 208) who argues that *zu* is a phrasal affix and fuses with the adjacent infinitive at Morphological Form. This is clearly a PF-operation and the basic intuition is arguably the same as in the present account; unfortunately, the workings of the operation are not spelled out in much detail so that it is not clear to what extent it can be applied to other cases of *zu*-placement (for instance, Hinterhölzl only discusses misplaced *zu* in 132 clusters like (11-c) but does not address 312 clusters like (12-a)). Furthermore, it is not clear whether the cross-linguistic variation can be derived in a natural way.

3.6.2. Realizational Approaches

In Bader (1995) and Vogel (2009), *zu* is treated as a phrasal affix/an instance of edge inflecton. It is not an independent syntactic head but rather a morphosyntactic feature assigned to an infinitival complement. Its realization on the last element of the verb cluster is the result of special realizational rules (an EDGE-feature in Bader's HPSG-approach and an alignment constraint in Vogel's OT-account). Both approaches successfully account for misplaced *zu/z*: *zu/z* is realized on the rightmost terminal node of the verb cluster. This captures in a very different way the intuition that the position of *zu* depends on the surface order in the verbal complex. The question that arises, though, is whether this rule of *zu*-placement successfully accounts for other cases of *zu*-placement as well. There are no problems with a single verb cluster as in (9-a). But clearly, the realization of *zu* on some dependent element of a non-finite complement must be restricted. For instance, it has to be ruled out that *zu* is realized on the rightmost element of a verb cluster that is embedded under the non-finite complement, as e.g. in (42) above and in the following example from Vogel (2009: 329):

(52) *Standard German*
*Ich bin froh, nicht haben hören ~~zu~~ müssen, dass du dich
I am happy not hav.INF hear.INF to must.INF that you self
geärgert **zu** hast.
annoyed to have.2SG
'I am happy that I did not have to hear that you were annoyed.'

Vogel (2009: 329) proposes that *zu* appears on the right-most element of the extended projection marked for *zu*. This correctly rules out (52) since the embedded clause constitutes a separate domain with its own extended projection. Other cases are less trivial. Consider, for instance, the placement of *zu* in VPR-cases like (14), repeated for convenience:

(53) 1X2 *Swiss German*
ohni mi [welle$_1$ [uf d bullesite **z** stelle$_2$]], im gegeteil,
without me want.INF on the cops.side to put.INF on.the contrary
aber ...
but
'without wanting to side with the cops, on the contrary, but ...'

Here, the complement of *welle* 'want' contains non-verbal material. In principle, VPR-complements can also contain external arguments and adverbials related to tense, which suggests that they are larger than bare VPs. There is no general consensus on their size apart from the fact that they are smaller than CPs but larger than VPs; for instance, based on West-Flemish data with expletive subjects within the VPR-complement, den Dikken (1996) proposes that they are TPs. Whether this constitutes a separate extended projection in Vogel's terms is hard to say. With respect to transparency effects like scrambling and pronoun fronting, VPR-complements tend to behave like monoclausal units, cf. e.g. Haegeman (1992: 110). This might indicate that VPR-complements do not constitute a separate domain, but note that this equates 'separate extended projections' with transparency effects and the presence of a CP. In other words, Vogel's approach predicts that *zu* must be realized within the CP containing the non-finite complement marked with *zu*. As pointed out in Haider (2011: 250), this may make the wrong prediction for restructuring infinitives like (10-a), repeated here for convenience:

(54) 321 *Standard German*
dass er [$_{\text{VP}_1}$ [$_{\text{VP}_2}$ [$_{\text{VP}_3}$ das Buch **zu** lesen$_3$] **zu** versuchen$_2$]
that he the book to read.INF to try.INF
versprach$_1$]
promised
'that he promised to try to read the book'

There is no doubt that the lower *zu*-phrase shows transparency effects; according to the reasoning above, this would suggest that it does not have any extended projections. It is not fully clear what Vogel's account would predict in this case. Since the structure is left-branching, one expects a *zu* on V$_1$ *versuchen* 'try' as it is the right-most element of the VP. The question is whether the embedded VP can be marked for *zu* at all if it does not have any functional projections above VP. Arguably, the realizational rule has to be adjusted: *zu* must be realized on the right-most element *within* the non-finite constituent marked vor *zu*, irrespective of whether the non-finite constituent bears any functional projections above VP. To rule out (52), one has to stipulate that *zu* has to be realized within the same CP that contains the non-finite XP marked for *zu*.[23] Another possible complication arises with the third construction as in (36-b) and the corresponding tree structure in (40-b). I repeat the final structure for convenience:

(55) VP-complement/3rd construction
ohne mich [$_{\text{vP}}$ [$_{\text{vP}}$ [$_{\text{FP}_1}$ [$_{\text{VP}_1}$ t$_{FP_2}$ **zu**+versuchen$_1$] t$_{zu_1}$]] [$_{\text{FP}_2}$ [$_{\text{VP}_2}$
without me to+try.INF
zu+mögen$_2$] t$_{zu_2}$]]
to+like.INF

Here it crucially depends on which XP is assigned the feature *zu*. If it is the projection that the extraposed VP is adjoined to, viz. vP as in (40-b), one would probably expect *zu* to remain unrealized in VP$_1$ because the right-most terminal of the non-finite XP assigned *zu* would be the right-most verb of the extraposed FP$_2$. Perhaps this can be avoided by the restriction introduced above that *zu* has to be realized within the phrase marked for *zu*, under the

---

[23] This stipulation can perhaps be avoided if CP-complements are extraposed, as was assumed in (45) above. Unfortunately, Vogel does not discuss the structural position of CP-complements. However, as we will see below, extraposition may still not be sufficient.

assumption that the adjoined FP$_2$ is not sufficiently part of FP$_2$.[24] As should have become clear, once other *zu*-placement phenomena are taken into account, the intuitively simple solution proposed in Vogel (2009) has to be modified and restricted in rather specific ways to attain observational adequacy.

Examples like (55) are particularly interesting given the observations about so-called missing *z* in Swiss German in Bader (1995). I will discuss this phenomenon in a separate subsection.

3.6.3. Missing z *in Swiss German*

Bader (1995: 22,26) discusses Swiss German verb clusters that seem structurally very parallel to the 3rd construction data in the standard language. They also contain two non-finite phrases marked for *zu*. What is remarkable in this construction is that the *zu* assigned to the higher VP can go missing (*schiine* 'seem' and *probiere* 'try' require a *zu*-infinitive):

(56) *Bernese German*

wüu	dr	Hans	sine	Fründe	schiint$_1$	probiere$_2$	z	häuffe$_3$
because	the	John	his.DAT	friends	seems	try.INF		to help.INF

'because John seems to try to help his friends'

As opposed to the examples from the standard language like (36-b), there is no *zu* on V$_2$ *probiere* 'try'. Bader accounts for both missing *z* and misplaced *z* by means of a realizational rule that is very smilar to the one proposed in Vogel (2009). Although the technical details differ, the result is the same: *zu/z* is realized on the last element of the XP assigned/marked with *zu*. This accounts for misplaced *zu*.

There is no provision in this system for preventing the feature from percolating downwards, and this is exactly what is exploited to account for missing *z*: In (56), the XP headed by *häuffe* 'help' is analyzed as a complement of *probiere* 'try'. The *zu*-feature assigned to the constituent headed by *häuffe* is, of course, realized on *häuffe*. Crucially, the *zu*-feature assigned to the constituent headed by *probiere* is also realized on *häuffe* because this is the rightmost element contained in that XP. In other words, missing *z* is not a separate phenomenon under this analysis, it is simply a side-effect of the workings of the realizational

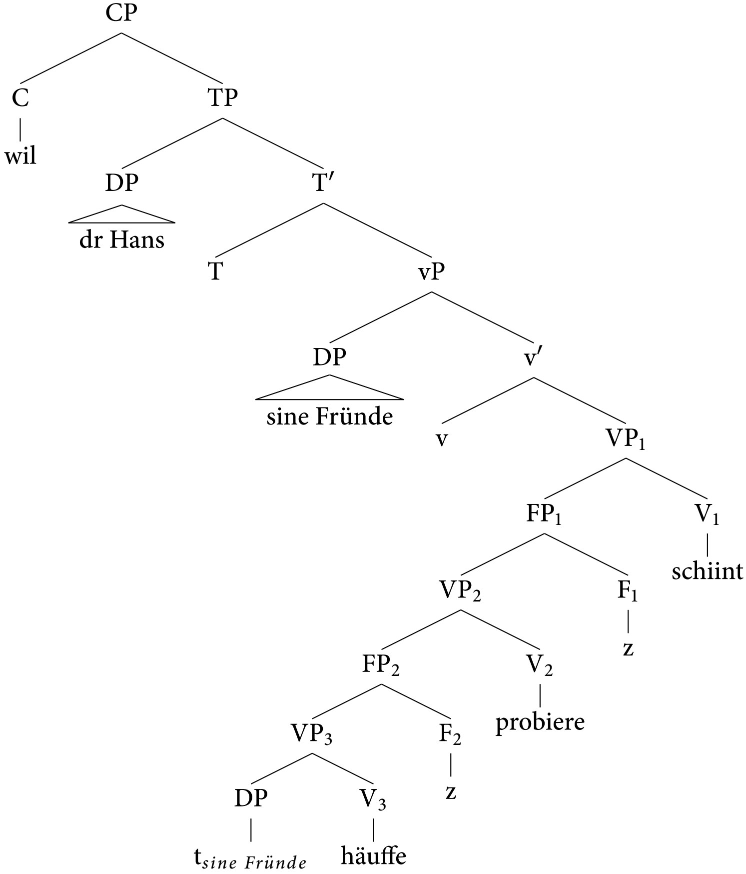
---

[24] An obvious alternative consists in adjoining the extraposed VP to a higher node, but this would have to be motivated independently.

rule. This seems to be a very elegant solution, but problems arise in other areas of *zu*-placement (basically as in Vogel's approach). Since downwards percolation is in principle unlimited, it could also wrongly derive (52) where *zu* assigned to the matrix VP ends up on the right-most verb of the embedded VP. Again, downwards percolation has to be restricted somehow. While the left-branching cases with two *zu*s as in (54) can probably be handled (the EDGE-feature has to be realized at the right edge of every constituent bearing the feature), problems also arise with the 3rd construction in Standard German, cf. (36-b) where one might also expect a missing *z*, contrary to fact. It seems that Bader would have to resort to extraposition of both VPs and CPs, which is perhaps not so obvious in his more surface-oriented HPSG-approach. In sum, then, realizational approaches to *zu*-placement may seem very elegant at first sight. However, once the entire empirical domain of *zu*-placement is taken into account, it becomes obvious that they need to be restricted in rather specific ways to derive the facts.

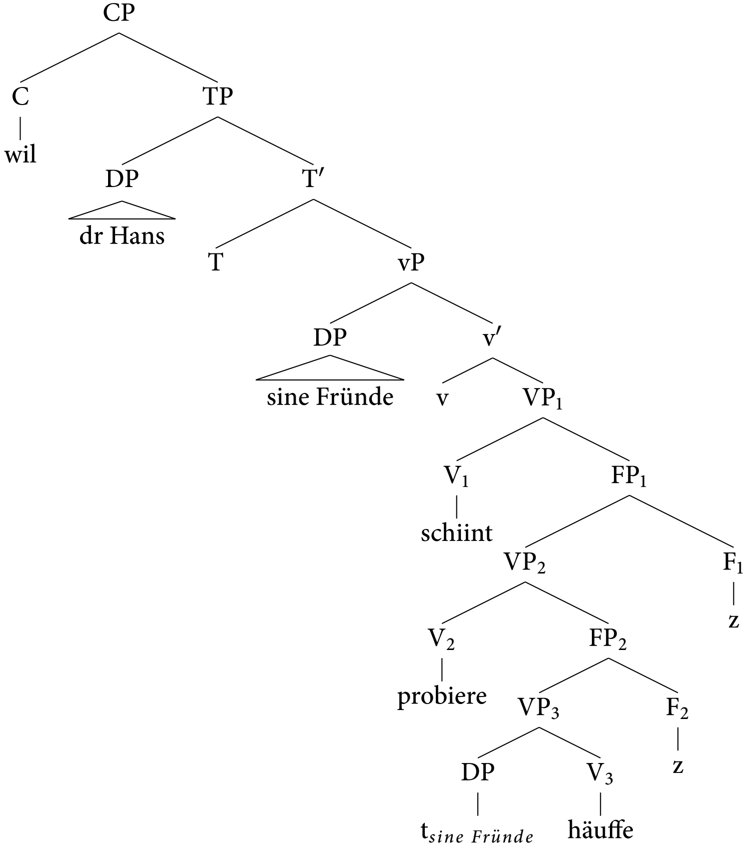
Before concluding, I would like to briefly discuss missing *z* in the rule-based framework adopted here. The starting point for the derivation of an example like (56) will again be a left-branching structure with two FPs as in the 3rd construction in Standard German (37-b):

(57) Missing *z* before inversion, cf. ex. (56)



Then suppose that the ascending structure is not derived by means of extraposition but by means of PF-inversion of V_1 with FP_1 and V_2 with FP_2:

(58) Missing *z* after VP-inversion, cf. ex. (56)



Note that the two *zu*s are now adjacent to each other. Suppose that a rule of haplology can reduce them into one *zu*. Then, *zu*-cliticization can apply in regular fashion, deriving the missing *z* example. The difference between the 3rd construction in Standard German and the missing *z*-construction would therefore reside in the operation that derives the ascending structure: extraposition vs. VP-inversion. Why there should be this difference remains to be investigated; perhaps, the possibility of VP-inversion with *zu*-infinitives in Swiss German can be related to the pervasiveness of ascending structures in these varieties.[25]

---

[25] Note that an analysis of the 3rd construction in Swiss German in terms of VP-inversion makes clear predictions with respect to scope reconstruction. Since no remnant movement is involved, we would expect scope reconstruction to be possible, as in VPR-structures like (48-a).

Before concluding, I would like to add a few empirical details concerning the missing *z* construction. The phenomenon appears to be very subtle. I have not been able to elicit it in an informal survey with linguists speaking a Swiss German dialect by means of a translation task, and in judgment tasks, examples with missing *z* were often rejected. This may be related to the above-mentioned preference for finite subordinate clauses in Swiss German. Many speakers also allow for the Standard German variant with two *zu*s, suggesting that extraposition may also be an option in their grammar. According to Cooper (1995: 188f.), who discusses the phenomenon in some detail, missing *zu* is limited to Verb Raising cases and is blocked whenever there is non-verbal material between V$_2$ 'try' and V$_3$. Since her example on p. 189 strikes me as very unnatural, I've constructed a pair based on (56) with the judgments taken from Cooper:

(59) a. VPR/3rd: 2x *z Swiss German*
De Hans schiint$_1$ *(z) probiere$_2$ [siine Fründe] z hälffe$_3$.
the John seems to try.INF his.DAT friends to help.INF
'John seems to try to help his friends.'

b. VR: 1 *z Swiss German*
De Hans schiint$_1$ [siine Fründe] (z) probiere$_2$ z hälffe$_3$.
the John seems his.DAT friends to try.INF to help.INF
'John seems to try to help his friends.'

I have not been able to replicate this contrast in my survey.[26] Data from the internet are only of limited help as there are only very few hits; in fact, I have been able to find only two relevant examples:

(60) a. … au ohni probiere z wohrsagere
also without try.INF to prophesy.INF
'without trying to prophesy'[27]

The facts are subtle, and I will limit myself to pointing out that Cooper (1995: 197, 199, fn. 39) argues that scope reconstruction *is* possible in the 3rd construction in Zurich German.

[26] The same goes for Cooper's examples 87 on p. 193 and 91 on p. 194f.; her data generally seem somewhat dubious to me.

[27] http://thats-me.ch/forum/em-gewinner/20/31, found on March 28, 2013.

b. … ohni öpe jeh mau säuber probiere, Dütsch z rede
without PRT ever once self try.INF German to speak.INF
'without ever trying to speak German oneself'[28]

Example (60-b) is a counterexample to Cooper's claim. If missing *z* were indeed restricted to VR-structures, they would constitute a problem for the present account since inversion of V_2 with FP_2 would have to depend on whether VP_2 contains non-verbal material; obviously, there is no simple way of ensuring this. Bader (1995) and Vogel (2009) also do not predict a VPR/VR-asymmetry. Future research will have to determine the precise properties of the construction, but for now I will simply conclude that it is another phenomenon that can be covered by the rule-based approach. The empirical coverage between the derivational/rule-based proposed here and the realizational approach is similar, but in my view the derivational account is superior in providing a more interesting account of the cross-linguistic variation and in requiring fewer stipulations to rule out unlimited downward percolation of the *zu*-feature.

3.7. PF-Rules in a Right-Branching Sructure

Until now, I have presupposed a left-branching structure as the input for the PF-operations. In this subsection, I will briefly evaluate the prospects of an analysis based on a right-branching structure. I will ignore the placement of non-verbal elements such as objects; for what follows, it is immaterial whether they obtain their preverbal position by means of movement as in strongly anti-symmetric approaches (e.g. Zwart 1994) or whether they are directly linearized as left-hand sisters of V as in approaches that employ linearization parameters that are sensitive to syntactic category (cf. e.g. Cooper 1995, Barbiers 2000, Schmid and Vogel 2004).

I will start with Dutch where things remain straightforward. For the simple ascending cases like (9-b) the starting point will be stacked right-branching VPs. As a language-particular property, all non-verbal constituents have to move out of the VP (for discussion of the evacuation operation, cf. Salzmann 2011).

---

[28] http://www.chefkoch.de/forum/2,22,296109/An-alle-CHer-Wir-zelebrieren-den-Kantoenligeist.html, found on March 28, 2013.

(61) *Standard Dutch*
zonder [$_{vP}$ het boek [$_{FP}$ te [$_{VP_1}$ hebben$_1$ [$_{VP_2}$ t$_{het\ boek}$ gelezen$_2$]]]]
without the book to have.INF read.PRT
‘without having read the book’

Te will then lower onto V$_1$. Dutch also allows descending orders to a limited degree, especially with participles. In those cases, *te* also appears on V$_1$:

(62) *Standard Dutch*
zonder het boek gelezen$_2$ **te** hebben$_1$
without the book read.PRT to have.INF
‘without having read the book’

This implies that *te*-lowering has to precede reordering – as under the left-branching analysis. Since descending orders in Dutch are as impenetrable as their German counterparts, I will assume that they are also the result of cluster-formation based on adjacency. Note that given a right-branching base, descending orders no longer involve string vacuous cluster formation but cluster formation with reordering. String-vacuous cluster formation is still needed for 312 clusters, cf. fn. 29.

Turning now to German and starting with simple descending structures like (9-a), repeated for convenience:

(63) 321 *Standard German*
Er dachte, das Buch [lesen$_3$ können$_2$ **zu** müssen$_1$].
he thought the book read.INF can.INF to must.INF
‘He thought he had to be able to read the book.’

The starting point will be a right-branching VP-structure with *zu* at the beginning of the verb cluster (as in Dutch, the object has scrambled out of the lexical VP):

(64) ... [$_{vP}$ das Buch [$_{FP}$ zu [$_{VP_1}$ müssen$_1$ [$_{VP_2}$ können$_2$ [$_{VP_3}$ t$_{das\ Buch}$
the book to must.INF can.INF
lesen$_3$]]]]]
read.INF

The major challenge for a right-branching account is the pre-final position of *zu*. I postulate an inversion rule that inverts *zu* (= F) with VP$_1$ so that it occurs after the last element of the verb cluster:

(65) … [$_{\text{vP}}$ das Buch [$_{\text{FP}}$ [$_{\text{VP}_1}$ müssen$_1$ [$_{\text{VP}_2}$ können$_2$ [$_{\text{VP}_3}$ t$_{\textit{das Buch}}$
the book must.INF can.INF
lesen$_3$]]] zu]]
read.INF to

The descending order in the verb cluster is derived by means of reordering cluster formation. Reordering has to precede *zu*-cliticization since *zu* ends up on V$_1$ in this case. The result is illustrated in the following example:

(66) [$_{\text{vP}}$ das Buch [$_{\text{FP}}$ [$_{\text{V}}$ lesen$_3$+können$_2$+**zu**+müssen$_1$] t$_{zu}$]]
the book read.INF+can.INF+to+must.INF

The cases of misplaced *zu* work similarly. I will illustrate the workings on the basis of the ascending VPR-example (14), repeated for convenience:

(67) 1X2 *Swiss German*
ohni mi [welle$_1$ [uf d bullesite **z** stelle$_2$]], im gegeteil,
without me want.INF on the cops.side to put.INF on.the contrary
aber …
but
'without wanting to side with the cops, on the contrary, but …'

The starting point is again a stacked VP-structure:

(68) 12, ex. (14) *Swiss German*
ohni mi [$_{\text{FP}}$ z [$_{\text{VP}_1}$ wele$_1$ [$_{\text{VP}_2}$ t$_{mi}$ uf d Bullesiite stelle$_2$]]]
without me to want.INF on the cops.side put.INF

Then, F is inverted with VP$_1$, placing *z* at the end of the VP:

(69) 12, ex. (14) *Swiss German*
ohni mi [$_{\text{FP}}$ [$_{\text{VP}_1}$ wele$_1$ [$_{\text{VP}_2}$ t$_{mi}$ uf d Bullesiite stelle$_2$]] z]
without me want.INF on the cops.side put.INF to

Then, *zu*-cliticization applies and "mis"places *zu* on V$_2$.[29]

The left-branching cases with two *zu*s as in (10-a) are next. I repeat the relevant example for convenience:

(70) 321 *Standard German*
dass er das Buch **zu** lesen$_3$ **zu** versuchen$_2$ versprach$_1$
that he the book to read.INF to try.INF promised
'that he promised to try to read the book'

The starting point would look as follows (with the object having scrambled to Specv):

(71) dass er [$_{vP}$ das Buch [$_{VP_1}$ versprach$_1$ [$_{FP_1}$ zu [$_{VP_2}$ versuchen$_2$ [$_{FP_2}$ zu
that he the book promised to try.INF to
[$_{VP_3}$ t$_{das\ Buch}$ lesen$_3$]]]]]]
read.INF

Then, F$_1$ inverts with VP$_2$ and F$_2$ inverts with VP$_3$ to put *zu* at the end of the respective non-finite phrase:

(72) dass er [$_{vP}$ das Buch [$_{VP_1}$ versprach$_1$ [$_{FP_1}$ [$_{VP_2}$ versuchen$_2$ [$_{FP_2}$ [$_{VP_3}$
that he the book promised try.INF
t$_{das\ Buch}$ lesen$_3$] zu]] zu]]]
read.INF to to

We argued above that (reordering) cluster formation takes place before *zu*-cliticization. However, if we first form the entire cluster consiting of V$_3$+V$_2$+V$_1$, we would end up with both *zu*s next to each other, which may trigger haplology. Then, the remaining *zu* would arguably be affixed onto V$_1$, which is, of course, the wrong result:

(73) 321 *Standard German*
*dass er das Buch lesen$_3$ versuchen$_2$ **zu** versprach$_1$
that he the book read.INF try.INF to promised
'that he promised to try to read the book'

---

[29] The analysis of the 'scandal construction' (12-a) is similar, but slightly more complex: To derive a 312 order from a linear structure, we have to allow for string-vacuous cluster formation between V$_2$ and V$_3$ followed by reordering cluster formation between V$_1$ and [V$_2$+V$_3$]. Both processes need to precede *zu*-cliticization, which is unproblematic under a cyclic PF-derivation.

It rather seems that we have to intersperse cluster formation with *zu*-cliticization so that *zu* ends up on V$_2$ and V$_3$. This is exactly the result that obtains under a cyclic bottom-up derivation: If we assume that the bracketed structure in (72) is still available and derive the structure bottom-up, the first step will be to cliticize F$_2$ onto V$_3$. Then, the complex *zu*$_2$+V$_3$ would undergo reordering cluster formation with V$_2$, leading to *zu*$_2$+V$_3$+V$_2$. Then, *zu*-cliticization of F$_1$ onto V$_2$ would apply, leading to *zu*$_2$+V$_3$+*zu*$_1$+V$_2$. Finally, the entire complex would undergo reordering cluster formation with V$_1$, producing the correct output *zu*$_2$+V$_3$+*zu*$_1$+V$_2$+V$_1$. Importantly, a cyclic derivation also produces the correct result for simple descending cases like (9-a) where the entire verbal complex is formed before *zu*-cliticization can apply.

The missing *z* facts in (56) can be handled quite straightforwardly: I repeat the relevant example from above:

(74) *Bernese German*
wüu dr Hans sine Fründe schiint$_1$ probiere$_2$ z häuffe$_3$
because the John his.DAT friends seems try.INF to help.INF
'because John seems to try to help his friends'

The starting point will be the following structure:

(75) wü dr Hans [$_{vP}$ sine Fründe [$_{VP_1}$ schiint$_1$ [$_{FP_1}$ z [$_{VP_2}$
because the John his.DAT friends seems to
probiere$_2$ [$_{FP_2}$ z [$_{VP_3}$ t$_{\textit{sine Fründe}}$ häuffe$_3$]]]]]]
try.INF to help.INF

Then, F$_1$ inverts with VP$_2$ and F$_2$ with VP$_3$:

(76) wü dr hans [$_{vP}$ sine Fründe [$_{VP_1}$ schiint$_1$ [$_{FP_1}$ [$_{VP_2}$ probiere$_2$
because the John his.DAT friends seems try.INF
[$_{FP_2}$ [$_{VP_3}$ t$_{\textit{sine Fründe}}$ häuffe$_3$] z]] z]]]
help.INF to to

Then, a haplological rule reduces the two *zu*s to one and *zu*-cliticization applies, deriving the desired result.

What remains to be discussed is the 3rd construction in Standard German with 2 *zu*s as in (36-b). I repeat the relevant example:

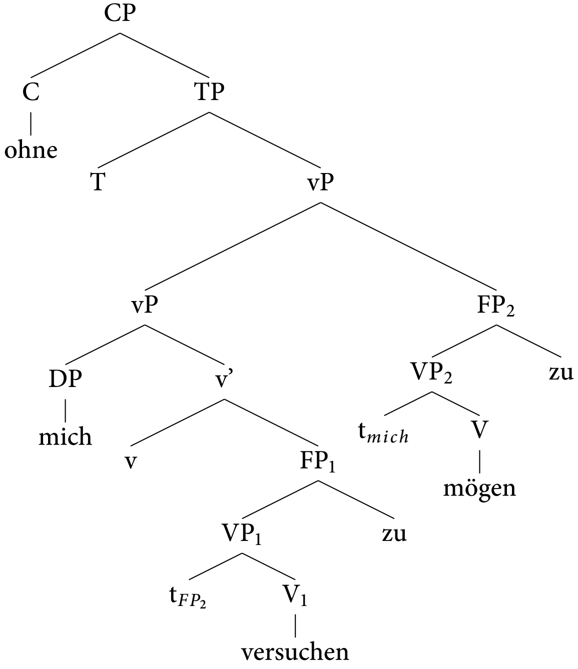
(77) VP-complement/3rd construction, *Standard German*
ohne mich [$_{VP}$ zu versuchen$_1$ [$_{VP}$ zu mögen$_2$]]
without me to try.INF to like.INF
'without trying to like me'

The basis will be the following structure:

(78) ohne [$_{vP}$ mich [$_{FP_1}$ zu [$_{VP_1}$ versuchen$_1$ [$_{FP_2}$ zu [$_{VP_2}$ t_{mich}
without me to try.INF to
mögen$_2$]]]]]
like.INF

It must not be derived like the missing *z* case because both *zu*s have to be retained. The only possibility to derive the correct result is extraposition of FP$_2$ to a position above FP$_1$, e.g. vP. Then, F$_1$ can invert with VP$_1$ and F$_2$ with VP$_2$, leading to the following tree structure:

(79) The 3rd construction in Standard German under a right-branching VP-structure



Then, *zu*-cliticization can apply in both VPs. Note that for the same reason extraposition must also be assumed for CP-complements because otherwise the *zu* assigned to matrix VP_1 would end up on the last verb of the complement CP, i.e. the problem is exactly the same as under the left-branching structure in (42).

As has been shown in this subsection, it is also possible to determine a non-contradictory ranking on the basis of a right-branching structure. The complete ordering is as follows:

(80) Ordering of operations (right-branching): synopsis

extraposition (RC, CP/FP_2) > topicalization (VP) >	syntax
te-lowering/*zu*-inversion > cluster formation (+/-inversion) > < *zu*-cliticization	PF

The result is quite similar to the one in table (51): The major difference is that VP-inversion is no longer necessary while a rule of *zu*-inversion had to be added. The relative ordering of *te*-lowering and *zu*-inversion cannot be determined for obvious reasons. As under a left-branching structure, the ordering in syntax follows from cyclicity, and the same goes for the relative ordering between cluster formation and *zu*-cliticization. In other words, the entire ordering in (80) is intrinsic: either because of cyclicity in syntax (extraposition vs. topicalization) and PF (reordering cluster formation and *zu*-cliticization) or because the operations take place in separate subcomponents of PF (*te*-lowering has to take place before the adjacency-sensitive operations). In the ordering in (51), however, the ordering between *te*-lowering and VP-inversion requires extrinsic ordering.

This may constitute slight advantage for a right-branching approach. But there are two aspects that seem suboptimal: First, extraposition for the 3rd construction and CP-complements is still necessary (at least in German) even though they can be directly linearized as right-hand sisters of V. This undermines one – independent – argument in favor of a right-branching structure (Zwart 1994); but since a left-branching structure requires obligatory extraposition as well (and is faced with the same questions w.r.t. a plausible trigger), this is probably not too detrimental. The only rule that seems quite stipulative and which can be avoided under a left-branching approach is *zu*-inversion. While this leads to a complication for German, it should be pointed out that generating the infinitival particle in a functional projection above (and thus before) the verb cluster has advantages for West-Flemish where *te* is not associated with the verb, recall (28-a). Under a left-branching approach, one probably has to postulate a corresponding rule of *te*-inversion to move it to the beginning of the verb cluster in West Flemish while no such rule is necessary under a right-branching account.

The following table lists the relevant phenomena with the rules required under both a left-branching and a right-branching approach ('svCF' stands for string-vacuous cluster formation, 'rCF' for reordering cluster formation, and 'hapl' for haplology; the other abbreviations should be self-explanatory):

(81) Phenomena and derivations

		left-branching	**right-branching**
32*zu*1	(9-a)	svCF > *zu*-clit	*zu*-inv > rCF > *zu*-clit
*zu*3*zu*21	(10-a)	*zu*-clit > svCF > *zu*-clit > svCF	*zu*-inv (2x) > *zu*-clit > rCF > *zu*-clit > rCF
13*zu*2	(11-c)	VP-inv > *zu*-clit > svCF	*zu*-inv > rCF > *zu*-clit
31*zu*2	(12-a)	rCF > svCF > *zu*-clit	*zu*-inv > svCF > rCF > *zu*-clit
*zu*1*zu*2	(36-b)	extrapos > *zu*-clit (2x)	extrapos > *zu*-inv (2x) > *zu*-clit (2x)
*zu*1CP	(42)	extrapos > *zu*-clit	extrapos > *zu*-inv > *zu*-clit
12*zu*3	(56)	VP-inv (2x) > hapl > *zu*-clit	*zu*-inv (2x) > hapl > *zu*-clit
*te*123	(9-b)	*te*-lowering> VP-inv (2x)	*te*-lowering
2*te*1	(62)	*te*-lowering > svCF	*te*-lowering > rCF

This table suggests that fewer operations are needed under a left-branching account. This tends to be correct for German and its varieties where descending orders are frequent. Once we look at Dutch, where ascending orders predominate, a right-branching VP-structure provides the better input for the PF-rules. Consequently, once the larger picture is taken into account, the price to be paid seems to be similar in both approaches. While a left-branching VP-structure provides the simpler solution for descending structures, a right-branching VP-structure is superior for ascending structures.

4. Conclusion

In this paper, I have addressed two puzzles in the domain of verb cluster formation in terms of rule ordering. I first discussed the extraposition paradox where extraposition to a VP that is part of a verb cluster is blocked when the VP is in-situ but not when it is topicalized. I have argued that the verbs have to form a complex head when adjacent in descending order. In contrast to earlier approaches, cluster formation takes place post-syntactically and is subject to an adjacency requirement. This explains why extraposition to the non-final VP in V-final structures leads to ungrammaticality: cluster formation is blocked as the adjacency is disrupted; extraposition thus bleeds cluster formation. Since cluster formation takes place after topicalization, nothing prohibits extraposing to the lexical VP if it is later moved to the beginning of the clause. In that case, topicalization bleeds cluster formation as the context for the latter operation is destroyed (there is no sequence of verbs anymore).

The second puzzling phenomenon discussed was the placement of the infinitival particle in Dutch and German. I have assumed that the particle is an independent syntactic element in both languages. The cross-linguistic differences result from the fact that the operation that associates the particle with the verb takes place at different points of the derivation in the two languages. While it is an early PF-process in Dutch that is still sensitive to hierarchical structure, it is a late process in German because it is sensitive to linear order and adjacency. Thus, while *te*-placement represents an instance of Lowering, *zu*-cliticization is best described as an instance of Local Dislocation in the framework of Embick and Noyer (2001).

I have shown that these two processes interact with other PF-rules such as cluster formation and VP-inversion. It turned out to be possible to determine a strict and non-contradictory ordering between these rules. Furthermore, with one exception under a left-branching VP-structure, the ordering is fully intrinsic – either because it follows from cyclicity (in syntax and PF) or because the operations take place in different components (syntax vs. PF or different subparts of PF). I take this to lend support to the rule-based approach pursued here. Furthermore, it can be seen as an initial attempt to provide insight into the articulation of the PF-component with earlier rules being more sensitive to hierarchical structure while later rules operate on linear structure.

On a more general theoretical level, a comparison between a left-branching and a right-branching approach has not revealed any significant advantages for either of the approaches. Rather, while a left-branching approach is best suited to derive descending orders, a right-branching approach provides a simpler account of ascending structures. Finally, a consistent account of *zu*-placement requires a remnant extraposition analysis of the 3rd construction (at least in Standard German) and more generally obligatory extraposition of right-hand VP and CP-complements of V.

References

Arregi, Karlos and Andrew Nevins (2012): *Morphotactics. Basque Auxiliaries and the Structure of Spellout*. Springer, Dordrecht.

Bader, Markus and Tanja Schmid (2009): 'Verb Clusters in Colloquial German', *The Journal of Comparative Germanic Linguistics* 12(3), 175–228.

Bader, Thomas (1995): Missing and Misplaced z' in Bernese Swiss German. *In:* Z. Penner, ed., *Topics in Swiss German syntax*. Peter Lang, Bern, pp. 19–28.

Barbiers, Sjef (2000): The Right Periphery in SOV Languages: English and Dutch. *In:* P. Svenonius, ed., *The Derivation of VO and OV*. John Benjamins, Amsterdam, pp. 181–218.

Barss, Andrew (1986): Chains and Anaphoric Dependencies. Doctoral dissertation, MIT.

Bayer, Josef, Tanja Schmid and Markus Bader (2005): Clause Union and Clausal Position. *In:* M. d. Dikken and C. M. Tortora, eds, *The function of function words and functional categories*. John Benjamins, Amsterdam, pp. 79–113.

Bech, Gunnar (1983): *Studien über das deutsche Verbum infinitum*. Niemeyer, Tübingen.

Bobaljik, Jonathan and Susi Wurmbrand (2005): 'The Domain of Agreement', *Natural Language & Linguistic Theory* 23(4), 809–865.

Broekhuis, Hans (1993): 'Verb Projection Raising', *Spektator* 22, 28–47.

Chomsky, Noam (1973): Conditions on Transformations. *In:* S. R. Anderson and P. Kiparsky, eds, *A Festschrift for Morris Halle*. Holt, Rinehart and Winston, New York, pp. 232–286.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Cooper, Kathrin E. (1995): *Topics in Zurich German Syntax*. Germanistisch Instituut, Groningen.

Dikken, Marcel den (1996): The Minimal Links of Verb (Projection) Raising. *In:* W. Abraham, S. D. Epstein, H. Thrainsson and C. J.-W. Zwart, eds, *Minimal Ideas. Syntactic Studies in the Minimalist Framework*. Benjamins, Amsterdam, pp. 67–96.

Dikken, Marcel den and Eric Hoekstra (1997): 'Parasitic Participles', *Linguistics* 35(6), 1057–1090.

Donaldson, Bruce C. (1993): *A Grammar of Afrikaans*. Mouton de Gruyter, Berlin/New York.

Embick, David and Rolf Noyer (2001): 'Movement Operations after Syntax', *Linguistic Inquiry* 32(4), 555–595.

Evers, Arnold (1975): The Transformational Cycle of Dutch and German. Doctoral dissertation, University of Utrecht.

Geilfuss-Wolfgang, Jochen (1991): *Verb- und Verbphrasensyntax*. Arbeitspapier des SFB 340 11, University of Tübingen, Tübingen.

Haegeman, Liliane (1992): *Theory and Description in Generative Syntax. A Case Study in West Flemish*. Cambridge University Press, Cambridge.

Haegeman, Liliane (1998): 'Verb Movement in Embedded Clauses in West Flemish', *Linguistic Inquiry* **29**(4), 631–656.

Haegeman, Liliane and Henk van Riemsdijk (1986): 'Verb Projection Raising, Scope, and the Typology of Rules Affecting Verbs', *Linguistic Inquiry* **17**(3), 417–466.

Haider, Hubert (1993): *Deutsche Syntax – Generativ: Vorstudien zur Theorie einer projektiven Grammatik*. Narr, Tübingen.

Haider, Hubert (2003): V-Clustering and Clause Union: Causes and Effects. *In:* P. Seuren and G. Kempen, eds, *Verb Constructions in German and Dutch*. John Benjamins, Amsterdam, pp. 91–126.

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

Haider, Hubert (2011): 'Grammatische Illusionen – Lokal wohlgeformt – global deviant', *Zeitschrift für Sprachwissenschaft* **30**(2), 223–257.

Halle, Morris and Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. *In:* K. Hale and S. J. Keyser, eds, *The view from building 20. Essays in linguistics in honour of Sylvain Bromberger*. MIT Press, Cambridge, pp. 111–176.

Hinterhölzl, Roland (2009): 'The IPP-Effect, Phrasal Affixes and Repair Strategies in the Syntax-Morphology Interface', *Linguistische Berichte* **2009**(218), 191–215.

Hodler, Werner (1969): *Berndeutsche Syntax*. Francke Verlag, Bern.

Höhle, Tilman N. (2006): Observing Non-Finite Verbs: Some 3v Phenomena in German-Dutch. *In:* P. Brandt and E. Fuß, eds, *Form, Structure, and Grammar. A Festschrift Presented to Günther Grewendorf on Occasion of His 60th Birthday*. Akademie Verlag, Berlin, pp. 55–77.

Kayne, Richard S. (1994): *The Antisymmetry of Syntax*. MIT Press, Cambridge, Mass.

Neeleman, Ad and Fred Weerman (1993): 'The Balance between Syntax and Morphology: Dutch Particles and Resultatives', *Natural Language & Linguistic Theory* **11**(3), 433–475.

Richards, Marc (2008): 'Desymmetrization: Parametric Variation at the PF-interface', *Cancadian Journal of Linguistics* **53**(2/3), 275–300.

Riemsdijk, Henk van (1998): 'Head Movement and Adjacency', *Natural Language & Linguistic Theory* **16**(3), 633–679.

Salzmann, Martin (2011): Resolving the Movement Paradox in Verb Projection Raising. *In:* O. Bonami and P. Cabredo Hofherr, eds, *Empirical Issues in Syntax and Semantics 8*. CNRS, Paris, pp. 453–485.

Salzmann, Martin (to appear): 'New Evidence for Postsyntactic Verb Cluster Formation and a Right-Branching Base Order', *Proceedings of NELS* **42**.

Sauerland, Uli and Paul Elbourne (2002): 'Total Reconstruction, PF Movement, and Derivational Order', *Linguistic Inquiry* **33**(2), 283–319.

Schallert, Oliver (2012): Untersuchungen zu Ersatzinfinitivkonstruktionen in den Vorarlberger und Liechtensteiner Dialekten. Doctoral dissertation, University of Marburg.

Schmid, Tanja and Ralf Vogel (2004): 'Dialectal Variation in German 3-Verb Clusters: A Surface-Oriented Optimality Theoretic Account', *The Journal of Comparative Germanic Linguistics* 7(3), 235–274.

Stechow, Arnim von (1990): Status Government and Coherence in German. *In:* G. Grewendorf and W. Sternefeld, eds, *Scrambling and barriers*. John Benjamins, Amsterdam, pp. 143–198.

Sternefeld, Wolfgang (1990): Scrambling and Minimality. *In:* G. Grewendorf and W. Sternefeld, eds, *Scrambling and barriers*. John Benjamins, Amsterdam, pp. 239–257.

Sternefeld, Wolfgang (2006): *Syntax: eine morphologisch motivierte generative Beschreibung des Deutschen*. Stauffenburg Verlag, Tübingen.

Vogel, Ralf (2009): 'Skandal im Verbkomplex. Betrachtungen zur scheinbar inkorrekten Morphologie in infiniten Verbkomplexen des Deutschen', *Zeitschrift für Sprachwissenschaft* 28(2), 307–346.

Weber, Albert (1987): *Zürichdeutsche Grammatik: ein Wegweiser zur guten Mundart*. Schweizer Spiegel Verlag, Zürich.

Williams, Edwin (2004): The structure of clusters. *In:* K. É. Kiss and H. v. Riemsdijk, eds, *Verb Clusters: A Study of Hungarian, German, and Dutch*. John Benjamins, Amsterdam, pp. 173–201.

Wurmbrand, Susanne (2004*a*): 'Two Types of Restructuring–Lexical vs. Functional', *Lingua* 114(8), 991–1014.

Wurmbrand, Susanne (2004*b*): West Germanic Verb Clusters: The Empirical Domain. *In:* K. . Kiss and H. v. Riemsdijk, eds, *Verb clusters: A study of Hungarian, German, and Dutch*. John Benjamins, Amsterdam, pp. 43–85.

Wurmbrand, Susanne (2005): Verb Clusters, Verb Raising, and Restructuring. *In:* M. Everaert and H. v. Riemsdijk, eds, *The Blackwell companion to syntax. Volume 5*. Blackwell, Oxford, pp. 227–341.

Wurmbrand, Susanne (2007): 'How Complex Are Complex Predicates?', *Syntax* 10(3), 243–288.

Zwart, Jan-Wouter (1994): 'Dutch is Head Initial', *The Linguistic Review* 11(3/4), 377–406.

The Linearization of Morphological Weight

Eva Zimmermann and Jochen Trommer[*]

Abstract

Affixation of moras is a standard analysis for many types of non-concatenative morphology (e.g. Lombardi and McCarthy 1991, Samek-Lodovici 1992, Davis and Ueda 2002*a*, Grimes 2002*b*, Davis and Ueda 2006, Álvarez 2005, Stonham 2007, Yoon 2008, Haugen and Kennard 2008). However, some basic questions about the nature of mora affixation have never been properly addressed, one of them being the question how moraic (and more generally prosodic) affixes are linearized with respect to their base. Based on a typological survey of mora affixation, we argue that morphological moras are assigned to a fixed position on their tier by morphology and cannot be dislocated by later processes. They are prefixed or suffixed to specific peripheral or prominent elements of their morphological bases on an affix-specific (i.e. phonologically arbitrary) basis. We thus extend the model of segmental affixation in Yu (2002, 2007) (cf. also Fitzpatrick 2004) where affixation targets a specific member of a set of crosslinguistically possible anchor points by lexical subcategorization to prosodic affixation. An important empirical prediction of the subcategorization-based system is that affix moras cannot move to different linear positions under the pressure of phonological constraints. We show that this prediction is correct and argue that apparent counterevidence in Keley-I gemination under the analysis of Samek-Lodovici (1992) is due to a morphological misinterpretation of the data. The alternative approach to the linearization of mora affixes based on the assumption of phonological morpheme dislocation suffers from a severe lack of empirical adequacy: it overgeneralizes and predicts pattern of mora affixation that are unattested and undergeneralizes since it cannot predict all existing patterns of mora affixation.

1. Introduction

One of the major assets of Autosegmental Phonology is that it allows to reduce procedural techniques of morphological exponence to a simple generalized

---

[*] For helpful comments and discussion, we are grateful to the audiences of the 'Phlunch' at UC Santa Cruz, the 'Phorum' at UC Berkeley and the linguistics colloquium at UC San Diego, especially Sharon Inkelas, Junko Ito, Sharon Rose, Eric Bakovic, Larry Hyman, Keith Johnson, Armin Mester, Grant McGuire, and Jaye Padgett.

Rule Interaction in Grammar, 123–161
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

concept of concatenation. In particular, the moraic approach to phonological length (Hayes 1989) gives rise to a maximally simple account of morphologically triggered gemination, vowel lengthening, and coda epenthesis: as affixation of a μ. Although mora affixation is a standard assumption in the literature on nonconcatenative morphology, some basic questions about the nature of mora affixation have never been properly addressed, one of them being their linearization.

In this paper, we argue that prosodic nodes are assigned to a fixed position on their tier by the morphology and cannot be dislocated by later processes. Prosodic nodes are prefixed or suffixed to specific peripheral or prominent elements of their morphological bases on an affix-specific (i.e. phonologically arbitrary) basis. We therefore extend the assumptions about segmental affixation in Yu (2002, 2007) (cf. also Fitzpatrick 2004) that affixation targets a specific member of a set of crosslinguistically possible anchor points by lexical subcategorization to prosodic affixation. An important empirical prediction of the subcategorization-based system is that the linearization of affix-μ's happens in a single step and it is impossible that an affix-μ moves to a different linear positions under the pressure of phonological constraints after the morphology has linearized it to a specific position in its base. Based on a typological survey of μ-affixation, we argue that this prediction is true and argue that apparent counterevidence such as Keley-I gemination under the analysis of Samek-Lodovici (1992) is due to a morphological misinterpretation of the data.

We proceed as follows: In section 2, we discuss the two major classes of affix linearization, one allowing the interaction/ordered application of morphological and phonological demands to linearize an affix, the other neglecting any affix dislocation to optimize phonological structure. An empirical survey lengthening morphology is presented in section 3 where we conclude that a fixed set of affixation pivots is able to predict all attested cases of μ-affixation. Several arguments against the alternative assumption that μ-affixes are linearized in a two-step procedure and can be dislocated under phonological pressure are presented in section 4. It is shown that such an approach overgeneralizes and undergeneralizes at the same time. We conclude in section 5.

2. Rule Ordering and Affix Linearization

2.1. Segmental Affixes

Theories for affix linearization can be divided into two major classes: **Phonological Dislocation** theories and **Morphological Pivot Affixation** theories. The former assume that affixes are prefixed or suffixed to their base but may infix under the pressure of phonological constraints (Moravcsik 1977, Prince and Smolensky 1993/2002, Stemberger and Bernhardt 1998, Halle 2003, Horwood 2002, Klein 2005) and the latter are based on the assumption that affixes are prefixes or suffixes to specific base positions (='pivots') and cannot be dislocated by phonological processes (Yu 2002, 2007). A classic instantiation of a phonological dislocation theory can be found in Prince and Smolensky (1993/2002) where it is argued that infixation in Tagalog (Bloomfield 1933, McCarthy and Prince 1993, Zoll 1996, Orgun and Sprouse 1999, Halle 2003, Klein 2005) is an instance of phonologically-motivated dislocation. The generalization for the placement of the actor focus affix *–um* is that it precedes all segmental material of a V-initial base (=*Prefixation*, (1-a)) but follows the first onset consonant for C-initial bases (=*Infixation*, (1-b)).

(1) *Tagalog* um-*Infixation* (Prince and Smolensky 1993/2002, Orgun and Sprouse 1999)

	BASE	ACTOR FOCUS	
a.	abot	'reach for'	**um**abot
	aral	'teach'	**um**aral
b.	sulat	'write'	s**um**ulat
	gradwet	'graduate'	gr**um**adwet

The analysis for these facts in Prince and Smolensky (1993/2002) is based on the assumption that *–um* is a prefix that may dislocate inside its base in order to avoid the creation of additional marked coda consonants. The OT-implementation of this intuition is briefly illustrated in the tableau (2). A morpheme-specific EDGEMOST (or ALIGN in, for example, McCarthy and Prince 1993, Zoll 1996, Orgun and Sprouse 1999) constraint demands that the affix *um–* must be realized as a prefix and is violated by any segment intervening between the *um–* and the left edge of the word. The markedness constraint *NOCODA penalizes any coda consonant and since it is ranked higher, dislocation of the affix (2-b) is predicted if it avoids an additional coda consonant.

(2) *Infixation as phonological dislocation* (Prince and Smolensky 1993/2002)

a. *V-initial Base*

um, abot	NoCoda	Edgemost(um,L)
☞ a. **u.ma**.bot	*	
b. a.**um**.bot	**!	*
c. a.b**u.m**ot	*	*!*

b. *C-initial Base*

um, tawag	NoCoda	Edgemost(um,L)
a. **um**.ta.wag	**!	
☞ b. t**u.ma**.wag	*	*
c. ta.**um**.wag	**!	**

This dislocation theory is crucially based on the simultaneous interaction of phonological and morphological rules/constraints that determine the place in the base where an affix is realized. Another implementation of phonological dislocation is the assumption that morphology linearizes an affix as prefix/suffix and that phonology applies afterwards, potentially dislocating the affix (Horwood 2002). An implementation of the Tagalog facts would be similar to the one in (2), with the only difference that the Edgemost/Align constraint is replaced with a Linearity constraint ensuring that the affix remains faithful to the underlying linearization as a prefix as can be seen in (3).

(3) *Infixation as Affixation und subsequent Phonological Dislocation* (Horwood 2002)

a. *V-initial Base*

um-abot	NoCoda	Linearity
☞ a. **u.ma**.bot	*	
b. a.**um**.bot	**!	*
c. a.b**u.m**ot	*	*!*

b. *C-initial Base*

um-tawag	NoCoda	Linearity
a. **um**.ta.wag	**!	
☞ b. t**u.ma**.wag	*	*
c. ta.**um**.wag	**!	**

Horwood (2002) argues that this account is superior to an EDGEMOST/ALIGN-based account since it avoids the misprediction of 'morphemic bitropism' where one morpheme can be forced to be aligned with the left and right edge of its base at the same time. In addition, typological tendencies about affix ordering (e.g. the preference for number markers to be more stem-inwards than case markers; Greenberg 1963) remain 'at best an accident, with cross-linguistically rerankable constraints in the phonological component positioning morphemes independently of any morphosyntactic or semantic principles' (Horwood 2002: 6+7).

Although phonological dislocation analyses capture an intuitive generalization about the Tagalog facts, there are a number of striking counterexamples to the claim that dislocation is phonologically optimizing (Fitzpatrick 2004, Yu 2007). A typical example is the nominalizing infix –*ni*– in Leti, an Austronesian language spoken on the island of Leti. The affix appears consistently after the first consonant of a C-initial base even though this makes syllable structure worse, not better. Thus the putative form ***ni**-*kaːti* avoids the complex onset of *k*-**ni**-*aːti*, and has otherwise the same amount of codas and onsets. Since complex onsets are well-established cases of phonologically marked structure, a phonological dislocation approach predicts that the Leti marker should not infix – contra to fact.

An alternative to phonological dislocation theories is the assumption of morphological pivot affixation as in Yu (2007). The only necessary assumption for an analysis of Tagalog is here that the affix is a prefix to the first base vowel. This simple subcategorization statement predicts the correct linearization of –*um* for all contexts.

(4) um ↔ $_{\text{Base}}[\ \ldots \_\_\_\ \text{V}$

A crucial feature of pivot theories is that the set of pivots is strictly limited. The exhaustive list of pivots that are assumed in Yu (2007) is given in (5) (see Fitzpatrick 2004 for a slightly different inventory of anchor points).

(5) *Possible pivots for affixation* (Yu 2007)

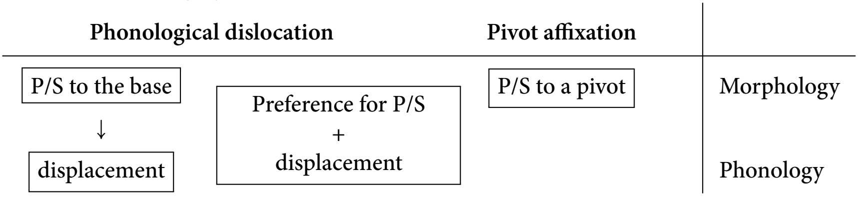
a. **Initial pivot**
   (i) First consonant/onset
   (ii) First vowel/nucleus
   (iii) First syllable

b. **Final pivot**
    (i) Final vowel/nucleus
    (ii) Final syllable
c. **Prominence pivot**
    (i) Stressed syllable
    (ii) Stressed vowel/nucleus

It has to be noted, though that the assumption of base-internal morphological pivots as a source of infixation is in principle independent of the possibility that in addition morphological prefixes and suffixes dislocate (infix) later due to phonological mechanisms. Yu (2007) explicitly rejects any phonological dislocation (or 'phonological readjustment') to account for infixation, and we think, with good reasons since a theory combining morphological *and* phonological infixation would not make any interesting empirical predictions. On the other hand, understanding morphological and phonological infixation as mutually exclusive theoretical options makes the morphological linearization of affixes an empirical area which provides evidence on the question whether morphological operations are carried out in parallel interaction with standard optimization (Wolf 2008, 2013), or are independent and derivationally prior to phonological optimization (Arregi and Nevins 2012). Whenever we refer in the following to morphological pivot affixation theories, we imply a restrictive theory of affixation that disallows any further dislocation at a later stage in the derivation.

The two major classes of affix linearization discussed above can be summarized as in (6) where the abbreviation 'P/S' stands for the morphological placement of an affix before or following all base material.

(6) *Theories of affix linearization*



In the domain of segmental affixation, Yu (2007) has argued convincingly for the morphological pivot affixation theory and against a phonological dislocation approach. In this paper, we investigate the linearization of affixes whose linearization properties are scarcely discussed in the theoretical literature: morphological μ's.

2.2. μ-Affixes

A striking argument for prosodic affixation in general and μ-affixation in specific is the existence of non-concatenative allomorphy, defined as the phenomenon that one morphological category is expressed through different non-concatenative operations. An instance of non-concatenative allomorphy can be found in the Shizuoka dialect of Japanese where the emphatic adjective is formed through either gemination (7-a), vowel lengthening (7-b), or insertion of an epenthetic nasal (7-c). The choice between these different allomorphs is phonologically predictable as is analyzed in detail in section 4.3.

(7) *Emphatic adjectives in Shizuoka Japanese* (Davis and Ueda 2006)

	Adjective		Emphatic Form			
a.	katai	'hard'	ka**t**ːai			
	osoi	'slow'	o**s**ːoi	CV.C̥…	⇒	CV.C̥ː…
	takai	'high'	ta**k**ːai			
b.	hade	'showy'	ha**n**de			
	ozoi	'terrible'	o**n**zoi	CV.C̬…	⇒	CV**N**.C̬…
	nagai	'long'	na**ŋ**gai			
c.	zonzai	'impolite'	z**o**ːnzai			
	supːai	'sour'	s**u**ːpːai	CVC.C…	⇒	CVːC.C…
	okːanai	'scary'	**o**ːkːanai			

The striking observation about the non-concatenative allomorphy in Shizuoka Japanese is now that all these three different strategies add prosodic weight to the first syllable: it is light in the normal adjective form but heavy in the emphatic adjective. The straightforward analysis for the Shizuoka Japanese emphatic adjective formation and similar patterns of length-manipulation is therefore the affixation of a morphological μ that must be realized through integrating it in the prosodic structure of its base Davis and Ueda (2002*b*, 2006).

Although there are numerous theories assuming μ-affixation (e.g. Lombardi and McCarthy 1991, Samek-Lodovici 1992, Davis and Ueda 2002*a*, Grimes

2002*b*, Davis and Ueda 2006, Álvarez 2005, Stonham 2007, Yoon 2008, Haugen and Kennard 2008), there is no consensus about the mechanism ensuring where in its base such an additional μ is realized. Even worse, the question of how the μ-affix is linearized is hardly ever explicitly discussed in detail with all its typological predictions. Before we undertake such a thourough discussion of the attested linearizations for μ-affixes in section 3, we briefly discuss the two major positions that exist in the literature to restrict a μ-affix to the position in its base where it is realized.

In the first analysis of morphological lengthening that assumed a μ-affix (Samek-Lodovici 1992), a version of a phonological dislocation theory for the μ-affix is assumed that is quite similar to the one in Prince and Smolensky (1993/2002) for segmental affixes. A constrain Left ensures that the μ is realized as close to the left edge of the word as possible: it is violated by every syllable intervening between the syllable that integrates the affix-μ and the left edge of the word. Similar implementations of phonological dislocation theories for μ-infixation can be found in the analysis for Shizuoka Japanese given in Davis and Ueda (2002*b*) and in Grimes (2002*a*) for gemination in Muskogean. Note that these Align-constraints, for example the definition in (8) from Davis and Ueda (2002*b*), are morpheme-specific.

(8) Align(μ_C,Wd) (Davis and Ueda 2002*b*: 4)
Align the emphatic mora with the beginning (left edge) of the word.

A slightly different version of μ-Align is sketched in the analysis for Hiaki in Haugen and Kennard (2008). A morphologically indexed constraint Affix-Left is violated by every segment between the segment dominated by the morphological μ and the left edge of the word. This is therefore a more restrictive dislocation approach since the constraint counts segments intervening between the μ and the word edge, not only syllables.

The theory of prosodic circumscription (Lombardi and McCarthy 1991, McCarthy and Prince 1990, McCarthy 2000), on the other hand, can be interpreted as a version of morphological pivot affixation in that affixes prefix or suffix to a specific anchor point in their base and this anchor point is potentially inside the base. The crucial assumption is that bases can be (recursively) delimited to certain prosodically defined portions and both parts (the 'outparsed' portion and the (extraprosodic) 'remainder') can then be targetted by further (morphological) operations. In the analysis of medial gemination in Choctaw in Lombardi and McCarthy (1991), a μ-affix is prefixed to a base form that is created through making the first μ of the base extraprosodic.

In the following sections, we explicitly argue against a phonological dislocation theory for μ-affixation (Samek-Lodovici 1992, Grimes 2002*a*, Davis and Ueda 2002*b*) where a morphological μ can dislocate inside its base in order to optimize phonological structure. The general prediction of a phonological dislocation approach to affixation is highly heterogeneous variation in the linearization options of single affixes, different affixes in specific languages and across languages since every combination of a high-ranked markedness constraint with bases and affixes of different phonological shape might lead to substantially different patterns of affix linearization. Thus constraints on vowel harmony might steer an infix containing a vowel to a position after a harmonizing syllable under affixation to a disharmonic base, whereas the same affix might undergo local dislocation due to syllabic well-formedness when attached to a harmonic base. We argue in section 4 that these predictions are problematic. In contrast, we argue for a morphological pivot affixation theory assuming a rather small set of possible pivots for μ-affixation and excluding any subsequent dislocation of affixes in the phonology. This predicts that infixation is rather stable inside and across languages. Neither the prosodic circumscription theory of Lombardi and McCarthy (1991) nor the ALIGNMENT-based theory of Davis and Ueda (2002*b*) make any predictions about *im*possible infixation patterns. Can a μ prefix/suffix to any prosodically delimited base? Can the ALIGNMENT constraint (8) refer to any prosodic constituent in the base?

The patterns of μ-affixation in the languages of the world are severely restricted to certain patterns and a theory allowing unrestricted infixation suffers from a serious overgeneralization problem as is discussed in the next section 3.

3. A Typology of Mora Affixation

In this section, we present an empirical survey of μ-affixation cases and conclude that a fixed set of pivots allows to predict all and only the attested cases of μ-affixation in the languages of the world.

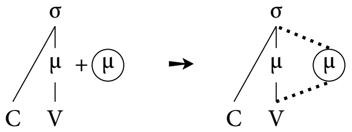
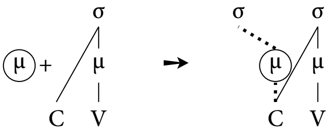
There are basically two types of morphological μ's. *First*, there are instances like Shizuoka Japanese where the augmentation of a μ is the sole exponent of a morpheme. *Secondly*, there are instances in the languages of the world where a segmental affix is always accompanied by lengthening of a base segment that is not phonologically predictable. An example is the plural suffix *–weʔ* in Zuni (9), that is always accompanied by lengthening of the preceding base vowel. This lengthening is not found in similar contexts with other suffixes and can therefore not be regarded as phonological process.

(9) *Plural suffix /–weʔ/ in Zuni* (Newman 1965, Saba Kirchner 2007)

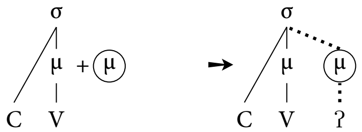
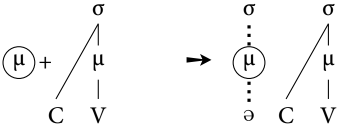
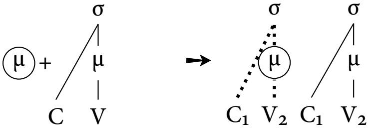
Base		Plural
lupa	'box of ashes'	lup**aː**weʔ
homata	'juniper tree'	homat**aː**weʔ
toːʃo	'seed'	toːʃoːw
kʔapa	'a deep container	kʔapaːw

For our empirical survey of the linearization of affix-μ's, we ignore such instances of lengthening-triggering segmental affixes and only concentrate on the former type of affix-μ's that constitute a morpheme on their own. For such an undertaking, a first obvious criterion is to classify the phenomenon in question as morphological and exclude any phonologically predictable reason for the lengthening. An empirical survey of μ-affixation is obviously a non-trivial task since the affixation of a μ as an abstract timing unit can result in different surface effects while not all specific surface outputs that may result from μ-affixation are necessarily the effect of a morphological μ. There are in principle five phonological strategies to realize an additional μ: vowel lengthening (10-a), consonant gemination (10-b), (coda) consonant epenthesis (10-c), vowel epenthesis (10-d), and reduplication (10-e).

(10) *Realization of a μ-affix (to be continued)*

a. **Vowel lengthening**	
b. **Gemination**	

(10) *Realization of a μ-affix (continued)*

c. C-Epenthesis	
d. V-Epenthesis	
e. Reduplication	

We classify vowel lengthening and consonant gemination as 'strictly μ-induced operations' since both trigger the longer realization of an underlying segment: inside a standard moraic theory, there is simply no other option than to represent those segments as augmented by an additional μ. Epenthesis and reduplication, however, are only 'potentially μ-induced' since additional segments are realized that were not present underlyingly. This actually opens up reasonable alternative theoretical explanations for these strategies that do not involve the addition of a μ. *First*, vowel-epenthesis and reduplication of at least one vowel could very well be the consequence of σ-affixation rather than μ-affixation (cf. Saba Kirchner 2010 for a recent discussion). And *second*, if a morphological category is expressed via segmental 'epenthesis' in all contexts, there is no way to distinguish these epenthetic segments from a 'normal' segmental representation for the morpheme in the lexicon. Since we wanted to reduce the empirical survey to cases where a μ-affixation analysis is one highly plausible theoretical account, we decided to include only those phenomena where at least one of the phonologically predictable allomorphs of a morpheme are 'strictly μ-induced' or where at least two 'potentially μ-induced' operations are phonologically predictable allomorphs of a morpheme. An example for an instance conforming to the former criterion would be Shizuoka Japanese (7) where a 'potentially μ-induced' process (coda epenthesis) alternates with two 'strictly μ-inducing' processes (gemination and vowel-lengthening); hence it is highly plausible to ascribe all of them to the affixation of a μ. Saanich, on the other hand is an example where multiple 'potentially μ-induced' allomorphs cooccur. The continuative is formed through ʔ-insertion and CV reduplication as can be seen in (11)[1] – one of these 'potentially μ-induced' allomorphs would not have been sufficient to include it in the sample but the coexistence of both qualifies it as plausible μ-affixation example.

(11) *Continuative allomorphy in Saanich* (Turner 2007, Kurisu 2001)

	NON-CONTINUATIVE		CONTINUATIVE
a.	*Reduplication*		
	’kʷey	‘to get hungry’	’kʷe’kʷə’y
	’ta’kʷ	‘to go home’	’ta’tə’kʷ
	kʷečəŋ	‘to yell’	kʷə kʷečəŋ’
b.	ʔ-*insertion*		
	weqəs	‘to yawn’	weʔqes
	hes-əŋ	‘to sneeze’	heʔsəŋ’
	ʔiɬən	‘to eat’	ʔiʔɬən

Cases of templatic morphology where a whole base is adjusted to conform to some C/V-template are ignored for this survey as well although they may involve lengthening, gemination, and/or segment insertion and μ-affixation may ultimately be (part of) the correct analysis for these data.[2] As for reduplication and/or epenthesis, alternative theoretical accounts involving the affixation of larger prosodic units than the μ are reasonable and in most cases even required for templatic morphemes. We also excluded languages where the length-manipulation affects the vowel in monosyllabic bases since nothing interesting about the linearization of a μ-affix can be deduced from such facts: every vowel is the first and the last of its base at the same time. An example for such excluded facts are length-manipulation in various Western Nilotic languages where bases are systematically monosyllabic.

The three criteria for the empirical survey of μ-affixation discussed above are summarized in (12).

---

[1] There are other continuative allomorphs in Saanich, one of them being a lexically listed ablaut pattern (Montler 1986, Turner 2007, Leonard and Turner 2010).

[2] For a recent discussion, see, for example, Bye and Svenonius (2012).

(12) The set of phonologically predictable allomorphs *A* expresses a morphological category *M*

- a. **μ-affixation**
  Either (i) or (ii) holds:
  - (i) a 'strictly μ-induced' operation (gemination, vowel lengthening) is one operation in *A*
    or
  - (ii) at least two different strategies from the set of 'potentially μ-induced' operations (C- or V-epenthesis, μ-sized reduplication) are part of *A*
- b. **Exclusion of templatic morphology**
  Not all forms expressing *M* through *A* conform to a prosodic shape that is not phonologically predictable.
- c. **Relevance for linearization**
  At least some bases to which *A* apply are polysyllabic.

Finally, we ensured that the sample is representative and typologically balanced and included a pattern only if there was not already a language belonging to the same language stock in the sample that shows the same non-concatenative pattern. For example, our sample includes a case of initial gemination in Marshallese – a non-concatenative phenomenon that can be found in various other Austronesian languages as well (e.g. in Woleaian and Chuukese; Kennedy 2002) that are not in the sample in order to be typologically balanced. In total, our empirical study contains 26 cases of μ-affixation in 24 languages. All of the languages in the sample are listed in (13) together with their classification according to AUTOTYP (Bickel and Nichols ongoing). It can be seen that the 24 languages are distributed over 19 different language stocks.

(13) *Language sample (to be continued)*

	Language	**Stock**	**Area**	**Continent**
I.	Shizuoka Japanese	Japanese	N Coast Asia	N-C Asia
II.	Alabama	Muskogean	E North America	EN America
III.	Zuni	Zuni	Basin and Plains	EN America
IV.	Lardil	Tangkic	N Australia	Australia

(13) *Language sample (continued)*

V.	Gidabal	Pama-Nyungan	S Australia	Australia
VI.	Arbizu Basque	Basque	Europe	W and SW Eurasia
VII.	Slovak	Slavic	Europe	W and SW Eurasia
VIII.	Hausa	Chadic	African Savannah	Africa
IX.	Asante Twi	Kwa	African Savannah	Africa
X.	Luganda	Benue-Congo	E Africa[3]	Africa
XI.	Aymara	Jaqui	Andean	S America
XII.	Quechua	Quechuan	Andean	S America
XIII.	Guajiro	Arawakan	NE South America	S America
XIV.	Southern Sierra Miwok	Yokuts-Utian	California	WN America
XV.	Nootka	Wakashan	Alaska-Oregon	WN America
XVI.	Diegueño	Yuman	California	C America
XVII.	Saanich	Salishan	Alaska-Oregon	WN America
XVIII.	Upriver Halkomelem	Salishan	Alaska-Oregon	WN America
XIX.	Hiaki	Uto-Aztecan	Mesoamerica	C America
XX.	Shoshone	Uto-Aztecan	Mesoamerica	C America
XXI.	Tepecano	Uto-Aztecan	Mesoamerica	C America
XXII.	Tawala	Austronesian	Oceania	NG and Oceania
XXIII.	Keley-i	Austronesian	Oceania	S/SE Asia
XXIV.	Marshallese	Austronesian	Oceania	S/SE Asia

The leading question in generating the sample was where in its base a μ-affix is realized. In Shizuoka Japanese (7), for example, the additional μ was realized in the first syllable, on the first vowel or the first coda consonant. What are possible other linearizations of μ-affixes in the languages of the world? A very locus for a μ-affix is the final vowel, as in Gidabal (14) where the imperative is marked through lengthening of the final vowel.

---

[3] This differs from the classification in the AUTOTYP database that lists Luganda as an South African language. This is clearly a mistake since it contradicts all standard classifications of the language, cf., for example, Clements (1986). Thanks to Larry Hyman for pointing this out to us.

(14) *Gidabal* (Geytenbeek and Geytenbeek 1971, Kenstowicz and Kisseberth 1977)

Base		Imperative
gida	‘to tell’	gid**aː**
ma	‘to put’	m**aː**
jaga	‘to fix’	jag**aː**
gaːda-li-wa	‘keep on chasing’	gaːdaliw**aː**

Another possible locus for the realization of the additional μ is after its base, i.e. through epenthesis at the right edge. An example can be found in Aymara, where final vowel lengthening and insertion of epenthetic *–ja* predictably alternate, the former occurring whenever two lengthening morphemes are expected to cooccur (15-c).

(15) *Aymara* (Beesley 2000, Hardman 2001, Kim 2003)

a.	sara	‘go’	sar**aː**	‘(I) will go’
b.	warmi	‘women’	warm**iː**	‘to be a women’
c.			warmi**jaː**	‘I will be a women’

Similarly, there are patterns where a morphological μ is realized on the first vowel (16-a) or its realization alternates between the first vowel and inserted elements preceding the first base segment (16-b).

(16) *Initial vowel lengthening*

a. *Hiaki* (Molina 1999, Haugen 2005, 2008, Harley and Leyva 2009)

stem		habitual
ivakta	‘embrace’	**iː**vakta
jepsa	‘arrive’	j**eː**psa
wokte	‘put on pants’	w**oː**kte

b. *Upriver Halkomelem* (Elmendorf and Suttles 1986, Galloway 1993, Suttles 2004, Shaw 2004)

non-continuative		continuative
i. Vowel lengthening		
ʔiməç	‘walk’	ʔ**iː**məç
hilt	‘roll sth. over’	h**iː**lt
hɛk’ʷələs	‘remember sth.’	h**ɛː**k’ʷələs

c.

NON-CONTINUATIVE		CONTINUATIVE
ii. CV-reduplication		
q’isət	‘tie sth.’	**q’i**q’əsət
jiq	‘fall (of snow)’	**ji**jəq
mat’əs	‘point, aim’	**ma**mət’əs
iii. CV-epenthesis		
məq’ət	‘swallow sth.’	**h**əmq’ət
wəq’^{w}	‘drift downstream’	**h**əwq’^{w}
jəq^{w}	‘burn’	**hɛ**jq^{w}

And finally, gemination can be located on the first consonant (17-a), the first coda consonant (17-b) or the final onset (17-c) or coda (17-d) consonant.

(17) *Loci for gemination*

a. *Luganda* (Clements 1986, Hyman and Katamba 1990, Kawahara 2007)

STEM		CLASS 5
kubo	‘path’	**k**ːubo
daːla	‘step’	**d**ːaːla
fumu	‘spear’	**f**ːumu

b. *Shoshone* (McLaughlin 1982, Crum and Dayley 1993, Haugen 2008)

STEM		DURATIVE
katɨ́	‘sit’	ka**t**ːɨ
jɨkwí	‘sit.pl’	jɨ**k**ːwi
nemi	‘travel’	ne**m**ːi

c. *Alabama* (Montler and Hardy 1988, Lombardi and McCarthy 1991)

STEM		IMPERFECT
balaː-ka	‘lie down’	bá**l**ːaːka
cokoː-li	‘sit down’	có**k**ːoːli
ilkowat-li	‘move’	ilkó**w**ːatli

d. *Asante Twi* (Paster 2010)

BASE		PAST (+OBJ)
nom	‘to drink’	no**m**ː

The table in (18) now summarizes all the different loci of μ-realization in our sample. The ■ indicates on which or between which segments the length manipulation is visible; if there is more than one ■ present in one line, this simply indicates that a μ-affix is realized through different non-concatenative allomorphs, the choice between them being phonologically predictable. Multiple listings of the same language simply encode the fact that some languages employ different μ-affixes. Note that according to the definition in (12), lexically listed allomorphs of one morpheme are regarded as patterns of their own as well, a pattern that appears in Hiaki.

(18) *Loci of μ-realization: summary*

	Language		#(C)		V		C	…	C	V	(C)#	
1.	Saanich	■		■		■						
2.	Tawala	■			■							
3.	U. Halkomelem	■			■							
4.	Luganda		■									
5.	Marshallese		■									
6.	Keley-i I		■									
7.	Hiaki I				■							
8.	Sh. Japanese				■	■	■					
9.	Tepecano						■					
10.	Keley-i II						■					
11.	Shoshone						■					
12.	Hiaki II						■					
13.	Alabama								■	■		
14.	Arbizu Basque									■		
15.	Gidabal									■		
16.	Zuni									■		
17.	Hausa									■		
18.	Diegeño									■		
19.	Slovak									■		
20.	Nootka									■		
21.	Asante Twi									■	■	
22.	Guajiro									■		■
23.	Quechua									■		■
24.	Lardil									■		■
25.	S. Sierra Miwok									■		■
26.	Aymara									■		■

The most important generalization one can draw from this summary is the simple fact that no μ-affix is realized on or between segments that are more inward in its base than the first coda consonant or last onset consonant. The table in (19) gives an overview over the analyses for all the μ-affixation patterns in terms of the pivot to which the μ-affix prefixes or suffixes. The affixed μ is circled in all depictions to distinguish it from the base μ's.[4]

(19) *Loci of μ-realization: pivots*

Language	**Pivot**	*Examples*		
Saanich	#μ __	μ (μ) s ə q	μ (μ) μ μ w e ʔ q ə s	μ (μ) μ q e q ə n
Tawala	#__μ	(μ)μ μ μ t ał t a w a	(μ) μ μ g e g a e	
Hiaki I	#__μ	(μ)μ μ μ ił v a k t a		
Upriver Halkomelem	#μ __	(μ)μ h i l t	(μ) μ μ q i q ə s ə t	(μ) μ h ə m q ə t
Luganda, Marshallese, Keley-i I	#__μ	(μ) μ μ k u b o		
Sh. Japanese	#μ __	μ (μ) μ h a n d e	μ (μ) μ k a tł ai	μ (μ) μ z o n z ai
Tepecano, Shoshone, Hiaki II, Keley-i II	#__μ	μ (μ) μ j ɨ kł w i		
Alabama	__μ #	μ (μ)μ b a l ał	μ (μ) μ c o bł a	
Gidabal, Zuni, Hausa, Diegeño, Slovak, Nootka	μ __#	μ μ(μ) j a g ał		
Asante Twi	μ __#	μ μ μ(μ) o b i s ał	μ (μ) n o mł	
Quechua, Lardil, S. Miwok, Aymara	μ __#	μ μ μ(μ) j o h k ał	μμ μ μ (μ) h ał j a ŋ k ɨ	

---

[4] Note that morphological pivot affixation for μ's presupposes that stems are equipped with a full prosodic structure at the point where affixation occurs. This follows in a theory assuming cyclic optimization (Trommer 2011, Kiparsky 2011, Bermúdez-Otero 2011, in preparation), a background assumption that is implicit in other subcategorizations approaches referring to prosodic structure as well (e.g. Paster 2005).

Our conclusion of this empirical survey is therefore simple: the locus of affix-μ's is restricted in a way that allows to sufficiently describe all and only the attested cases of μ-affixation cases with the two pivots in (20).

(20) *Pivots for μ-affixation*
first μ
last μ

Any dislocation that follows the morphological placement of a μ-affix is excluded in our theory. As was discussed in section 2.1, this ban on phonological dislocation is in principle independent from the adoption of morphological pivot affixation but a crucial part of the original proposal of morphological pivot affixation in Yu (2007). In addition, we take it for granted that reordering is generally impossible on the segmental as well as on the prosodic level, i.e. there is no true metathesis. All surface effects of metathesis are consequently the result of deletion and insertion instead of true reordering (Zimmermann 2009).[5] From these assumptions it follows that the phonological operations triggered by a μ-affix are always located exactly on/adjacent to the μ/σ-pivot to which the μ affixes.

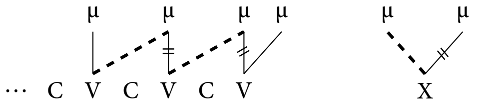
4. Against Phonological μ-Dislocation

In this section, we present several arguments against the alternative approach that linearization of μ-affixes is due to phonological dislocation. We argue that such an approach is too unrestrictive and predicts various instances of μ-affixation that are not attested in the languages of the world (21-a+b) but

---

[5] A final crucial assumption we make is the general absence of μ-shifting, i.e. configurations where a μ newly associates to a segment X while a μ that was underlyingly associated to X deassociates and associates newly to an adjacent segment Y, depicted in (i-a). We take it for granted that such configurations are generally excluded since Gen fails to generate structures violating the principle of 'No vacuous μ-association' (i-b).

(i) *a. μ-Shifting* *b. No vacuous μ-association*



empirically inadequate at the same time since it fails to predict all attested patterns of μ-affixation (21-c+d).

(21) *Arguments against phonological μ-dislocation*
  a. Lack of non-local infixation (subsection 4.1)
  b. Lack of variable infixation (subsection 4.2)
  c. Cases of fixed infixation (subsection 4.3)
  d. Morphologically contrastive μ-affixes (subsection 4.4)

Before we discuss these points in detail, let's briefly recap how a phonological dislocation approach for μ-linearization works. The approaches in Samek-Lodovici (1992), Grimes (2002*a*), Davis and Ueda (2002*b*), and Haugen and Kennard (2008) all assume that a morpheme-specific version of ALIGN demands that a certain morpheme must be realized at the left or right edge but that phonological markedness constraints can be higher-ranked that this preference for being a prefix or suffix. In (22) +(23), two simple exemplifying derivations for phonological dislocation of morphological μ's are given. (22) derives a language as Gidabal (14) with final vowel lengthening and (23) a language as Shoshone (17-b) with gemination of the first coda consonant. The superscript μ in all following candidates is the affix-μ. In the following, we define the morpheme-specific ALIGN as in Haugen and Kennard (2008) meaning that it is violated by every segment between the left/right edge and the segment associated with the μ. In Gidabal, consonant gemination is impossible and thus no available strategy to realize the additional μ – the winning candidate (22-a) consequently realizes the μ on the last vowel under perfect satisfaction of ALIGN(μ_{IMP},R).

(22) *Long vowels in Gidabal*

gida, μ	*Cː	ALIGN(μ_{IMP},R)	*Vː
☞ a. gida^{μ} [gidaː]		*	*
b. gid^{μ}a [gidːa]	*!	**	

In Shoshone on the other hand, vowel lengthening is excluded by high-ranked constraints and since initial geminates are illicit in the language (high-ranked *#Cː), the affix-μ dislocates insides its base to be realized on the first coda consonant. This strategy (23-c) violates ALIGN(μ_{DUR},R) twice since two segments intervene between the *k* and the left edge of the word.

(23) *Geminates in Shoshone*

μ, maka	*#Cː	*Vː	Align(μ_{Dur},L)	*Cː
a. m$^{\mu}$aka [mːaka]	*!			*
b. ma$^{\mu}$ka [maːka]		*!	*	
☞ c. mak$^{\mu}$a [makːa]			**	*

4.1. Lack of Non-Local Infixation

A first overgeneralization problem for phonological disfixation accounts is that they inherently predict non-local μ-infixation. In section 3, we argued that the pivots first and last μ and first σ are sufficient to describe all attested patterns of μ-affixation. A pattern where a segment further right than the second base vowel or further left than the penultimate vowel is affected by the affix-μ cannot be derived with these four pivots but is predicted in a phonological dislocation account. An example for such an unattested pattern of non-local μ-affixation is illustrated in (24) with the pseudo-language Shoshone′ where it is always the second base vowel that is lengthened in the derived form.

(24) *Non-local gemination in unattested Shoshone′*

Base	μ-affixed form
gadali	gada**lː**i
pukalimbu	puka**lː**imbu
sanagumkilte	sana**gː**umkilte

In the tableau (26) it is shown how such a pattern is predicted in a theory assuming that a morphological μ can in principle dislocate inside its base. It is a well-known fact that some prominent position inside a word are especially resistent against phonological changes, amongst them being the first syllable (Beckman 1997, 1998). If now the positional faithfulness constraint (25) preserves the initial syllable from changing the length-value of all segments inside the first syllable and if in the language gemination is the strategy to realize additional μ's, then a non-local pattern as in Shoshone′ emerges.

(25) Dep-AL$_{\#\sigma}$

Assign a violation mark for every new association line between a segment in the first syllable and a μ.

(26) **Shoshone′*

μ, sanagumkil	*Vː	Dep-AL$_{\#\sigma}$	Align(μ,L)	*Cː
a. sa$^{\mu}$nagumkil (saːnagumkil)	*!	*	*	
b. san$^{\mu}$agumkil (sanːagumkil)		*!	**	*
c. sana$^{\mu}$gumkil (sanaːgumkil)	*!		***	
☞ d. sanag$^{\mu}$umkil (sanagːumkil)			****	*

On the basis of our empirical survey of attested μ-affixation cases (cf. (18)), we claim that this is a serious misprediction since such patterns of non-local μ-infixation are unattested in the languages of the world.

4.2. Lack of Variable Infixation

A second general pattern of μ-linearization that is predicted by phonological dislocation approaches but is unattested in the languages of the world is variable μ-infixation, i.e. the realization of an affix-μ in different positions in its base, depending on the shape of the base. An example for a language displaying variable infixation is given in (27). In this pseudo-language Shoshone″, the leftmost consonant that can be geminated without creating an illicit coda cluster is lengthened.

(27) **Shoshone″*

Base	μ-affixed form
petali	pe**tː**ali
mantaku	manta**kː**u
malkuftika	malkufti**kː**a

The tableau in (29) sketches how such a pattern is predicted under a phonological dislocation approach, in fact it is predicted from the very ranking assumed in (23) for the attested gemination in Shoshone: high-ranked *Vː demands infixation of the μ to a coda consonant, under violation of Align(μ,L) (29-I). This ranking predicts that the affix-μ should dislocate further into its base as soon as it is attached to a base where the leftmost base consonant cannot be geminated without creating an illicit CVCC-syllable (29-II+III). We take it for granted that initial geminates are illicit, as in the real Shoshone (23). *#Cː is therefore again taken as undominated and not included in the tableaux.

(28) *Complex (Kager 1999)
Assign a violation mark for every complex syllable margin.

(29) *Derivation of Shoshone″*

			*Complex	*Vː	Align(μ,L)	*Cː
I.	μ, petali					
a.	pe$^{\mu}$ta.li	(petːali)		*!	*	
☞ b.	pet$^{\mu}$a.li	(petːali)			**	*
II.	μ, mantaku					
a.	ma$^{\mu}$n.tan.ku	(maːkantu)		*!	*	
b.	man$^{\mu}$ta.ku	(manːtaku)	*!		**	*
☞ c.	mantak$^{\mu}$u	(mantakːu)			*****	*
III.	μ, malkuftika					
a.	ma$^{\mu}$l.kuf.ti.ka	(maːlkuftika)		*!	*	
b.	mal$^{\mu}$kuf.ti.ka	(malːkuftika)	*!		**	*
c.	mal.kuf$^{\mu}$ti.ka	(malkufːtika)	*!		*****	*
☞ d.	mal.kuf.tik$^{\mu}$a	(malkuftikːa)			********	*

A morphological pivot affixation approach where any dislocation of a morphological μ is impossible, cannot predict such a 'wandering μ' pattern, We argue that this prediction is borne out since no such language as Shoshone″ exists.

A potentially problematic case that apparently exhibits exactly this pattern is morphological gemination in Keley-i under the analysis of Samek-Lodovici (1992). There it is argued that verbs in the language exhibit morphological non-perfect (present and future) gemination of the leftmost consonant that can be geminated, i.e. the first intervocalic consonant in its base. The examples in (30) show cases of gemination in future forms. Note that as in many Philippine languages, Keley-i verbs show inflection indicating whether the clause 'focusses' on a.) the subject, b.) the object or a specific other DPs: c.) 'accessory' focus (=Accs.) indicates focus on instrumentals, d.) 'referent' focus (=Ref) is used for objects 'in a particular region of time or space' (Hohulin and Kenstowicz 1979: 243), and finally there is e.) a 'beneficial' focus.

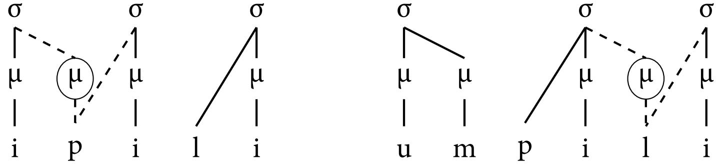
(30) *Non-perfect gemination* (Hohulin and Kenstowicz 1979)

Accs.Focus	Ben.Foc
ʔi-**p**ːili	ʔi-**p**ːili-ʔan
ʔi-**d**ːujag	ʔi-**d**ːujag-an

Subj.Focus	Obj.Focus	Ref.Foc
um-pi**l**ːi	pi**l**ːi-ʔen	pi**l**ːi-ʔan
um-du**j**ːag	du**j**ːag-en	du**j**ːag-an

The examples in (30) indeed suggest the interpretation given in Samek-Lodovici (1992): the leftmost consonant that can be geminated without creating an illicit CVCC-syllable is lengthened in order to distinguish future and present from the past. Whenever the prefix ʔ*i*– precedes a base, the base-initial consonant can be geminated as in ʔ*imːapili* (31-a). However, if a prefix like *um*– ending in a coda consonant is attached to a C-initial base, the base-medial consonant is geminated as in *umpilːi* (31-b) since a structure like **umpːili* is illicit in Keley-i. Such a state of affairs cannot be derived given the assumption of morphological pivot affixation for μ's.

(31) *Gemination in Keley-i*



a. Initial gemination

b. Medial gemination

A closer look at the morphological system, however, reveals the fact that root-initial and root-medial gemination in Keley-i are two different morphophonological processes (Hohulin 1971, Hohulin and Kenstowicz 1979, Archangeli 1987, Lombardi and McCarthy 1991). Root-initial gemination shows up only immediately after the prefix ʔ*i*–, an affix with a highly heterogeneous distribution in Keley-i. In finite non-stative forms, it is restricted to non-perfect (present and future) verb forms of the accessory and beneficial focus, in imperatives it appears only in the non-perfect accessory focus, whereas it extends to perfect (past forms) in the stative paradigm. Root-initial gemination occurs in a subset of these contexts, the non-past uses of ʔ*i*–.

(32) *Non-perfect root-initial gemination* (Hohulin and Kenstowicz 1979)

	Access.Focus	Ben.Foc	
Fut	ʔi-**pː**ili	ʔi-**pː**ili-ʔan	
Past	ʔim-pili	ʔim-pili-ʔan	‘to choose’
Pres	ke-ʔi-**pː**ili	ke-ʔi-**pː**ili-ʔi	
Fut	ʔi-**dː**ujag	ʔi-**dː**ujag-an	
Past	ʔin-dujag	ʔin-dujag-an	‘to pour’
Pres	ke-ʔi-**dː**ujag	ke-ʔi-**dː**ujag-i	

Root-medial gemination has a more straightforward distribution and occurs regularly in (stative and non-stative) non-past subject, object and referent focus forms of roots with a single root-medial consonant.

(33) *Non-perfect root-medial gemination* (Hohulin and Kenstowicz 1979)

	Subj.Focus	Obj.Focus	Ref.Foc	
Fut	um-pi**lː**i	pi**lː**i-ʔen	pi**lː**i-ʔan	
Past	p-imː-ili	p-in-ili	p-in-ili-ʔan	‘to chose’
Pres	ka-ʔum-pi**lː**i	ke-pi**lː**i-ʔa	ke-pi**lː**i-ʔi	
Fut	um-du**jː**ag	du**jː**ag-en	du**jː**ag-an	
Past	d-imː-ujag	d-in-ujag	d-in-ujag-an	‘to pour’
Pres	ka-ʔum-du**jː**ag	ka-du**jː**ag	ka-du**jː**ag-i	

What generates the appearance that root-initial and root-medial gemination are due to the same affix-μ is their partially complementary morphological distribution. In finite non-stative forms, root-initial gemination is found in the accessory and beneficial focus, and root-medial gemination in the subject, object and referent focus. This is summarized in (34) where the lightly shaded background indicates the cells where initial gemination occurs and the darker shaded background those contexts where medial gemination occurs. Since all Keley-i non-stative verb forms belong to one of these five foci, there seems to be only one gemination process. However in verb forms such as stative non-perfect forms the complementarity of distribution breaks down, and both gemination processes show up. An example for such a contexts is given in (35).

(34) *Morphological analysis for Keley-i*

	Focus					
	Access.	**Ben.**	**Sbj.**	**Obj.**	**Ref.**	stative
Pst						ʔi–
Prs	ʔi–	ʔi–	ʔum–	ke–	ke–	ʔi–
Fut	ʔi–	ʔi–	ʔum–			ʔi–

(35) *Initial and medial gemination in Keley-i* (Hohulin and Kenstowicz 1979)

		Pst	**Prs**	**Fut**
bitu	‘to put’	ne-ʔi-bitw-an	ke-ʔi-**bːitː**u-ʔan	me-ʔi-**bːitː**u-ʔan

We can therefore conclude that it is not one μ in Keley-i that is realized in different positions in its base but two morphological μ’s with different subcategorization requirements, one prefixes to the initial base-μ, the other suffixes to the first base-σ. The description in Hohulin and Kenstowicz (1979), the only source for morphological facts of Keley-i, is unfortunately rather sketchy and the lexical entries given in (36) for the two different μ’s are consequently rather preliminary and include simply the different foci where the respective μ occurs listed as disjunctives.

(36) *There are two μ-affixes!*

I. μ/ [__μ ↔ [−pst, Access ∨ Ben ∨ Stat]

II. μ/ [σ__ ↔ [−pst, Sbj ∨ Obj ∨ Ref ∨ Stat]

As a matter of fact, the present tense forms of subject, benificial, and accessory focus reveals a second fatal problem for Samek-Lodovici’s claim that there is a single μ-affix in Keley-i verb forms. As can be seen in the data in (32) and (33), two prefixes are added in these contexts: *ke–ʔi–* and *ka–ʔum–* respectively. Under Samek-Lodovici’s analysis that a single affix-μ is simply realized on the first consonant that can be geminated, we would expect that the glottal stop of the prefixes *ʔi–* and *ʔum–* geminates in such forms resulting in, for example, _*ke–ʔːi–pili_ or _*ka–ʔːum–pili_ instead of *ke–ʔi–pːili* and *ka–ʔum–pːili*.

4.3. Cases of Fixed Infixation: Shizuoka Japanese

In the last two subsections, we discussed overgeneralization problems that a phonological dislocation approach to μ-affixation faces: it predicts patterns of μ-affixation that are unattested in the languages of the world. In the following two subsections, we argue that such an approach is also empirically inadequate in that it is unable to predict patterns that are attested.

The first argument is that a phonological dislocation approach is unable to predict instances of fixed infixation. A case at hand is μ-affixation in Shizuoka that was already discussed in the beginning of section 2.2. Recall that emphatic adjectives in Shizuoka Japanese (Davis and Ueda 2002*a*, 2006) are marked through one of three different phonological processes targeting roughly the edge between the first and the second syllable of the base: gemination of a voiceless intervocalic obstruent (37-a), insertion of a homorganic nasal coda before a voiced obstruent and after a short vowel (37-b), and lengthening of a vowel preceding a nasal coda or a geminate (37-c).

(37) *Emphatic adjectives in Shizuoka Japanese*Davis and Ueda 2006

	Adjective		Emphatic Form			
a.	katai	'hard'	ka**tː**ai			
	osoi	'slow'	o**sː**oi	CV.C̥. . .	⇒	CV.C̥ː. . .
	takai	'high'	ta**kː**ai			
b.	hade	'showy'	ha**n**de			
	ozoi	'terrible'	o**n**zoi	CV.C̬. . .	⇒	CV**N**.C̬. . .
	nagai	'long'	naŋgai			
c.	zonzai	'impolite'	z**oː**nzai			
	supːai	'sour'	s**uː**pːai	CVC.C. . .	⇒	CVːC.C. . .
	okːanai	'scary'	**oː**kːanai			

As was already mentioned in section 2.2, the choice between these three allomorphs to realize an additional μ is phonologically predictable. Whenever the first syllable of the base is closed by a coda consonant, vowel lengthening occurs (37-c). If the first base-syllable is open, the nature of the second onset consonant determines the choice of the allomorph: if the second syllable starts with a sonorant, nasal-insertion takes place (37-b) and if its starts with an obstruent, gemination surfaces (37-a). Before we turn to the linearization of the μ-affix, we briefly illustrate how this choice of allomorphs follows in (a simplified version of) the OT-analysis given in Davis and Ueda (2002*a*). σ-Cond in the following tableaux stands for a constraint stratum in which several standard markedness constraints on syllable structure are ordered; especially the demand that onsets are not moraic (*$_{\sigma}$[C$_{\mu}$), that voiceless obstruents should not be geminates (*C̬ː), and the ban against complex codas (*CC]$_{\sigma}$).

(38) *Choice between the allomorphs in Shizuoka Japanese*

a. *Gemination*

μ, katai	σ-Cond	*Vː	Dep n	*Cː
☞ a. kat$_\mu$ai (katːai)				*
b. ka n$_\mu$ tai (kantai)			*!	
c. ka$_\mu$tai (kaːtai)		*!		

b. *Nasal insertion*

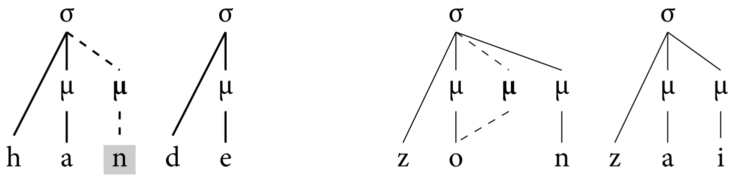
μ, hade	σ-Cond	*Vː	Dep n	*Cː
a. had$_\mu$e (hadːe)	*C̬ː	*		*
☞ b. ha n$_\mu$ de (hande)			*	
c. ha$_\mu$de (haːde)		*!		

c. *Vowel lengthening*

μ, zonzai	σ-Cond	*Vː	Dep n	*Cː
a. zonz$_\mu$ai (zon.zːai)	$*_\sigma[C_\mu$!	*		*
b. zon n $_\mu$zai (zonn.zai)	$*CC]_\sigma$!		*	
☞ c. zo$_\mu$nzai (zoːn.zai)		*		

Now we can turn to question of where this additional μ is realized. Crucially, as shown in (39), the additional mora expressing emphasis appears after the first mora of the base. This becomes especially obvious from the nasal insertion case (39-a) where the vocalic stem mora clearly intervenes between the additional affix mora and the left edge of the base.[6]

(39) *μ-affixation in Shizuoka Japanese (to be continued)*



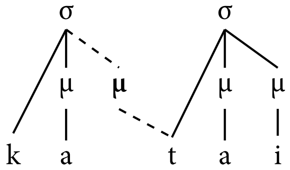
a. *Nasal insertion* b. *Vowel lengthening*

---

[6] The emphatic cannot be formed by infixation before the final mora of the base since the second syllable sometimes contains one (*hade*) and sometimes 2 moras (*nagai*).

(39) *μ-affixation in Shizuoka Japanese (continued)*

c. *Gemination*



An analysis for these facts under a morphological pivot affixation approach is straightforward: the μ is assumed to suffix to the first μ of its base.

(40) *μ-affix in Shizuoka Japanese*

μ ↔ $_{\text{Base}}[$ μ ___

And how are these facts derived under the alternative assumption of phonological dislocation? Basically this question reduces to the placement of the constraint Align(μ,L) demanding a prefix position of the affix-μ. It can easily be demonstrated that the phonological dislocation theories run into a ranking paradox to predict the correct placement of the affix-μ. Let's start with contexts where a nasal is inserted. We know that the affix-μ must dislocate in these cases since it is realized as coda-μ following the first base-μ. Consequently, Align(μ,L) must be ranked below *Vː as can be seen in the comparison between I. and II. in (41).

(41) *Nasal insertion for CVO̮V*

a. *I. Wrong ranking: vowel lengthening is predicted*

μ-hade	σ-Cond	Align(μ,L)	*Vː	Dep n	*Cː
a. had$^{\mu}$e (hadːe)	*C̮ː!	**			*
☞ b. ha n$^{\mu}$ de (hande)		**!		*	
☛ c. ha$^{\mu}$de (haːde)		*	*		

b. *II. Correct ranking: nasal insertion is predicted*

μ-hade	σ-Cond	*Vː	Align(μ,L)	Dep n	*Cː
a. had$^{\mu}$e (hadːe)	*C̮ː!		**		*
☞ b. ha n$^{\mu}$ de (hande)			**	*	
c. ha$^{\mu}$de (haːde)		*!			

However, if we turn to another context, this ranking of *Vː over Lin-μ makes a fatal misprediction. For bases starting with a closed syllable, Lin-μ must be ranked above *Vː to block gemination beyond the first σ (42).

(42) *Vowel lengthening for CVN.OV*

a. *I. Correct ranking: vowel lengthening is predicted*

	μ-kata	σ-Cond	Align(μ,L)	*Vː	Depn	*Cː
	a. onz$^{\mu}$okutai (on.zːokutai)	*$_{\sigma}$[C$_{\mu}$!	**			*
	a’. onzok$^{\mu}$utai (on.zokːutai)		*!**			*
	b. on n$^{\mu}$ zai (onn.zokutai)	*CC]$_{\sigma}$!			*	
☞	c. o$^{\mu}$nzokutai (oːn.zokutai)			*		

b. *II. Wrong ranking: non-local gemination is predicted*

	μ-kata	σ-Cond	*Vː	Align(μ,L)	Depn	*Cː
	a. onz$^{\mu}$okutai (on.zːokutai)	*$_{\sigma}$[C$_{\mu}$!		**		*
☛	a’. onzok$^{\mu}$utai (on.zokːutai)			****		*
	b. on n$^{\mu}$ zai (onn.zokutai)	*CC]$_{\sigma}$!		*	**	
☞	c. o$^{\mu}$nzokutai (oːn.zokutai)		*			

In their analysis of Japanese, Davis and Ueda (2002*a*) employ a non-standard alignment constraint which fixes the position of the affix-μ to the initial syllable of a base; it is violated if the μ is realized beyond the first syllable. The constraint has a similar effect as fixing the position of the affix-μ with respect to the μ's of the base word. Hence in effect, this constraint amounts to a subcategorization requirement.

(43) Align-L(μ$_{e}$,Wd)
Align the emphatic mora with the beginning (left edge) of the word.

4.4. Morphologically Contrastive μ-Affixes

Under a phonological dislocation approach, the realization of an affix-μ follows from the general phonology of the language, for example, from the preference for geminates or long vowels. Under the assumption that there are no morpheme specific mechanisms (indexed constraints Pater 2009, cophonologies Inkelas and Zoll 2005), it follows that any μ that is affixed to a certain base will be realized in exactly the same base. In this section we argue that this is a serious drawback for a phonological dislocation approach since there are indeed languages where different μ-affixes result in different outputs (Guerssel and Lowenstamm 1990, Lowenstamm 2003). A case in point are the first two Binyanim of Classical Arabic.[7] Whereas Binyan II geminates the first non-initial consonant, Binyan III lengthens the first vowel.

(44) *Binyanim in Classical Arabic* (McCarthy 1979, McCarthy and Prince 1990)

	‘write’	‘do’
Binyan I	katab	faʔal
Binyan II	ka**tː**ab	faʔ**ː**al
Binyan III	k**aː**tab	f**aː**ʔal

Under the assumption that all Binyanim are derived from fully vocalized Binyan I forms (Ussishkin 2003, 2005), both Binyan II and III could be captured by μ-affixation (Davis and Ueda 2006). The crucial problem which emerges is why the Binyan II-μ attaches to the consonant and the Binyan III-μ to the vowel. We will call this problem ‘Lowenstamm’s dilemma’, who was the first to clearly notice that the two Binyanim undergo roughly the same type of prosodic augmentation that differs minimally by the attachment of affixal prosody to segmental material which is crucially dependent on the morphological category involved (Guerssel and Lowenstamm 1990, Lowenstamm 2003). There are two standard solutions to Lowenstamm’s dilemma: Lowenstamm himself associates the lengthening prosody (which for him is not a μ, but a skeletal CV unit) by morphological stipulation, a move which has become canonical for analyses of templatic morphology in government phonology (e.g. Bendjaballah 2001, Bendjaballah and Haiden 2003, Rucart 2001, 2006, Lahrouchi 2009, Arbaoui 2010).

Davis and Ueda (2002*a*) suggest to capture the fact that the Binyan III-μ associates to a V, not to a consonant (the default option for affixal μ’s according

---

[7] We have excluded templatic morphology from our sample to restrict it to clear-cut cases of μ-affixation. However Semitic verbal root-and-pattern morphology is one of the the areas where an affixational analysis is highly plausible. See Ussishkin (1999, 2003, 2005), Trommer and Zimmermann (2011) for affix-based analyses of vocalic patterns in Modern Hebrew, Trommer (2005) for a mora-affixation analysis of gemination in Amharic, and Davis and Ueda (2006) on a similar proposal for Classical Arabic.

to Davis and Ueda's (2002*a*) assumptions) by positing "that the input mora indicating Form 3 would be subscripted as a vocalic mora" (p.17). Both approaches are conceptually costly since they add technical devices (phonological subscripting and morphological association) to phonological theory which are not motivated outside of nonconcatenative morphology. Under the assumption that μ-affixation is morphological pivot-affixation, however, Lowenstamm's dilemma is completely unproblematic and rather an expected pattern: the affix-μ's in Binyan II and III are simply pronounced in different positions in their base because they are affix to different pivots.

(45) *Two μ-affixes in Classical Arabic*
  a. Binyan II ↔ μ / [μ __
  b. Binyan III ↔ μ / [__ μ

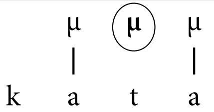
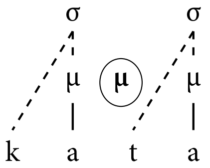
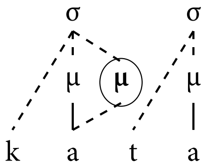
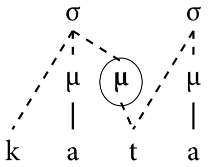
The first μ-affix (45-a) suffixes to the first μ of its base. In this position, it can in principle associate to three segments without crossing association lines: to the first or second base vowel or the coda consonant of the first syllable. We assume that gemination is the preferred option to realize a morphological μ in Classical Arabic, ensured through ranking *Vː higher than *Cː. In the tableau (47), it is shown how this ranking then ensures that the suffixed μ will always result in gemination. We included whole autosegmental representations to illustrate the rather subtle differences between the two μ-affixes and we also added the constraints (46-b+c) that ensure proper realization of an affix-μ in the first place.

(46) a. $*\times$ Assign a violation mark for every pair of crossing association lines.

  b. $\begin{matrix}\sigma\\ \uparrow\\ \mu\end{matrix}$ Assign a violation mark for every μ that is not dominated by a σ.

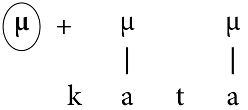
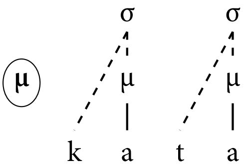
  c. $\begin{matrix}\mu\\ \downarrow\\ \bullet\end{matrix}$ Assign a violation mark for every μ that does not dominate any segment.

(47) *Binyan II: Gemination*

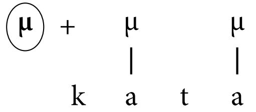
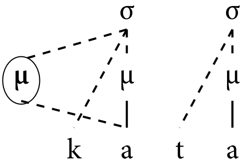
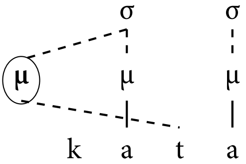
	*×	σ ↑ μ	μ ↓ ●	*Vː	*Cː
a. 		*!	*		
b. 				*!	
☞ c. 					*

The second μ-affix, on the other hand, is taken to be a prefix to the first μ (46-b). Although vowel-lengthening is the dispreferred option to realize an additional μ in Classical Arabic, this μ-prefix will always result in vowel-lengthening simply because there is no consonant the μ could associate to without creating a crossing association line configuration (given that initial geminates are generally illicit).

(48) *Binyan III: Vowel Lengthening (to be continued)*

	*×	σ ↑ μ	μ ↓ ●	*Vː	*Cː
a. 		*!	*		

(48) *Binyan III: Vowel Lengthening (continued)*

	*×	σ ↑ μ	μ ↓ •	*Vː	*Cː
☞ b. 				*	
c. 	*!				*

5. Conclusion

We have argued that μ-affixation is pivot-affixation and explicitly reject any phonological dislocation (or 'phonological readjustment') to account for infixation. Affixation of a morphological μ is therefore a morphological operation that is independent and derivationally prior to phonological optimization. An alternative account assuming phonological dislocation is empirically inadequate and too unrestrictive at the same time. It predicts unattested instances of non-local infixation and variable infixation but fails on the other hand to predict attested instances of fixed infixation or moraic distinctiveness.

References

Álvarez, José (2005): Vocalic Mora Augmentation in the Morphology of Guajiro/Wayuunaiki. In: *Proceedings from the Eighth Workshop on American Indigenous Languages (2005)*. Santa Barbara Working Papers in Linguistics, vol. 16.

Arbaoui, Nora (2010): Les dix formes de l'arabe classique à l'interface phonologie-syntaxe- pour une déconstruction du gabarit. PhD thesis, Université Paris 7.

Archangeli, Diana (1987): Consonant Assimilation in Keley-i. In: *Coyote Papers 6*. University of Arizona.

Arregi, Karlos and Andrew Nevins (2012): *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Springer, Dordrecht.

Beckman, Jill (1998): Positional Faithfulness. PhD thesis, University of Massachusetts at Amherst.

Beckman, Jill N. (1997): 'Positional Faithfulness, Positional Neutralisation and Shona Vowel Harmony', *Phonology* 14, 1–46.

Beesley, Kenneth R. (2000): 'A Note on Phonologically Conditioned Selection of Verbalization Suffixes in Aymara', Technical Report, Xerox Research Centre Europe, July.

Bendjaballah, Sabrina (2001): 'The Negative Preterite in Kabyle Berber', *Folia Linguistica* 34, 185–223.

Bendjaballah, Sabrina and Martin Haiden (2003): 'Templatic Architecture', *Recherches linguistiques de Vincennes* 32(157-168).

Bermúdez-Otero, Ricardo (2011): Cyclicity. *In:* M. van Oostendorp, C. J. Ewen, E. Hume and K. Rice, eds, *The Blackwell Companion to Phonology*. Wiley Blackwell, Malden MA, chapter 85.

Bermúdez-Otero, Ricardo (in preparation): *Stratal Optimality Theory*. Oxford University Press, Oxford.

Bickel, Balthasar and Johanna Nichols (ongoing): 'AUTOTYP, Developing an International Network of Typological Linguistic Databases', http://www.spw.uzh.ch/autotyp/.

Bloomfield, Leonard (1933): *Language*. New York: Holt, Rinehart & Winston.

Bye, Patrick and Peter Svenonius (2012): Non-Concatenative Morphology as Epiphenomenon. *In:* J. Trommer, ed., *The Morphology and Phonology of Exponence: The State of the Art*. Oxford University Press, Oxford, pp. 426–495.

Clements, George (1986): Compensatory Lengthening and Consonant Gemination in LuGanda. *In:* L. Wetzels and E. Sezer, eds, *Studies in Compensatory Lengthening*. Foris, Dordrecht, pp. 37–77.

Crum, Beverly and Jon Dayley (1993): *Western Shoshoni grammar*. Boise State University, Boise.

Davis, Stuart and Isao Ueda (2002*a*): 'Mora Augmentation Processes in Japanese', *Journal of Japanese Linguistics* 18, 1–23.

Davis, Stuart and Isao Ueda (2002*b*): 'The Typology of Mora Augmentation', Proceedings of LP 2005.

Davis, Stuart and Isao Ueda (2006): 'Prosodic vs. Morphological Mora Augmentation', *Lexicon Forum* 2, 121–143.

Elmendorf, William and Wayne Suttles (1986): 'Pattern and Change in Halkomelem Salish', *Anthropological Linguistics* 2, 1–35.

Fitzpatrick, Justin (2004): A Concatenative Theory of Possible Affix Types. *In:* I. A. Salanova, ed., *Papers from EVELIN I*. MIT Working Papers in Linguistics.

Galloway, Brent (1993): *A Grammar of Upriver Halkomelem*. University of California Press, Berkeley.

Geytenbeek, Brian and H. Geytenbeek (1971): *Gidabal Grammar and Dictionary*. Australian Institute of Aboriginal Studies, Canberra.

Greenberg, Joseph H. (1963): Some Universals of Grammar with Special Reference to The Order of Meaningful Elements. *In:* J. H. Greenberg, ed., *Universals of Language*. Cambridge MA: MIT Press, pp. 58–90.

Grimes, Stephen (2002*a*): 'Mora Augmentation in the Alabama Imperfective: An Optimality Theoretic Perspective', Ms. Indiana University, online available at http://pweb.ldc.upenn.edu/~sgrimes/papers/alabama.pdf.

Grimes, Stephen (2002*b*): Morphological Gemination and Root Augmentation in Three Muskogean Languages. Ms., Linguistic Data Consortium, UPenn.

Guerssel, Mohand and Jean Lowenstamm (1990): The Derivational Morphology of the Classical Arabic Verbal System. Université du Québec à Montréal and Université Paris 7.

Halle, Morris (2003): Infixation versus Onset Metathesis in Tagalog, Chamorro and Toba Batak. In: *Ken Hale: A Life in Language*. MIT Press, Cambridge, pp. 153–168.

Hardman, Martha J. (2001): *Aymara*. LINCOM.

Harley, Heidi and Maria Flores Leyva (2009): 'Form and Meaning in Hiaki (Yaqui) Verbal Reduplication', *International Journal of American Linguistics* 75, 235–74.

Haugen, Jason (2005): Reduplicative Allomorphy and Language Prehistory in Uto-Aztecan. *In:* B. Hurch, ed., *Studies on reduplication*. de Gruyter, Berlin, pp. 315–350.

Haugen, Jason (2008): *Morphology at the Interfaces. Reduplication and Noun Incorporation in Uto-Aztecan*. John Benjamin.

Haugen, Jason and Cathy Hicks Kennard (2008): 'Morphological Moras and Morphological Doubling Theory', LSA Annual Meeting, San Francisco.

Hayes, Bruce (1989): 'Compensatory Lengthening in Moraic Phonology', *Linguistic Inquiry* 20, 253–306.

Hohulin, Lou and Michael Kenstowicz (1979): 'Keley-i Phonology and Morphophonemics', *South-East Asia Linguistic Studies* 4, 241–254.

Hohulin, R. M. (1971): Cohesive Organisation in Keley-I Kallahan. *In:* R. M. Hohulin and L. Hoholin, eds, *Papers in Philippine Linguistics*. Vol. 4, Australian National University, pp. 1–17.

Horwood, Graham (2002): Precedence Faithfulness Governs Morpheme Position. *In:* L. Mikkelsen and C. Potts, eds, *Proceedings of WCCFL 21*. Cascadilla Press, Somerville, MA, pp. 166–179.

Hyman, Larry and Francis Katamba (1990): 'Final Vowel Shortening in Luganda', *Studies in African Linguistics* 21, 1–59.

Inkelas, Sharon and Cheryl Zoll (2005): *Reduplication: Doubling in Morphology*. Cambridge: Cambridge University Press.

Kager, René (1999): *Optimality Theory*. Cambridge University Press, Cambridge.

Kawahara, Shigeto (2007): Sonorancy and Geminacy. *In:* L. Bateman, M. O'Keefe and A. Werle, eds, *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*. GLSA, pp. 145–186.

Kennedy, Robert (2002): Stress and Allomorphy in Woleaian Reduplication. In: *Proceedings of the Texas Linguistic Society*. The University of Texas at Austin, Austin.

Kenstowicz, Michael and Charles Kisseberth (1977): *Topics in Phonological Theory*. Academic Press, New York.

Kim, Yuni (2003): 'Vowel Elision and the Morphophonology of Dominance in Aymara', Manuscript, UC Berkeley.

Kiparsky, Paul (2011): Compensatory Lengthening. *In:* C. Cairns and E. Raimy, eds, *Handbook on the Syllable*. Brill, Leiden, pp. 33–69.

Klein, Thomas B. (2005): 'Infixation and Segmental Constraint Effects: UM and IN in Tagalog, Chamorro, and Toba Batak', *Lingua* 115(7), 959–995.

Kurisu, Kazutaka (2001): The Phonology of Morpheme Realization. PhD thesis, University of California at Santa Cruz. Available as ROA 490-0102.

Lahrouchi, Mohamed (2009): 'A Templatic Approach to Gemination in the Imperfective Stem of Tashlhiyt Berber', *Studies in African Linguistics* 37(1), 21–60.

Leonard, Janet and Claire Kelly Turner (2010): Predicting the Shape of Senćoten Imperfectives. In: *A Festschrift for Thomas M. Hess on the Occasion of his 70th Birthday*. WA: Whatcom Museum Publication, pp. 82–112.

Lombardi, Linda and John J. McCarthy (1991): 'Prosodic Circumscription in Choctaw Morphology', *Phonology* 8, 37–71.

Lowenstamm, Jean (2003): 'À propos de gabarits', *Recherches linguistiques de Vincennes* 32, 7–30.

McCarthy, J. (1979): Formal Problems in Semitic Phonology and Morphology. PhD thesis, Massachusetts Institute of Technology.

McCarthy, John (2000): Faithfulness and Prosodic Circumscription. *In:* J. Dekkers, F. van der Leeuw and J. van de Weijer, eds, *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford University Press, pp. 151–189.

McCarthy, John and Alan Prince (1993): 'Generalized Alignment', *Yearbook of Morphology* pp. 79–153.

McCarthy, John J. and Alan Prince (1990): 'Foot and Word in Prosodic Morphology: The Arabic Broken Plural', *Natural Language and Linguistic Theory* 8, 209–283.

McLaughlin, John (1982): Two or Three (or Four) Points about Adverbs and Aspect in Central Numic (Uto-Aztecan). In: *Kansas Working Papers in Linguistics*. University of Kansas. Linguistics Graduate Student Association.

Molina, Felipe (1999): *Hippocrene Standard Dictionary Yoeme-English, English-Yoeme*. Hippocrene Books.

Montler, Timothy (1986): *An Outline of the Morphology and Phonology of Saanich, North Straits Salish*. University of Montana.

Montler, Timothy and Heather Hardy (1988): 'Imperfective Gemination in Alabama', *International Journal of American Linguistics* 54, 399–415.

Moravcsik, Edith A. (1977): *On Rules of Infixing*. Indiana University Linguistics Club, Bloomington.

Newman, Stanley (1965): *Zuni Grammar*. University of New Mexico Publications.

Orgun, Cemir Orhan and Ronald L. Sprouse (1999): 'From MPARSE to CONTROL', *Phonology* **16**, 191–224.

Paster, Mary (2005): Subcategorization vs. Output Optimization in Syllable-Counting Allomorphy. In: *Proceedings of the 24th West Coast Conference on Formal Linguistics*. pp. 326–333.

Paster, Mary (2010): 'The Verbal Morphology and Phonology of Asante Twi', *Studies in African Linguistics* **39**, 77–120.

Pater, Joe (2009): Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution. *In:* S. Parker, ed., *Phonological Argumentation: Essays on Evidence and Motivation*. Equinox, London, pp. 123–154.

Prince, Alan and Paul Smolensky (1993/2002): 'Optimality Theory: Constraint Interaction in Generative Grammar', [first circulated as Prince & Smolensky (1993) Technical reports of the Rutgers University Center of Cognitive Science], ROA 537-0802.

Rucart, Pierre (2001): The Vocalism of Strong Verbs in Afar. *In:* A. Simpson, ed., *Berkeley Linguistics Society: Proceedings of the Annual Meeting - Special Session on Afroasiatic Languages, University of California, Berkeley, 22-25 mars 2001*. Vol. 27, Berkeley Linguistics Society, pp. 99–108.

Rucart, Pierre (2006): Morphologie gabaritique et interface phonosyntaxique. PhD thesis, Paris 7, Sciences du Langage.

Saba Kirchner, Jesse (2007): 'The Phonology of Lexical Underspecification', Ms. University of California, online available at http://jessesabakirchner.com/docs/2007-phonology-of-lexical-underspecification.pdf.

Saba Kirchner, Jesse (2010): Minimal Reduplication. PhD thesis, University of California at Santa Cruz. Available as ROA 1078-0610.

Samek-Lodovici, Vieri (1992): A Unified Analysis of Crosslinguistic Morphological Gemination. In: *Proceedings of CONSOLE 1*. Utrecht, pp. 265–283.

Shaw, Patricia (2004): Reduplicant Order and Identity: Never Trust a Salish CVC Either? *In:* D. B. Gerdts and L. Matthewson, eds, *Studies in Salish linguistics in Honor of M. Dale Kinkade*. University of Montana Working Papers in Linguistics 17, Missoula, pp. 329–354.

Stemberger, Joseph and Barbara Bernhardt (1998): 'Contiguity, Metathesis and Infixation', *Proceedings of the WCCFL 17* pp. 610–624.

Stonham, John (2007): 'Metathesis as Prosodic Repair', *Studies of Phonetics, Phonology and Morphology*, 13, 3–24.

Suttles, Wayne (2004): *Musqueam Reference Grammar*. UBC Press, Vancouver.

Trommer, Jochen (2005): Amharic Roots and Patterns. Talk given at "Flexionstheorie im Osten", Leipzig, July 2005; slides available under http://www.uni-leipzig.de/~jtrommer/rup.pdf.

Trommer, Jochen (2011): 'Phonological Aspects of Western Nilotic Mutation Morphology', Habil. University of Leipzig.

Trommer, Jochen and Eva Zimmermann (2011): 'Overwriting as Optimization', *NLLT* **29**(2), 561–580.

Turner, Claire Kelly (2007): 'The Sencoten Resultive Construction', *Northwest Journal of Linguistics* **1**, 1–92.

Ussishkin, Adam (1999): 'The Inadequacy of the Consonantal Root: Modern Hebrew Denominal Verbs and Output-Output Correspondence', *Phonology* **16**(3), 401–442.

Ussishkin, Adam (2003): Templatic Effects as Fixed Prosody: The Verbal System in Semitic. *In:* J. Lecarme, J. Lowenstamm and U. Shlonsky, eds, *Research in Afroasiatic Grammar III*. Amsterdam: John Benjamins, pp. 511–530.

Ussishkin, Adam (2005): 'A Fixed Prosodic Theory of Nonconcatenative Templatic Morphology', *Natural Language and Linguistic Theory* **23**(1), 169–218.

Wolf, Matthew (2008): Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model. PhD thesis, University of Massachusetts Amherst.

Wolf, Matthew (2013): Lexical Insertion Occurs in the Phonological Component. *In:* B. Tranel, ed., *Understanding Allomorphy: Perspectives from Optimality Theory*. London: Equinox. To appear.

Yoon, Junghyoe (2008): 'Mora Augmentation in Korean Noun Subcompounding', IULC Working Papers online.

Yu, Alan C. L. (2002): Understanding Infixes as Infixes. Handout of a talk given at NAPhC 2. ROA-523-0602/523-0602.

Yu, Alan C. L. (2007): *A Natural History of Infixation*. Oxford: Oxford University Press.

Zimmermann, Eva (2009): Metathesis without Reordering. Master's thesis, University of Leipzig.

Zoll, Cheryl (1996): Parsing below the Segment in a Constraint-Based Framework. PhD thesis, UC Berkeley.

Rule Interaction in Kleverlandish Diminutive Formation

Barbara Stiebels[*]

Abstract

This paper discusses the various patterns of diminutive formation in Kleverlandish, a German variety of Low Franconian. Kleverlandish displays four diminutive allomorphs (default: /-kə/; further variants: /-skə/, /-əkə/, /-tjə/), whose distribution is determined by the final segment of the base. Whereas noun stems with a final coronal sonorant exhibit a complex pattern of allomorphic distribution, those with a final dorsal consonant evoke a complex interaction of rules in diminutives. I will argue that the four allomorphs should be analyzed as separate lexical items and that they are not related by generalized rules of /s/- or schwa epenthesis.

1. Introduction

The Kleverlandish dialect, which is a variety of Low Franconian in the transition zone to Low German and High/Standard German, displays an interesting allomorphy of diminutive formation.[1] Apart from the default form /-kə/, the following allomorphs occur: /-skə/, /-tjə/ and /-əkə/. Their distribution is mainly determined by the final stem segment, which I will demonstrate below.

The transitional character of Kleverlandish manifests itself in at least two properties of the diminutives: first, unlike the diminutives of Standard Dutch, which is the major representative of Low Franconian, those of Kleverlandish trigger umlaut, thus following the pattern of the High German diminutive

---

[*] I would like to thank my parents Roswitha and Wilhelm Stiebels for their help with the data and Fabian Heck, Nina Topintzi and Marzena Żygis for discussion. For the compilation of possible forms I have also consulted two dictionaries (Schönberner 1998 and VABW 1995).

[1] In the following I am referring to the dialect variety spoken in Kleve (Cleves), Germany. Note that this dialect does not have a standardized orthography. Long vowels are indicated by ⟨VV⟩, short vowels in closed syllables by complex orthographic codas (⟨VCC⟩).

Rule Interaction in Grammar, 163–176
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

-*chen*. Umlaut applies without exceptions and leads to a fronting of back stem vowels.[2]

(1) *Umlaut patterns*[3]

	Base	Diminutive	
/u/ ~ /y/	*Bluus*	*Blüüs-ke*	‘shirt’
/ʊ/ ~ /ʏ/	*Blumm*	*Blümm-ke*	‘flower’
/o/ ~ /ø/	*Poot*	*Pööt-je*	‘paw’
/ɔ/ ~ /œ/	*Flokk*	*Flökk-ske*	‘(snow) flake’
	Oos [ɔ:]	*Öös-ke* [œ:]	‘carrion, bitch’
/a/ ~ /æ/	*Danz*	*Dänz-ke*	‘dance’
/aʊ/ ~ /ɔʏ/	*Frau*	*Fräu-ke*	‘woman’

Secondly, in contrast to High German, Kleverlandish strongly avoids zero-marked plural forms; the plural of diminutives is marked by /-s/ throughout all variants of the diminutive (e.g. *Blümm-ke-s* ‘little flowers’ – cf. High German *Blüm-chen-Ø*).

Some diminutives exhibit an unexpected lengthening or shortening of the stem vowel; the respective cases of stem vowel lengthening seem to be paralleled in the plural of the base forms:[4]

(2) *Lengthening of the stem vowel*

	Base		Diminutive	Plural	
/ɔ/	*Hoff*	[œ:]	*Hööf-ke*	*Hööf*	‘yard’
/ʏ/	*Hüss*	[y:]	*Hüüs-ke*	*Hüüs*	‘house’
/ʊ/	*Muss*	[y:]	*Müüs-ke*	*Müüs*	‘mouse’

The shortening of stem vowels seems to be restricted to the diminutive:

---

[2] Note that there is also an interesting, though lexicalized case of multiple umlaut: the diminutive of *Auto* [aʊto] ‘car’ is *Äutö-ke* [ˈɔʏtœkə]/[ˈɔʏtəkə].

[3] Kleverlandish does not exhibit a strict correlation between tenseness/laxness of vowels and their respective length; one can find short tensed vowels such as /e/ (e.g., *Lecht* ‘light’) and long lax vowels such as /ɔ:/ or /œ:/.

[4] See Booij (1995: 72) for similar cases in Standard Dutch.

(3) *Shortening of the stem vowel*

	Base		Diminutive	Plural	
/u:/	*Kuuk*	[ʏ]	*Kükk-ske*	*Küük*	'cake'
	Fuut		*Fütt-je*	*Füüt*	'foot'
/o:/	*Hoop*	[œ]	*Höpp-ke*	*Hööp*	'heap'
/ɔ:/	*Schoop*	[œ]	*Schöpp-ke*	*Schööp*	'sheep'
/i:/	*Brief*	[ɪ]	*Briff-ke*	*Briev-e*	'letter'
	Piep		*Pipp-ke*	*Piep-e*	'pipe'

But since lengthening/shortening does not occur systematically and seems to be a lexicalized property of certain nouns, I do not pursue this issue further.

I will first discuss the default form of the diminutive and then the other allomorphs, which are suffixed to stems that end in coronal or dorsal consonants. Note that I do not intend to provide a complete picture of the phonological rules in Kleverlandish that interact with diminutive formation; I will confine myself to the aspects that are relevant for diminutives.

2. The Default Form /-kə/

The default form of the diminutive (/-kə/) is used after stems that end in (i) vowels or glides, (ii) labial consonants or (iii) coronal fricatives (/s, z/):[5]

(4) *Default form of the diminutive*

Stem-final segment	Base	Diminutive	
/V/	*Ei*	*Ei-ke*	'egg'
/j/	*Floj*	*Flöj-ke*	'flea'
/ʀ/	*Fenger* [ɐ]	*Fenger-ke*	'finger'
/m/	*Flamm*	*Flämm-ke*	'flame'
/p/	*Dropp*	*Dröpp-ke*	'drop'
/v/	*Duuv*	*Düüv-ke*	'pigeon'
/f/	*Hoff*	*Hööf-ke*	'yard'
/s/	*Foss*	*Föss-ke*	'fox'
/z/	*Nöös*	*Nöös-ke*	'nose'

---

[5] The coronal fricative /ʃ/ is restricted to word-initial position. The French loanword *blamaasch* 'disgrace' is the only exception to this generalization. With sufficient contextual support my informants accepted the diminutive *Blamääsch-ke*.

Due to final devoicing (FDV), which applies to syllable-final obstruents, the voicing contrast in the final segment of the stem is neutralized in the base form as well as in the diminutive. FDV may be stated as follows:

(5) (Wiese 1996: 201)
FDV (Final devoicing): $[+\text{obstruent}] \rightarrow [-\text{voice}]\ /\ \_\_\ ]_{\sigma}$

Note also that the uvular rhotic /ʀ/ is vocalized in post-vocalic position (i.e., realized as [ɐ]).[6] Therefore, stems that end in /ʀ/ seem to pattern like vowel-final stems and not like stems with a final dorsal consonant (see section 4). I will adopt Wiese's simplified /ʀ/-vocalization rule, which is based on an underspecification analysis of the respective phonemes; the details of /ʀ/-vocalization are not relevant for the discussion of Kleverlandish diminutives. (6) applies to /ʀ/ in rhyme position.

(6) (Wiese 1996: 256)
RV (/ʀ/-vocalization):

$$\begin{bmatrix} +\text{continuant} \\ -\text{obstruent} \end{bmatrix} \rightarrow [\text{-consonantal}]\ /\ \overset{\overset{\text{Rhyme}}{|}}{\_\_}$$

3. Stems with a Final Coronal Consonant

Apart from stems that end in coronal fricatives other coronal-final stems do not take the default form of the diminutive. The least restricted allomorph is /-tjə/, which is chosen after stems with final coronal non-continuants:

(7)	/l/	*Deckel*	*Deckel-tje*	'lid'
		Läpel	*Läpel-tje*	'spoon'
	/n/	*Boon*	*Böön-tje*	'bean'
		Kaploon	*Kaplöön-tje*	'chaplain'
	/d/	*Fodd* [fɔt]	*Född-je* [fœt.jə]	'rags'
		Hond	*Hönd-je*	'dog'
	/t/	*Katt*	*Kätt-je*	'cat'
		Pott	*Pött-je*	'pot'

As the examples illustrate, there is cluster simplification in the sequence of coronal (oral) stops, thus leading to /t/-deletion, which is part of a general

---

[6] If /ʀ/ is followed by /t/, the rhotic is realized as a fricative.

degemination rule (see Booij 1995: 68f. for Standard Dutch and Wiese 1996: 229ff. for Standard German). Stems with complex final clusters consisting of a fricative and /t/ show a peculiar behavior, to which I will return in section 5.

A subset of the stems with a final coronal sonorant takes another diminutive allomorph, namely /-əkə/. This diminutive allomorph is subject to the prosodic context requirement that the stem vowel be short; polysyllabic stems must show final stress in addition. Monosyllabic stems with this pattern are quite common. Note that these nouns take /-ə/ as plural marker, which, however, is not confined to this class of nouns (see also cases in (3)):[7]

(8) *Monosyllabic stems*

	Base	Diminutive	Plural	
/n/	*Dänn*	*Dänn-eke*	*Dänn-e*	'fir'
	Pann	*Pänn-eke*	*Pann-e*	'pan'
/l/	*Brell*	*Brell-eke*	*Brell-e*	'glasses'
	Mull	*Müll-eke*	*Mull-e*	'mouth, trap'

Polysyllabic nouns are mostly loanwords. They also tend to mark the plural of the base form with /-ə/.

(9) *Polysyllabic stems*

	Base	Diminutive	Plural	
/n/	*Maˈschinn*	*Maschinn-eke*	*Maschinn-e*	'machine'
	Garˈdinn	*Gardinn-eke*	*Gardinn-e*	'curtain'
/l/	*Kaˈpäll*	*Kapäll-eke*	*Kapäll-e*	'chapel'
	Moˈdäll	*Modäll-eke*	*Modäll-e*	'model'
	Kaˈnonn	*Kanönneke*	*Kanonn-e*	'canon'
	Karuˈssäll	*Karusäll-eke*	*Karusäll-s*	'carousel'
	Kaseˈroll	*Kaseröll-eke*	*Kaseroll-e*	'casserole'

One could analyze the allomorph /-əkə/ as a combination of the default form /-kə/ with additional schwa epenthesis – as some kind of reflex of a prosodic minimality condition. However, I am not aware of any other morphological

---

[7] Kleverlandish plural forms also seem to be subject to a prosodic minimality requirement (being at least bimoraic). However, the additional requirement that there be an overt exponent for plural leads to the insertion of further segmental or super-segmental material if the base already fulfills this minimality condition (e.g. umlaut, suffixation of /ə/, /s/ etc.).

derivation in Kleverlandish that would require schwa epenthesis just with stems of this makeup.

A comparison with Standard Dutch may be instructive. Standard Dutch (Booij 1995, Gussenhoven and Jacobs 1998) has a diminutive form that seems to be composed of an epenthetic schwa and the default form /-tjə/, namely /-ətjə/. This allomorph, however, is suffixed to all sonorant-final stems, as the following examples illustrate.[8]

(10) *Dutch schwa-epenthesizing diminutives*
(Gussenhoven and Jacobs 1998)

/ŋ/	slɑŋ	slɑŋ-ətjə	'snake'
/m/	bɔm	bɔm-ətjə	'bomb'
/n/	pɑn	pɑn-ətjə	'pot'
/l/	bɑl	bɑl-ətjə	'ball'
/r/	snɔr	snɔr-ətjə	'moustache'

Even with this more general pattern of schwa epenthesis/prosodic minimality requirement, the question arises as to why short vowel stems with final obstruents should project a different prosodic structure than short vowel stems with final sonorants (see, for instance, Botma and Torre 2000 for such a proposal). Given that the Kleverlandish /-əkə/-diminutives only occur with a subset of the sonorant-final stems, which renders a relation between segmental and prosodic structure even more questionable, I assume that there is no productive rule of schwa epenthesis; I take the allomorph to be the fixed sequence /-əkə/ with the combined segmental-prosodic input condition mentioned above.

4. Stems with a Final Dorsal Consonant

At first sight the allomorph /-skə/ seems to occur after velar-final stems, as evidenced by stems ending in /k/ or /ŋ/:

---

[8] The Dutch dialect of Sittard does not exhibit any diminutive allomorph with initial schwa, although it resembles Kleverlandish in most other respects (see Gussenhoven and Jacobs 1998: 109-112).

(11)	/k/	*Bukk*	*Bükk-ske*	'book'
		Dakk	*Däkk-ske*	'roof'
		Pläkk	*Pläkk-ske*	'spot'
	/ŋ/	*Schlang*	*Schläng-ske*	'snake'
		Chaiselongue	*Chaiselöng-ske*	'chaise longue'
		Kartong	*Kartöng-ske*	'cardboard'

The picture gets more complex with nouns that exhibit an underlying /g/ as final stem segment. This velar is subject to a spirantization rule (/g/-spirantization, Wiese 1996: 206-09), which applies in syllable-final position. (12) illustrates /g/-spirantization; the plural form provides evidence for the underlying representation as velar plosive:

(12) /g/-*Spirantization*

Singular		Plural		
Saag	[χ]	*Saag-e*	[g]	'saw'
Oog		*Oog-e*		'eye'

The rule for /g/-spirantization may be stated as follows:

(13) (Wiese 1996: 207)

$$\textbf{GSP}\text{(/g/-spirantization): [continuant]} \rightarrow \text{[+continuant]} \; / \left[\begin{array}{c} \_\_ \\ \text{+voice} \\ \text{+obstruent} \\ \text{Dorsal} \end{array}\right]\Bigg]_{\sigma}$$

As the examples in (12) demonstrate, Kleverlandish /g/-spirantization is not confined to contexts with preceding /i/ – unlike Standard German (Standard German: *König* [ç] 'king'). Moreover, /g/-spirantization feeds another well-known rule for consonants, namely dorsal fricative assimilation (DFA), often dubbed "*ich-ach* alternation" (see Wiese 1996, Robinson 2001): depending on the preceding segment, the dorsal fricative surfaces as [ç], [x] or [χ].

Before demonstrating the distribution of the allophones let me point out that Kleverlandish has only few lexical items with underlying dorsal fricatives. This results from the fact that Kleverlandish has not undergone the High German consonant shift (spirantization of plosives), which created many forms with dorsal fricatives. The following examples illustrate the difference:

(14) *Fricative-stop correspondence*

High German		Kleverlandish		
*Da**ch***	[χ]	*Da**kk***	/k/	'roof'
*Kü**ch**e*	[ç]	*Köö**k***		'kitchen'
*Schi**ff***	/f/	*Sche**pp***	/p/	'ship'
*Stra**ß**e*	/s/	*Stroo**t***	/t/	'street'

Underived dorsal fricative phonemes are confined to positions before /t/.

(15) *Dorsal fricatives*

[χ]	*a**ch**ter*		'behind'
	Macht		'power'
	Dochter		'daughter'
[ç]	*Knä**ch**t*	[æ]_	'farmhand'
	Löcht	[ø]_	'candlestick'
	Plecht	[e]_	'duty'
	Trechter		'funnel'

These examples already indicate that [ç] surfaces after front vowels. In contrast to Standard German there are no Kleverlandish items in which an underlying dorsal fricative is preceded by one of the sonorants /l, ʀ, n/, which then would trigger the realization of the fricative as [ç].[9] [χ] is realized after non-high back vowels; the third allophone [x] appears after high back vowels (see (18-a) and (20)).

Since the allophone [ç] does not have a wider distribution than the two other allophones and since most dorsal fricatives result from /g/-spirantization I analyze [x/(ɣ)] as basic and derive [ç/(ʝ)] and [χ/(ʁ)] via fronting or lowering, respectively. I adopt Wiese's (1996: 23) featural system for describing the inventory of dorsal consonants in Kleverlandish, which is given in table 1.

---

[9] In my search for such items I managed to elicit a diminutive for the High German loanword *Molch* 'newt': *Mölch-ske.*

	ç/(ʝ)	k/g	x/(ɣ)	ŋ	χ/(ʁ)	ʀ
obstruent	+	+	+	−	+	−
dorsal	+	+	+	+	+	+
continuant	+	−	+	−	+	+
nasal	−	−	−	+	−	−
front	+	−	−	−	−	−
high tongue position	+	+	+	+	−	−
low tongue position	−	−	−	−	+	+

Table 1: Dorsal consonants in Kleverlandish

The two rules of dorsal fricative assimilation can be formulated as follows:

(16) a. **DFF** (dorsal fricative fronting):

$$\begin{bmatrix} \text{Dorsal} \\ \text{+obstruent} \\ \text{+continuant} \end{bmatrix} \rightarrow [\text{+front}] \; / \begin{bmatrix} -\text{consonantal} \\ \text{+front} \end{bmatrix} \_\_$$

b. **DFL** (dorsal fricative lowering):

$$\begin{bmatrix} \text{Dorsal} \\ \text{+obstruent} \\ \text{+continuant} \end{bmatrix} \rightarrow [\text{+low}] \; / \begin{bmatrix} -\text{consonantal} \\ -\text{front} \\ -\text{high} \end{bmatrix} \_\_$$

Nouns with a final /g/ segment are subject to a complex interaction of phonological rules. As Kiparsky (1968) already pointed out, /g/-spirantization must be ordered before final devoicing because it would otherwise be bled by final devoicing, as illustrated for *Oog* 'eye' in (17). /g/-spirantization feeds the dorsal fricative assimilation rule complex (DFF/DFL):

(17) a. /o:g/ $\rightarrow_{GSP}$ [o:ɣ] $\rightarrow_{DFL}$ [o:ʁ] $\rightarrow_{FDV}$ [o:χ]
b. /o:g/ $\rightarrow_{FDV}$ [o:k] $^{*}\rightarrow_{GSP}$

In umlaut contexts the dorsal fricative alternation shifts the derived fricative to its front dorsal counterpart [ç], as shown for the following contrast between *Zug* 'train' vs. *Züüg* 'trains'. In the context of a preceding high back vowel (see (18-a)), the fricative is realized as the velar allophone because neither DFL nor DFF apply:

(18) a. /tsʊg/ $\rightarrow_{GSP}$ [tsʊɣ] $\rightarrow_{FDV}$ [tsʊx]
b. /tsy:g/ $\rightarrow_{GSP}$ [tsy:ɣ] $\rightarrow_{DFF}$ [tsy:ɟ] $\rightarrow_{FDV}$ [tsy:ç]

Diminutive formation of nouns with /g/-final stems follows the pattern in (18-b); the sequence of rules must be enriched by diminutive suffixation (DIM) and umlaut (UML):

(19) /… V… g/ $\rightarrow_{DIM}$ [… V… g-skə$^{[+\text{front}]}$] $\rightarrow_{UML}$ [… V$_{[+\text{front}]}$… g-skə]
$\rightarrow_{GSP}$ [… V$_{[+\text{front}]}$… ɣ-skə] $\rightarrow_{DFF}$ [… V$_{[+\text{front}]}$… ɟ-skə]
$\rightarrow_{FDV}$ [… V$_{[+\text{front}]}$… ç-skə]

I assume that the diminutive suffix has a floating feature [+front], which is linked to the stem vowel. (20) shows some respective diminutive forms.

(20) *Nouns with final /g/*

Singular		Plural		
Saag	[χ]	*Sääg-ske*	[ç]	‘saw’
Oog	[χ]	*Öög-ske*		‘eye’
Plugg	[x]	*Plügg-ske*		‘plough’

If one would conclude from the examples in (11) that the diminutive allomorph /-skə/ is restricted to stems that end in velar consonants, the forms in (20) would appear to be cases of counterbleeding (see Baković’s 2011 formulation of counterbleeding in (21)), i.e., umlaut and dorsal fricative fronting would change the right edge of the stem in such a way that it could not be the input for /-skə/ suffixation.

(21) B counterbleeds A if B eliminates potential inputs to A and A precedes B.

However, the context for /-skə/ is best characterized as an underspecified final dorsal stem segment. Note that the set of dorsal consonants includes the rhotic /ʀ/, which is vocalized in rhyme position. However, since the syllabic position of /ʀ/ is only determined after morphological processes such as diminutive formation, one cannot assume vocalization prior to diminutive suffixation. Therefore, /ʀ/ has to be excluded from the set of possible right stem edges for /-skə/, which means that the context specification for /-skə/ should be set to [Dorsal,+obstruent].

Since there are no equivalent cases of /s/-epenthesis in other domains of Kleverlandish morphology, I refrain from postulating a rule of /s/-epenthesis for the default form of the diminutive (but see Gussenhoven and Jacobs 1998 for such an SPE-like treatment of the /-skə/-allomorph in various Dutch dialects). I will return to this issue in the next section.

5. A Complex Case of Cluster Simplification

The last pattern of diminutive formation to be considered is the complication that arises with stems that end in complex clusters of the type fricative and /t/. Quite unexpectedly, these stems do not take the allomorph /-tjə/, which is generally used after other /t/-final stems. Instead, the cluster is simplified by deletion of the stem-final plosive and the diminutive is marked with the allomorph that fits the derived environment.[10]

Diminutive-induced cluster simplification is found systematically with the coda cluster /st/; the two other contexts are only evidenced by a few lexical items. The simplification of the underlying coda clusters /st/ and /ft/ yields a context for the default allomorph /-kə/. If the cluster-initial fricative is dorsal, the allomorph /-skə/ is used.

(22) Cluster simplification

/st/ → [s]	*Kast*	*Käs-ke* **Käs(t)-je*	'closet'
	Fust	*Füs-ke*	'fist'
	Knust	*Knüs-ke*	'chunk (of bread)'
/ft/ → [f]	*Geschäft*	*Geschäf-ke*	'business'
/[Dorsal]t/ → [ç]	*Gesecht*	*Gesech-ske*	'face'
	Lecht	*Lech-ske*	'light'

These diminutives exhibit a rule ordering paradox: the final plosive is deleted in the context of diminutive formation (and possibly other obstruent-initial suffixes). Therefore, diminutive formation should apply before cluster simplification. But the latter bleeds the context for the diminutive allomorph that would be used in the first place. With regard to the forms that take the default allomorph one could solve the indicated dilemma by assuming a more specific

---

[10] There is a general (optional) rule of /t/-deletion in Standard Dutch (see Booij 1995: 152-154): if a complex coda ending in /t/ is followed by an obstruent, /t/ may be deleted. Since I have not studied this process in Kleverlandish more thoroughly, I cannot provide information on the full range of /t/-deletion contexts.

input condition for the allomorph /-tjə/: it only attaches to stems in which the coronal non-continuant is preceded by a sonorant.

However, the situation is more complex with the forms that take the allomorph /-skə/. Given the general character of /t/-deletion in such phonological contexts, it is not plausible to assume that the final /t/ has undergone spirantization (/t/ → [s]). It is also not a very elegant solution to add a disjunction to the input specification, which includes /[Dorsal]t/ as a further context for /-skə/. A possible, though not very attractive solution would be to assume that /s/-epenthesis functions as some kind of repair mechanism in contexts in which the default form of the diminutive is realized in adjacency to a dorsal consonant on the surface. As already indicated above, there are no further parallel contexts for /s/-epenthesis. This rule would be confined to diminutives and would actually only be needed in the very few relevant cases in (22). A final, also not very attractive solution would be to split up diminutive formation in three steps:

(23) "abstract diminutive formation" (/C.../$^{[+\text{front}]}$ ⇒ phonological rules ⇒ allomorphy selection

Those aspects of the diminutive that are required to feed or trigger the respective phonological rules are part of an abstract diminutive morpheme (i.e., the initial consonant and the umlaut feature); after the application of the phonological rules, the specific allomorphs are chosen. However, such a solution, which is usually not implemented in the morpheme-based lexical approach to morphology that I am pursuing here, should be justified on more phenomena than a few problematic cases of diminutive formation.

6. Summary

The previous sections have shown that Kleverlandish diminutive formation is affected by various general phonological rules: final devoicing, /g/-spirantization, /ʀ/-vocalization and dorsal fricative assimilation. Since the diminutives trigger umlaut, a further umlaut/fronting rule comes into play. The four diminutive allomorphs are targeted differently by the various rules: the default form /-kə/ is affected least, whereas the allomorph /-skə/ is involved in a complex interaction of rules due to the fact that the coda of its base is subject to various rules. Stems ending in coronal consonants show the greatest variation in diminutive formation in that all four allomorphs may occur – if the cases discussed in section 5 are included.

Two of the allomorphs, namely /-əkə/ and /-skə/, could be derived by respective epenthesis rules from the default form, whereas there is no obvious phonological relation between the allomorph /-tjə/ and the default form. However, since these epenthetic rules are diminutive-specific, their explanatory power is not much higher than the assumption of four lexical alternatives with specific input conditions. The lexical specification of the allomorphs is given in (24). 'X' denotes a segment, 'V́' a stressed short vowel. Following general assumptions on input specificity, the allomorph with the most specific compatible input is selected in Kleverlandish diminutive formation.

(24) *Lexical specification of the allomorphs*
/-kə/$^{[+\text{front}]}$; $]_{\text{N.stem}}$ —
/-skə/$^{[+\text{front}]}$; $X_{[\text{Dorsal},+\text{obstruent}]}]_{\text{N.stem}}$ —
/-əkə/$^{[+\text{front}]}$; $\acute{V}X_{[\text{Coronal},-\text{obstruent},-\text{continuant}]}]_{\text{N.stem}}$ —
/-tjə/ $^{[+\text{front}]}$; $X_{[-\text{obstruent}]}X_{[\text{Coronal},-\text{continuant}]}]_{\text{N.stem}}$ —

Note that the specification for /-skə/ does not take into account the problematic /t/-deletion cases. Without further evidence that this pattern is productive in Kleverlandish, I leave the discussion to the aspects mentioned in section 5.

References

Baković, Eric (2011): Opacity and Ordering. *In:* J. Goldsmith, A. Yu and J. Riggle, eds, *The Handbook of Phonological Theory*. 2nd edn, Wiley-Blackwell, Oxford, pp. 40–67.

Booij, Geert (1995): *The Phonology of Dutch*. Oxford University Press, Oxford.

Botma, Bert and Erik Jan van der Torre (2000): The Prosodic Interpretation of Sonorants in Dutch. *In:* H. d. Hoop and T. v. d. Wouden, eds, *Linguistics in the Netherlands 2000*. John Benjamins, Amsterdam, pp. 17–29.

Gussenhoven, Carlos and Haike Jacobs (1998): *Understanding Phonology*. Hodder Education, London.

Kiparsky, Paul (1968): Linguistic Universals and Linguistic Change. *In:* E. Bach and R. T. Harms, eds, *Universals in Linguistic Theory*. Holt, Reinhart, and Winston, New York, pp. 170–202.

Robinson, Orrin W. (2001): *Whose German? The* ach/ich *Alternation and Related Phenomena in 'Standard' and 'Colloquial'*. John Benjamins, Amsterdam.

Schönberner, Egon (1998): *Wortschatz des unteren Niederrheins*. Vol. 1, Verlag Erika Schönberner, Kleve.

Van Aewerdäss bes Wilewaj: Mundartwörterbuch (1995). Boss, Kleve. [cited as VABW].

Wiese, Richard (1996): *The Phonology of German*. Oxford University Press, Oxford.

The Super-Strong Person Case Constraint: Scarcity of Resources by Scale-Driven Impoverishment

Aaron Doliana[*]

Abstract

Kambera, a Malayo-Polynesian language, shows a new version of the Person Case Constraint (PCC), disallowing any combination of phonologically weak objects except the one where the indirect object is 1st/2nd person and the direct object is 3rd person. Recent minimalist accounts fail to capture this new pattern, which, I claim, indicates the existence of a continuum within the constraint's typology. In this paper, I am going to account for this new version as a syntactic rule-interaction effect between Agree and scale-driven Impoverishment. I claim that with this mechanism, set along the lines of an Optimality Theoretic version of the Minimalist Programm, the whole typology of the PCC can be accounted for.

1. Introduction

This paper aims to show by means of the *super-strong* version of the Person Case Constraint that there is a continuum in the typology of the PCC, and thus to account for the full typology as a syntactic rule-interaction effect between Agree and scale-driven Impoverishment. The PCC is a constraint on combinations of phonologically weak objects in ditransitive constructions, depending on their person feature specifications. The super-strong version allows only combinations of 1st/2nd person indirect object and 3rd person direct object. In spite of their ability to derive other versions, this version of the

---

[*] For valuable comments and suggestions I am grateful to the participants of the syntax colloquium at the University of Leipzig, to the audience of the *Workshop on Interaction of Syntactic Primitives* at the annual DGfS meeting 2013 in Potsdam, and especially to Anke Assmann, Doreen Georgi, Gereon Müller and Philipp Weisser, who have helped me through my previous drafts. The research was carried out in the project *Argument Encoding in Morphology and Syntax* as part of the DFG research group 742 *Grammar and Processing of Verbal Arguments*.

Rule Interaction in Grammar, 177–202
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

PCC cannot be derived in existing minimalist approaches such as Anagnostopoulou (2005), Adger and Harbour (2007), Richards (2008) or Nevins (2007). These build on the intuition that the PCC arises in "two-arguments-against-one-head situations" where the functional head entering Agree with the two arguments lacks the resources to check the features of both arguments. I will from here on call this intuition *scarcity of resources*. In this paper I will attempt to rescue this assumption by relativising it to the effects of syntactic Impoverishment following from the Harmonic Alignment of markedness scales (cf. Keine and Müller 2008, 2009, Keine 2010). With the additional assumption that the operation Agree is split up into two sub-operations, *Copy* and *Check*, the PCC follows in three derivational steps: (i) the π-features of the goals are copied and transferred onto the probe by Copy; (ii) Impoverishment applies to the probe, due to the harmonic alignment of markedness scales interacting with a faithfulness constraint protecting the copied π-features on the probe; (iii) the scarcity of resources caused by Impoverishment bleeds Check, which deletes uninterpretable features under the feature identity of probe and goal, and the derivation crashes. Consequently, the intuition of the scarcity of resources on the probe is saved (although relativised to the effects of Impoverishment), PCC effects can be linked to Hale/Silverstein hierarchy effects and the full typology of the PCC can be derived without having to assume asymmetries between the representations of 1st/2nd and 3rd person.

This paper is divided into three sections. In section 2 I will summarise the background of the Person Case Constraint, describe the super-strong version of it and discuss the pros and cons of Anagnostopoulou's (2005), Adger and Harbour's (2007) and Haspelmath's (2004) approaches to the PCC. In section 3 I will introduce the theoretical background and in section 4 I will present my assumptions, propose my approach to the PCC and show some of its consequences. In conclusion I will discuss the consequences more generally and direct to open questions that need further research.

2. The Person Case Constraint

2.1. Background

The Person Case Constraint, also known as the **me-lui* Constraint, is a restriction on possible combinations of phonologically weak elements. This restriction was first reported for French by Perlmutter (1971). A ditransitive construction is grammatical if the indirect object (IO) is local person (i.e. 1st, 2nd person and reflexive pronouns) and the direct object (DO) is 3rd person as shown in (1).

(1) On me le montrera.
one 1.DAT 3.ACC show.FUT
'They will show it to me.'[1]

However, the same sentence is ungrammatical if the indirect object is 3rd person and the direct object is local person, cf. (2).

(2) *On me lui montrera.
one 1.ACC 3.DAT show.FUT
'They will show me to him.'

The PCC was first thoroughly analysed by Bonet (1991, 1994) who noticed the following properties: (i) it applies in a large range of unrelated languages; (ii) it applies only to phonologically *weak* elements, i.e. clitics, agreement affixes and weak pronouns; (iii) it applies only to combinations of phonologically weak elements; (iv) it also applies to combinations where the DO is a reflexive element; (v) it only affects constructions with an external argument.

Apart from the super-strong version of the PCC, two further versions[2] have been found so far: the *strong* version and the *weak* version of the PCC. The former disallows local person direct objects in double-object constructions in general, whereas the latter disallows local person direct objects only when the indirect object is 3rd person.

Furthermore, the PCC is argued to pattern together with other phenomena constraining certain combinations of person features or combinations of certain person features with certain ϕ-features. On the one hand the PCC is argued to have the same syntactic origin as Dat-Nom constructions in Icelandic (Anagnostopoulou 2005), case syncretism in Kiowa and French (Adger and Harbour 2007), defective Agree in Russian (Richards 2008) or limited plural

---

[1] Abbreviations are as follows: ACC (accusative), DAT (dative), 1 (1st person), 2 (2nd person), 3 (3rd person), SG (singular), PL (plural), THM (theme), REC (recipient), FUT (future). I will furthermore adopt the notation <x, y>, where x = IO and y = DO.

[2] See Nevins (2007) for more versions of the PCC, which will not be discussed in this paper. Those are versions of the PCC where 1st and 2nd person do not pattern together as local person (*me-first* PCC and *strictly-descending* PCC). But see also Sturgeon et al. (2011) for an approach deriving the strictly-descending PCC as a linearisation effect.

agreement in Pazar Laz (Blix 2012). On the other hand it is argued by Haspelmath (2004, 2011) to be a reflex of Hale/Silverstein hierarchies (Hale 1972, Silverstein 1976), whose effects can best be seen in inverse systems and limited plural marking.

2.2. The Super-Strong PCC

Haspelmath (2004) introduces the super-strong version of the PCC, found in the Malayo-Polynesian language Kambera. The following data from Klamer (1997: 903-904) show that in Kambera ditransitive constructions only the configuration <1/2, 3> is allowed. Thus, in addition to the combinations prohibited in the strong PCC, the super-strong version also prohibits <3, 3> combinations as can be seen in (3-c).

(3) a. Na-wua-ngga-nya
3SG.AG-give-1SG.REC-3SG.THM
'He gives it to me.' – ✓ <1, 3>
b. Na-wua-nggau-nja
3SG.AG-give-2SG.REC-3PL.THM
'He gives them to you.' – ✓ <2, 3>
c. *Na-wua-nja-nya
3SG.AG-give-3PL.REC-3SG.THM
'He gives it to them.' – *<3, 3>
d. *Na-wua-ngga-nggau
3SG.AG-give-1SG.REC-2SG.THM
'He gives you to me.' – *<loc, loc>

In ditransitive constructions in Kambera only combinations of local person indirect objects with 3rd person direct objects are allowed[3].

Including the super-strong version and maybe language types such as German, which, for most verbs, allow any combination,[4] we see that there is a continuum within the typology of the PCC. This can be seen in Table 1. This complicates the phenomenon, as it can no longer be analysed as a constraint

---

[3] It may be worth noticing that in ditransitive constructions in Kambera both objects bear the dative case (cf. Georgi 2008 for a detailed analysis of argument encoding in Kambera). Furthermore, the super-strong PCC has so far only been found in Kambera.

[4] But see also Anagnostopoulou (2008), who argues that there are PCC effects in German in non-default word orders.

IO	DO	super-strong	strong	weak	German
1/2	3	✓	✓	✓	✓
3	3	✗	✓	✓	✓
1/2	1/2	✗	✗	✓	✓
3	1/2	✗	✗	✗	✓

Table 1: Typology of the Person Case Constraint

against a certain person feature in a certain context. Hence, unless one treats the patterns in these languages as epiphenomena of further constraints, one must analyse the PCC as a continuum. This is why the minimalist accounts I have mentioned cannot derive the super-strong version of the PCC (or a zero version as in German). Nonetheless, I would like to contend that their ideas offered fundamental insight that should be maintained. Therefore, these approaches will be the basis of a new account which comprises the idea of scarcity of resources, Optimality Theoretical modelling of scales and a rule interaction effect with Agree.

2.3. Existing Approaches

2.3.1. Scarcity of Resources

Anagnostopoulou (2005) proposes scarcity of syntactic resources as the trigger for the PCC. In general this means that there is a "two arguments against one head situation". More precisely, in this approach it consists of one functional head (viz. little *v*) entering Agree with both objects, but having only one set of ϕ-features to give them, which means that only one argument can have its full set of ϕ-features (viz. its person feature) checked. The indirect object, being closer (in terms of c-command) to the functional head, undergoes Agree first and gets its person feature checked. All that is left for the direct object to agree with is number, since Anagnostopoulou assumes that the indirect object does not undergo Agree for number. Crucially, Anagnostopoulou makes another assumption. As can be seen in (4)[5], the person feature system

---

[5] [Participant] stands for discourse participant. [+Part] is thus local person and [−Part] 3rd person. [Author] stands for author or narrator. [+Auth] means 1st person, [−Auth] 2nd and 3rd.

she assumes exhibits an asymmetry. 1st and 2nd person are always specified for their person features, whereas 3rd person can be optionally underspecified.

(4) 1: [+Author, +Participant]
2: [−Author, +Participant]
3: [−Participant]; [ ]

Anagnostopoulou justifies this asymmetry with "contextual salience". 3rd person can be underspecified when not salient, but is required to be specified for person when salient. Thus in ditransitive constructions 3rd person indirect objects, being salient, are always specified for person, whereas 3rd person direct objects can be underspecified.

The strong version of the PCC follows because whenever the probe enters Agree with its goals, it first checks the person feature on the indirect object, leaving only number for the direct object. Local person direct objects are ruled out because they necessarily have a person feature. Since it cannot be checked, double object combinations with a local direct object lead to a crash in the derivation. Languages with the weak version of the PCC, on the other hand, are argued to have an optional Multiple Agree mechanism, which allows the probe to check the person features of the two goals simultaneously. If the two features are identical (e.g. [+Part] and [+Part]; [Auth] plays no role here) the derivation converges and <loc, loc> configurations are saved. The super-strong version of the PCC, however, cannot be accounted for: both objects would either need to be specified or underspecified for [Part] in order to derive the ungrammaticality of <3, 3>, not being able to check the person feature on the direct object or leaving an uninterpretable person feature on the probe. This is ruled out by the fact that the two objects cannot both be salient to the same degree, resulting in different feature specifications.

2.3.2. Domain-Specific Restrictions

In Adger and Harbour (2007) the PCC also arises from a "two arguments against one head situation", but the system proposed is slightly different. The two phonologically weak objects merge with an Appl-head[6]: the direct object as its complement and the indirect object as its specifier. In addition, the Appl-head has the ability to ban a feature in its complement domain and to re-

---

[6] Appl stands for applicative in the sense of Pylkkänen (2002).

quire the same feature in its specifier, the value of the feature being irrelevant. Adger and Harbour assume the same person feature system as Anagnostopoulou in (4). They motivate it with observations on case syncretisms. The strong version of the PCC follows if the feature banned and required by the Appl-head is [±Part]. The 3rd person can be both direct or indirect object as it can be underspecified for [Part] and escape the ban when being the complement, and be specified for [Part] and fulfil the requirement when it is the specifier. Local person, though, can only be the indirect object as it has to be specified for [Part] and can never escape the ban on its feature in the complement domain. The weak version of the PCC is not considered in Adger and Harbour's approach because there seems to bee too much variance between regions and speakers as to which combinations of <local, local> are allowed.[7] The following points remain unclear: how the ban and requirement are modelled on the head, which features can be banned and required and what lies behind the asymmetry between [Participant] and [Empathy], both entailing semantic animacy, but only [Part] being responsible for the PCC.

2.3.3. Markedness Scales

Haspelmath (2004) is a diachronic, frequency-based approach to the PCC. The focus does not lie on combinations of person and case, but on combinations of person and semantic roles; although this difference is irrelevant for what follows. Haspelmath argues for a grammaticalisation effect, where over time only the more frequent structures are grammaticalised. In this case only the more frequent pronoun combinations are grammaticalised into clitic combinations, whereas their less frequent counterparts are not and are hence ungrammatical as clitics. The frequency of the pronoun combinations is related to Silverstein/markedness scales: indirect objects (or recipients) tend to be 1st or 2nd person and direct objects (or themes) tend to be 3rd person. The unmarked combination in double object constructions is therefore <loc, 3>, which is allowed in almost all languages exhibiting the PCC. The most marked combination, on the other hand, is <3, loc>, which is forbidden in almost all

---

[7] Interestingly, the super-strong version of the PCC, which is also not taken into consideration in their paper, could possibly be derived in their system. If the feature banned and required by the Appl-head were [Author], the only grammatical combination would be <loc, 3>. However, this only holds if 3rd person could not bear the feature [Author] at all, a point which remains unclear in their appendix to the person feature specifications.

languages obeying the PCC. Although it does not aim to explain how the PCC works in synchronic grammars and can therefore give no answer to that question, this approach succeeds in motivating the existence of the super-strong, strong and weak versions of the PCC. It also predicts a fourth version of the PCC, disallowing <3, 3> and <3, loc>. However, to the best of my knowledge, this version has not been attested in any language so far.

3. Theoretical Background

3.1. Impoverishment

Impoverishment (Halle and Marantz 1993, Noyer 1998, Keine and Müller 2008, 2009, Keine 2010, Bank, Sappir and Trommer 2012) is a post-syntactic feature deletion operation. It was first introduced within the framework of Distributed Morphology (DM, Halle and Marantz 1993), where it has the form of transformational rules and deletes certain features in certain contexts. DM operates under the assumptions of the *Subset Principle* and *Specificity*.[8] The former states that a vocabulary item V_1 is inserted in a functional head F when its features form a subset of the functional head's features and V_1 is more specific than any other compatible vocabulary item V_i. The latter standardly states that a vocabulary item V_1 is more specific than a vocabulary item V_2 iff V_1 has more features than V_2. Thus, whenever Impoverishment applies, deleting certain features, a vocabulary item, otherwise the most specific, may no longer fit, giving room for the insertion of a less specific exponent. A typical example of Impoverishment, shown in (5), is the deletion of the feature [+object] in the context of singular neuter nouns in several Indo-European languages such as German. This leads to a syncretism between the nominative and the accusative case on singular neuter nouns because the distinctive feature [+object] is deleted.

(5) /[+obj]/ → ∅ / [−mask, −fem, −pl]

In the approach to be developed here, however, I will follow Keine and Müller (2008, 2009) and Keine (2010), who, building on the work by Aissen (1999,

---

[8] It also operates under the assumptions of *Late Insertion*, i.e. the morphological exponents are inserted after all syntactic processes have terminated, and *Syntactic Hierarchical Structure all the Way Down*, i.e. syntactic hierarchical structure does not stop at the word level, but rather goes down all the way to morphemes.

2003), developed a more restrictive theory of Impoverishment, ultimately driven by ranked and violable constraints in an Optimality Theoretic fashion. In this approach faithfulness constraints penalising featural changes (viz. deletion) compete with markedness constraints penalising the presence of certain features (hence demanding deletion). Consequently, the ranking between these two types of constraints determines whether or not Impoverishment applies. This is achieved in Keine and Müller (2008, 2009), who posit Harmonic Alignment of markedness scales at its base. Finally, Keine (2010) takes Impoverishment to apply in syntax, allowing it to interact freely with other syntactic operations such as Agree. Since these two assumptions play a major role in the following approach to the PCC, I shall explain briefly the mechanisms involved and give their theoretical background.

3.2. Optimality Theory and Harmonic Alignment of Scales

Optimality Theory (OT) was originally introduced as a phonological framework by Prince and Smolensky (1993, 2004). Since then it has also been adopted in syntactic analyses (cf. Kiparsky 1999, Wunderlich 2000, Stiebels and Wunderlich 2000, Stiebels 2002, Lee 2003). The main idea of OT is that grammatical constraints are *ranked*, *violable* and *universal*. Consequently, not satisfying a constraint does not strictly lead to ungrammaticality. Rather, it is the competition between different potential outputs that gives linguistic expressions grammatical status: an output is well-formed if it is optimal with respect to a given constraint ranking, i.e. it fares better than all its competitors. Whether an output *A* fares better than its competitor output *B*, depends on their constraint profiles. Output *A* has a better constraint profile if it violates a given constraint less often than its competitor and there is no higher ranked constraint which *A* violates, but *B* doesn't. This is important because constraints in OT are ranked strictly, which means that an output becomes suboptimal (and therefore ungrammatical) as soon as it violates a higher ranked constraint more often than another output, regardless of their relative violations of lower ranked constraints.

Moreover, within the framework of OT, two mechanisms to model hierarchical scales were given by Prince and Smolensky (Prince and Smolensky 2004, Smolensky 1993, 1995, 2006): *Harmonic Alignment* and *Local Conjunction*.

Harmonic alignment was first introduced to model sonority hierarchies in Phonology, but soon used to model Hale/Silverstein scales, too (cf. Aissen

1999, 2003). The two mechanisms are defined in (6) and (7). Basically, the first element of a binary scale is aligned with the elements of another scale, starting with the edge it is best associated with. Then the same is done for the second element of the binary scale, starting from the opposite edge. Two harmonically aligned scales result, with the most harmonic combination at its left edge and progressively less harmonic combinations towards the right edge. Furthermore, constraints can be gained from these scales by prohibiting the inverse order of the Harmonic alignment scales.

(6) *Harmonic alignment* (Prince and Smolensky 2004: 161)
Suppose given a binary dimension D_1 with the scale X > Y on its elements {X,Y}, and another dimension D2 with a scale a > b > …> z on its elements {a,b,…,z}. The *harmonic alignment* of D_1 and $D_2$is the pair of Harmony scales H_X, H_Y:
a. H_X: X/a ≻ X/b ≻ …≻ X/z
b. H_Y: Y/z ≻ …≻ Y/b ≻ Y/a

The *constraint alignment* is the pair of the following constraint hierarchies C_X, C_Y:

(7) a. *X/z ≫ …≫ *X/b ≫ *X/a
b. *Y/a ≫ *Y/b ≫ …≫ *Y/z

Local conjunction, on the other hand, is the creation of a new constraint, made up of the combination of two existing constraints. The new constraint is violated whenever both of the constraints which it comprises are violated within a given domain. Furthermore, it is inherently ranked higher than its combined parts.

(8) *Local conjunction* (Smolensky 1995: 4)
The local conjunction of C_1 and C_2 in domain D, C_1 & C_2, is violated when there is some domain of type D in which both C_1 and C_2 are violated.

Universally, the local conjunction of two constraints C_1 and C_2 outranks the individual constraints C_1 and C_2; in other words: C_1 & C_2 ≫ C_1, C_2.

3.3. Agree

The present approach is couched within the framework of an optimality-theoretic version of the *Minimalist Program* (Chomsky 2000, Adger 2003, Heck and Müller 2007) with realisational morphology. Agree is – along with Merge – one of the two structure-building operations of the framework. The operation Agree checks features under c-command, allowing the deletion of uninterpretable features and thus preventing a crash of the derivation. When certain features are involved – such as e.g. ϕ, case or tense – checking happens by *valuation*. The interpretable ϕ-features of the c-commanded element (goal) are copied and transferred to the functional head (probe) yielding the corresponding uninterpretable feature. The probe is valued by the transferred copy and its uninterpretable feature may delete once it has been checked.

In accordance with much recent work, where Agree (cf. Di Sciullo and Isac 2003, Arregi and Nevins to appear, Bhatt and Walkow to appear, Bobaljik 2008), Move (cf. Chomsky 1995, 2000), or syntactic operations in general (cf. Hornstein 2009) are decomposed into more fine-grained operations, I will split Agree into the two sub-operations Copy and Check, with the former copying and transferring the goal's features onto the probe (and thus valuing it), and the latter checking uninterpretable features under feature identity of probe and goal. This is necessary for Impoverishment to apply between the valuation (copying) and the checking of the probe, which is the key assumption of the new approach.

4. A New Approach

4.1. Assumptions

In what follows, I will make the following assumptions.

[A1] There is only one probe entering Agree with both phonologically weak elements in ditransitive constructions. The probe is made up of an ordered tuple of uninterpretable feature bundles (viz. <[uϕ], [uϕ]>) that need valuation and checking by entering Agree with two elements providing interpretable features. The ordered tuple is valued in an order related to c-command closeness, thus, roughly speaking, resulting in the form <IO, DO>. This is more or less as in Anagnostopoulou (2005).

[A2] 3rd person is always fully specified (Nevins 2007).

[A3] Impoverishment applies in syntax and is thus able to interact with operations such as Agree (Keine 2010).

[A4] Impoverishment is scale-driven: markedness constraints penalising less likely feature combinations interact in an optimality-theoretic fashion with a faithfulness constraint penalising the deletion of the features involved (Keine and Müller 2008, 2009, Keine 2010).

[A5] Impoverishment may target probes just as it may target goals.

[A6] Optimisation happens in a strictly derivational fashion (so-called "extremely local optimization"; Müller 2004, 2009, Heck and Müller 2007), only ever targeting one derivational step at a time. The step optimised in the present approach occurs between the applications of Copy and Check.

[A7] Crucially, Agree is made up of two sub-operations, cf. (9).

(9) *Agree*:
Agree is a process containing the following operations.
a. Copy: The operation copying and transferring the goal's features onto the probe.
b. Check: The operation deleting uninterpretable features under feature identity.
They apply in the only logical order Copy > Check.

4.2. Impoverishment of the Probe

The feature combinations interacting in the PCC are at least Case and Person.[9] I will further assume that cases are decomposed in binary features (Bierwisch 1967), e.g.: Nominative [–obl(ique), –obj(ect)]; Accusative [–obl, +obj]; Dative [+obl, +obj]; Genitive [+obl, –obj]. The decomposition of person is also possible, but not necessary for this account as I am focusing on PCC types that make a distinction only between local and 3rd person. The relevant scales are thus the case-feature scale in (10) and the person scale in (11), which will be the

---

[9] See Subsection 4.5 for PCC effects with further ϕ-features.

basis of the constraints at work. The first scale shows that [+obl]-arguments are more prominent than [−obl]-arguments. The second scale shows that 1st and 2nd person – patterning together as local person – are more prominent than 3rd person.

(10) *Case-feature scale*
[+oblique] > [−oblique][10]

(11) *Person scale*

$$\overbrace{\text{1st person} > \text{2nd person}}^{\text{local person}} > \text{3rd person}$$

These two scales are combined by harmonic alignment to give rise to the harmony scales in (12-a) and (12-b).The more harmonic (viz. less marked) combinations are on the left edge of the scales, whereas the less harmonic (viz. more marked) combinations are on the right edge. The OT constraints following from the prohibition against the reversed order of the harmonic scales in (12) can be seen in (13). The prohibition against less harmonic combinations is ranked higher, which in OT means that it is more difficult to violate in a well-formed output.

(12) *Harmony scales*
a. [+oblique]/local > [+oblique]/3
b. [−oblique]/3 > [−oblique]/local

(13) *Constraint alignment*
a. *[+oblique]/3 ≫ *[+oblique]/local
b. *[−oblique]/local ≫ *[−oblique]/3

As the PCC applies only to combinations, both the indirect and the direct object are relevant for triggering Impoverishment and the rankings in (13) have to be combined. This is achieved by local conjunction in (14). Recall that as defined in (8), local conjunction of two constraints C_1 and C_2 is violated whenever both constraints are violated within a given domain (by assumption the syntactic head, i.e. the probe).

---

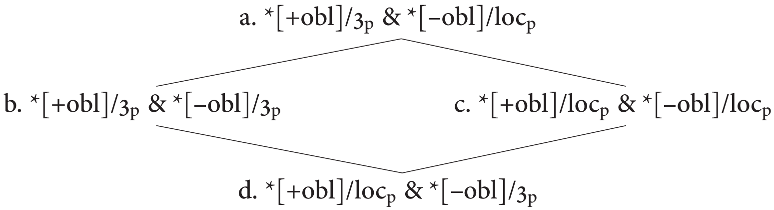
[10] As noted in footnote 3, ditransitive constructions in Kambera actually have dative case for both objects. Therefore [−obl] does actually not apply to a dative-case object, as it bears [+obl]. This could however simply be a morphological effect. The case relevant here is the syntactically assigned case, not its morpho-phonological realisation.

(14) *Local conjunction*

a. $*[+\mathrm{obl}]/3_{\mathrm{p}}\ \&\ *[-\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}} \gg *[+\mathrm{obl}]/3_{\mathrm{p}}\ \&\ *[-\mathrm{obl}]/3_{\mathrm{p}}$
b. $*[+\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}}\ \&\ *[-\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}} \gg *[+\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}}\ \&\ *[-\mathrm{obl}]/3_{\mathrm{p}}$
c. $*[-\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}}\ \&\ *[+\mathrm{obl}]/3_{\mathrm{p}} \gg *[-\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}}\ \&\ *[+\mathrm{obl}]/\mathrm{local}_{\mathrm{p}}$
d. $*[-\mathrm{obl}]/3_{\mathrm{p}}\ \&\ *[+\mathrm{obl}]/3_{\mathrm{p}} \gg *[-\mathrm{obl}]/3_{\mathrm{p}}\ \&\ *[+\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}}$

For example, the first constraint in (14-a) is violated if both a 3rd person with [+obl] case and a local person with [−obl] are present in the relevant domain. This will be the case if the two objects trigger agreement on the same verbal head, justifying the assumption that the relevant domain of the locally conjoined constraints in (14) is the probe. Furthermore, the rankings in (14) correspond to markedness in terms of Hale/Silverstein hierarchies: $*[+\mathrm{obl}]/3_{\mathrm{p}}\ \&\ *[-\mathrm{obl}]/\mathrm{loc}_{\mathrm{p}}$ is ranked higher than $*[+\mathrm{obl}]/3_{\mathrm{p}}\ \&\ *[-\mathrm{obl}]/3_{\mathrm{p}}$ as local person direct objects are less canonical than 3rd person direct objects. This can be shown graphically in (15).

(15) *Inherent ranking of markedness constraints*:



As Impoverishment arises from the interaction of markedness and faithfulness constraints, I am going to introduce the faithfulness constraint Max, which penalises deletion. More precisely, the faithfulness constraint will be relativised to the relevant feature and domain to avoid false predictions. The relevant feature is π and the relevant domain is the probe. The result is a constraint that penalises deletion of π-features on probes, cf. (16).

(16) *Max-*π_{probe}
Penalise deletion of π-features on probes.

The relative ranking of this faithfulness constraint and the markedness constraints determines whether a certain feature combination is deleted or not. It also gives rise to the different versions of the PCC, as will be demonstrated in subsection 4.3.

4.3. Derivation of the PCC

As outlined previously, the PCC is accounted for by scale-driven Impoverishment causing scarcity of resources on the probe and consequent bleeding of Check. I will illustrate this with one grammatical and one ungrammatical example of each version of the PCC. Generally, the derivation may unfold into two different directions, as I will demonstrate on the two abstract examples in (17) and (18). The first example shows why certain phonologically weak object combinations lead to a crash of the derivation; the second why, on the contrary, others lead to grammaticality.

Crash: In any case, the first step of the derivation is the copying of the goal's interpretable features onto the probe. This is triggered by the uninterpretable feature on the probe, which may only be deleted by Check if the feature identity between the probe and its goals was established. The result of copying is a valued probe with an uninterpretable feature yet to be checked, (17-a→b). The copying of certain features onto the probe may then feed Impoverishment.[11] Impoverishment of the copied features on the probe applies whenever the markedness constraint that penalises a given feature combination on the probe is ranked higher than the faithfulness constraint protecting the probe from feature deletion. This can be seen abstractly in the tableau in (17). Whenever this is the case, the copied features are deleted and the derivation continues with an empty probe. As a consequence, Check is bled, because the feature identity between the probe and its goals cannot be established. Since Check is bled, it can no longer delete the uninterpretable feature on the probe, which leads to a crash of the derivation, cf. (17-c→d).

(17) a. [$_{\text{v}}$ [uPers: <□, □>]] [$_{\text{IO}}$ [Pers: x]] [$_{\text{DO}}$ [Pers: y]] COPY →
b. [$_{\text{v}}$ [uPers: <x, y>]] [$_{\text{IO}}$ [Pers: x]] [$_{\text{DO}}$ [Pers: y]] impov. fed →

	[$_{v}$ [uPers: <x, y>]]	*<x, y>$_{\text{P}}$	MAX-π_{P}
	[$_{v}$ [uPers: <x, y>]]	*!	
☞	[$_{v}$ [uPers: < , >]]		*

c. [$_{\text{v}}$ [**uPers**: < , >]] [$_{\text{IO}}$ [Pers: x]] [$_{\text{DO}}$ [Pers: y]] CHECK bled →
d. Ungrammaticality

---

[11] Feeding, bleeding, counter-feeding and counter-bleeding are all understood as in the sense of Kiparsky (1976).

Convergence: On the other hand, if the faithfulness constraint is ranked higher than the markedness constraint, Impoverishment is not triggered. As a result, the probe maintains its valued features, cf. the tableau in (18). This has the consequence that Check may apply, because the feature identity between the probe and its goals can be established, cf. (18-c). Hence, Check deletes the uninterpretable feature on the probe and the derivation converges, cf. (18-c→d).

(18) a. [$_{\mathrm{v}}$ [uPers: <□, □>]] [$_{\mathrm{IO}}$ [Pers: x]] [$_{\mathrm{DO}}$ [Pers: y]] COPY →
b. [$_{\mathrm{v}}$ [uPers: <x, y>]] [$_{\mathrm{IO}}$ [Pers: x]] [$_{\mathrm{DO}}$ [Pers: y]] impov. bled →

	[$_{v}$ [uPers: <x, y>]]	MAX-π_{p}	*<x, y>$_{\mathrm{p}}$
☞	[$_{v}$ [uPers: <x, y>]]		*
	[$_{v}$ [uPers: < , >]]	*!	

c. [$_{\mathrm{v}}$ [~~uPers~~: <x, y>]] [$_{\mathrm{IO}}$ [Pers: x]] [$_{\mathrm{DO}}$ [Pers: y]] CHECK fed →
d. Grammaticality

4.3.1. The Super-Strong Version of the PCC

The ranking specific to languages instantiating the super-strong version is the one in (19) (where *<x, y>$_{\mathrm{d}}$ stands for *[+obl]/x$_{\mathrm{p}}$ & *[−obl]/y$_{\mathrm{p}}$). The faithfulness constraint is ranked lower than the markedness constraints penalising the ungrammatical combinations, but higher than the markedness constraint penalising the grammatical combination <loc, 3>.

(19) *Super-strong PCC Impoverishment ranking:*
*<3, loc>$_{\mathrm{p}}$ ≫ *<loc, loc>$_{\mathrm{p}}$ ≫ *<3, 3>$_{\mathrm{p}}$ ≫ **MAX-π_{p}** ≫ *<loc, 3>$_{\mathrm{p}}$

The derivation unfolds as previously described. In the first case, the markedness constraint prohibiting the combination involved is ranked higher than the faithfulness constraint. This triggers deletion because the empty probe is optimal – as shown by the pointing finger in front of the optimal candidate and the exclamation mark signalling that the constraint violation caused by the competitor was fatal. As a consequence, the feature identity of goal and probe cannot be established and Check is bled, leading to ungrammaticality.

(20) *Deriving *<3, 3>:*

a. [$_\text{v}$ [uPers: <□, □>]] [$_\text{IO}$ [Pers: 3]] [$_\text{DO}$ [Pers: 3]] Copy →
b. [$_\text{v}$ [uPers: <3, 3>]] [$_\text{IO}$ [Pers: 3]] [$_\text{DO}$ [Pers: 3]] impov. fed →

	[$_v$ [uPers: <3, 3>]]	*<3, loc>$_\text{p}$	*<loc, loc>$_\text{p}$	*<3, 3>$_\text{p}$	Max-π_p	*<loc, 3>$_\text{p}$
	[$_v$ [uPers: <3, 3>]]			*!		
☞	[$_v$ [uPers: < , >]]				*	

c. [$_\text{v}$ [**uPers**: < , >]] [$_\text{IO}$ [Pers: 3]] [$_\text{DO}$ [Pers: 3]] Check bled →
d. Ungrammaticality

In the second case, the copying of the features <loc, 3> does not lead to their deletion, because the faithfulness constraint Max-π_p is ranked higher than the markedness constraint *<loc, 3>$_\text{p}$. Therefore, the output with the full probe is optimal, which means that Check may apply and that the derivation converges.

(21) *Deriving ✓<loc, 3>:*

a. [$_\text{v}$ [uPers: <□, □>]] [$_\text{IO}$ [Pers: loc]] [$_\text{DO}$ [Pers: 3]] Copy →
b. [$_\text{v}$ [uPers: <loc, 3>]] [$_\text{IO}$ [Pers: loc]] [$_\text{DO}$ [Pers: 3]] impov. bled →

	[$_v$ [uPers: <loc, 3>]]	*<3, loc>$_\text{p}$	*<loc, loc>$_\text{p}$	*<3, 3>$_\text{p}$	Max-π_p	*<loc, 3>$_\text{p}$
☞	[$_v$ [uPers: <loc, 3>]]					*
	[$_v$ [uPers: < , >]]				*!	

c. [$_\text{v}$ [~~uPers~~: < loc, 3>]] [$_\text{IO}$ [Pers: loc]] [$_\text{DO}$ [Pers: 3]] Check fed →
d. Grammaticality

4.3.2. *The Strong Version of the PCC*

The ranking specific to languages obeying the strong version of the PCC is the one in (22): the faithfulness constraint is ranked lower than the markedness constraints penalising the ungrammatical combinations, but higher than those penalising the grammatical combinations.

(22) *Strong PCC Impoverishment ranking:*
*<3, loc>$_\text{p}$ $\gg$ *<loc, loc>$_\text{p}$ $\gg$ **Max-π_p** $\gg$ *<3, 3>$_\text{p}$ $\gg$ *<loc, 3>$_\text{p}$

The derivation of the strong version of the PCC behaves just like the previous one. The first example shows how <loc, loc> combinations are ruled out; the second how <3, 3> combinations can emerge as grammatical.

(23) *Deriving *<loc, loc>:*

a. [$_{\text{v}}$ [uPers: <□, □>]] [$_{\text{IO}}$ [Pers: loc]] [$_{\text{DO}}$ [Pers: loc]] Copy →
b. [$_{\text{v}}$ [uPers: <loc, loc>]] [$_{\text{IO}}$ [Pers: loc]] [$_{\text{DO}}$ [Pers: loc]] impov. fed →

	[$_{v}$ [uPers: <loc, loc>]]	*<3, loc>$_{\text{p}}$	*<loc, loc>$_{\text{p}}$	Max-π_{p}	*<3, 3>$_{\text{p}}$	*<loc, 3>$_{\text{p}}$
	[$_{v}$ [uPers: <loc, loc>]]		*!			
☞	[$_{v}$ [uPers: < , >]]			*		

c. [$_{\text{v}}$ [**uPers**: < , >]] [$_{\text{IO}}$ [Pers: loc]] [$_{\text{DO}}$ [Pers: loc]] Check bled →
d. Ungrammaticality

(24) *Deriving ✓<3, 3>:*

a. [$_{\text{v}}$ [uPers: <□, □>]] [$_{\text{IO}}$ [Pers: 3]] [$_{\text{DO}}$ [Pers: 3]] Copy →
b. [$_{\text{v}}$ [uPers: <3, 3>]] [$_{\text{IO}}$ [Pers: 3]] [$_{\text{DO}}$ [Pers: 3]] impov. bled →

	[$_{v}$ [uPers: <3, 3>]]	*<3, loc>$_{\text{p}}$	*<loc, loc>$_{\text{p}}$	Max-π_{p}	*<3, 3>$_{\text{p}}$	*<loc, 3>$_{\text{p}}$
☞	[$_{v}$ [uPers: <3, 3>]]				*	
	[$_{v}$ [uPers: < , >]]			*!		

c. [$_{\text{v}}$ [~~**uPers**~~: < 3, 3>]] [$_{\text{IO}}$ [Pers: 3]] [$_{\text{DO}}$ [Pers: 3]] Check fed →
d. Grammaticality

4.3.3. The Weak Version of the PCC

The ranking specific to languages exhibiting the weak version of the PCC is the ranking in (25). Once more, the faithfulness constraint is ranked higher than the constraints against the grammatical combinations <loc, 3>, <3, 3> and <loc, loc>, and lower than the constraint against the only ungrammatical combination <3, loc>.

(25) *Weak PCC Impoverishment ranking:*
*<3, loc>$_{\text{p}}$ ≫ **Max-π_{p}** ≫ *<loc, loc>$_{\text{p}}$ ≫ *<3, 3>$_{\text{p}}$ ≫ *<loc, 3>$_{\text{p}}$

The derivation of the weak version of the PCC unfolds as in the other versions. In the first case, the copying of features onto the probe leads to their deletion, to a bleeding of Check and thus to ungrammaticality. In the second case, the markedness constraint prohibiting the combination involved is ranked lower than the faithfulness constraint: deletion by Impoverishment is avoided and Check may apply, leading to grammaticality.

(26) *Deriving *<3, loc>*:

a. $[_{\text{v}}$ [uPers: <□, □>]] $[_{\text{IO}}$ [Pers: 3]] $[_{\text{DO}}$ [Pers: loc]] Copy →
b. $[_{\text{v}}$ [uPers: <3, loc>]] $[_{\text{IO}}$ [Pers: 3]] $[_{\text{DO}}$ [Pers: loc]] impov. fed →

	$[_{v}$ [uPers: <3, loc>]]	*<3, loc>$_{p}$	Max-π_{p}	*<loc, loc>$_{p}$	*<3, 3>$_{p}$	*<loc, 3>$_{p}$
	$[_{v}$ [uPers: <3, loc>]]	*!				
☞	$[_{v}$ [uPers: < , >]]		*			

c. $[_{\text{v}}$ [**uPers**: < , >]] $[_{\text{IO}}$ [Pers: 3]] $[_{\text{DO}}$ [Pers: loc]] Check bled →
d. Ungrammaticality

(27) *Deriving ✓<loc, loc>*:

a. $[_{\text{v}}$ [uPers: <□, □>]] $[_{\text{IO}}$ [Pers: loc]] $[_{\text{DO}}$ [Pers: loc]] Copy →
b. $[_{\text{v}}$ [uPers: <loc, loc>]] $[_{\text{IO}}$ [Pers: loc]] $[_{\text{DO}}$ [Pers: loc]] impov. bled →

	$[_{v}$ [uPers: <loc, loc>]]	*<3, loc>$_{p}$	Max-π_{p}	*<loc, loc>$_{p}$	*<3, 3>$_{p}$	*<loc, 3>$_{p}$
☞	$[_{v}$ [uPers: <loc, loc>]]			*		
	$[_{v}$ [uPers: < , >]]		*!			

c. $[_{\text{v}}$ [~~uPers~~: < loc, loc>]] $[_{\text{IO}}$ [Pers: loc]] $[_{\text{DO}}$ [Pers: loc]] Check fed →
d. Grammaticality

4.4. Rule Interaction

As shown in the previous subsection, there are two paths that the derivation can take:

1. The features copied onto the probe are penalised by a constraint ranked higher than the faithfulness constraint. The context for feeding Impoverishment is given because the output with the empty probe is optimal. As a consequence Check is bled, leading to ungrammaticality.

2. The features copied onto the probe are penalised by a constraint ranked lower than the faithfulness constraint. The context for Impoverishment is not given and the output with the full probe is optimal. As a consequence Check is fed, leading to grammaticality.

As a consequence, the following two general patterns in (28) emerge.

(28) *Consequent ordering of processes and interaction:*
a. Copy —feeds → deletion —bleeds → Check $\Longrightarrow$ ✗
b. Copy —feeds → Check $\Longrightarrow$ ✓

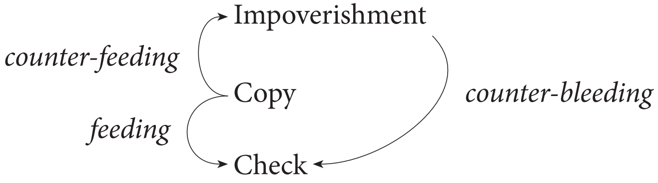
Moreover, the ordering of the three operations adopted so far (Copy > Impoverishment > Check) is the only logical one if PCC effects are to be explained this way. In fact, if the rule ordering were different – and Agree must be split for this ordering to be possible – no PCC effects would follow. Since there is only one logical ordering of Copy and Check, there are two further possible orderings involving Impoverishment: (29-b) and (29-c).

(29) *Possible rule orderings:*
a. Copy > Impoverishment > Check
b. Impoverishment > Copy > Check
c. Copy > Check > Impoverishment

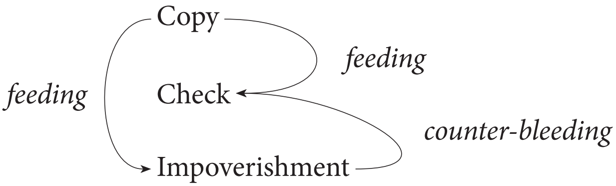
If (29-b) holds, Impoverishment will never take place. In fact, the probe would still be empty as Copy has not applied yet, meaning that the context for Impoverishment is not given yet. This would create an opaque case of rule interaction: at the surface we would ask ourselves why Impoverishment was not triggered by what appears to be a suitable context in the surface structure. The answer would be that Impoverishment may only apply at a point where the context is not given yet: a classic case of counter-feeding. Furthermore, exactly the same logic is applicable to the interaction between Impoverishment and Check. Since Impoverishment may only apply at a point where there are no features on the probe, nothing can be deleted, and with no deletion Check will never be bled. Therefore, with the rule ordering of (29-b) we have counter-feeding between Impoverishment and Copy, and counter-bleeding between Impoverishment and Check, causing no PCC effects at all.

If (29-c) holds, Impoverishment will behave just as in the rule ordering I assume for all well-formed cases in (29-a): there is no opaque interaction between Copy and Impoverishment. However, Check can no longer be bled by Impoverishment because it may only apply too late. This is a case of counter-bleeding between Impoverishment and Check and their interaction is thus opaque. In fact, the features on the probe may be deleted in certain contexts and the derivation would still converge – with the consequences for the morphology or semantics interfaces remaining unclear. Still, no PCC effects would follow. The rule interaction effects resulting from these two different rule orderings can be shown graphically as in (30-a) and (30-b) respectively.

(30) a. *Rule interaction under ordering Impoverishment > Copy > Check:*



b. *Rule interaction under ordering Copy > Check > Impoverishment:*



In sum, I tried to show here that only one of the three possible rule orderings of Copy, Check and Impoverishment leads to a successful application of Impoverishment: only in that rule ordering can Impoverishment differentiate between the ungrammatical and the grammatical combinations of agreeing objects in the languages obeying the PCC. On the contrary, in the other two orderings, and especially the one in (30-a), Impoverishment does not seem to serve any purpose. This is why I exclude those two rule orderings from playing a role in PCC effects. However, I do not exclude their existence completely, as Impoverishment might show further interactions with other operations. This may well give sense to a rule ordering such as in (30-a). In conclusion, in the case of languages exhibiting the PCC, Impoverishment must apply as soon as it can, i.e. just after Copy. In fact, this is necessary for Impoverishment to bleed Check in the right contexts and differentiate between grammatical and ungrammatical Person-Case combinations.

4.5. Consequences

Positing scale-driven Impoverishment at the basis of the PCC has the consequence that the constraint typology of Impoverishment automatically and restrictively determines the typology of the PCC as well, cf. (31). Therefore all existing PCC language types are accounted for, with a mechanism able to derive other ϕ-feature sensitive phenomena (cf. Keine 2010). However, three versions of the PCC, that have not been discussed in this paper, arise: what might be called the giga version in (31-a), the other-strong version in (31-d) and the zero version in (31-f). A language instantiating the giga version would have an absolute prohibition against double-object constructions with two phonologically weak elements. As pointed out to me by Thomas Graf (p.c.), Cairene Arabic (Shlonsky 1997: 207) is one of those languages.[12] A language with the zero version is one allowing any combination, such as German. A language with the other-strong version, on the other hand, would prohibit only the phonologically weak combinations <3, 3> and <3, loc>. Given the present assumptions, this version has to be treated as an accidental gap, as no language with that pattern has been attested so far, unless Spanish might be identified as an other-strong language, but I leave a definite answer to this question to further research. In fact, <3, 3> combinations in Spanish are only grammatical if the IO is expressed by the reflexive clitic *se* – also known as the *spurious se*. As mentioned in the background section, reflexive elements pattern together with local person. The <se, 3> combination could thus also be analysed as a repair strategy to avoid the combination *<3, 3> by replacing it with a <loc, 3> combination of the same meaning. If this were the case, Spanish would fit the other-strong version of the PCC for those speakers who allow <loc, loc> combinations. All in all, the following typology is predicted:

(31) *Rankings*

a. Giga version of the PCC: (Cairene Arabic)
$*\langle 3, \text{loc}\rangle_{\text{p}} \gg *\langle \text{loc}, \text{loc}\rangle_{\text{p}} \gg *\langle 3, 3\rangle_{\text{p}} \gg *\langle \text{loc}, 3\rangle_{\text{p}} \gg \textbf{Max}\text{-}\pi_{\text{p}}$

b. Super-strong version of the PCC: (Kambera)
$*\langle 3, \text{loc}\rangle_{\text{p}} \gg *\langle \text{loc}, \text{loc}\rangle_{\text{p}} \gg *\langle 3, 3\rangle_{\text{p}} \gg \textbf{Max}\text{-}\pi_{\text{p}} \gg *\langle \text{loc}, 3\rangle_{\text{p}}$

c. Strong version of the PCC: (French, Greek, Kiowa)
$*\langle 3, \text{loc}\rangle_{\text{p}} \gg *\langle \text{loc}, \text{loc}\rangle_{\text{p}} \gg \textbf{Max}\text{-}\pi_{\text{p}} \gg *\langle 3, 3\rangle_{\text{p}} \gg *\langle \text{loc}, 3\rangle_{\text{p}}$

d. Other-strong version of the PCC: (Spanish?)
$*\langle 3, \text{loc}\rangle_{\text{p}} \gg *\langle 3, 3\rangle_{\text{p}} \gg \textbf{Max}\text{-}\pi_{\text{p}} \gg *\langle \text{loc}, \text{loc}\rangle_{\text{p}} \gg *\langle \text{loc}, 3\rangle_{\text{p}}$

---

[12] See also Graf (2012) for an algebraic account of the PCC, which also predicts the existence of a giga version.

e. Weak version of the PCC: (Italian, Catalan, Old Occitan)
$*\langle 3, loc\rangle_p \gg$ **Max**-$\pi_\mathbf{p}$ $\gg *\langle loc, loc\rangle_p \gg *\langle 3, 3\rangle_p \gg *\langle loc, 3\rangle_p$

f. Zero version of the PCC: (German, Dutch)
Max-$\pi_\mathbf{p}$ $\gg *\langle 3, loc\rangle_p \gg *\langle 3, 3\rangle_p \gg *\langle loc, loc\rangle_p \gg *\langle loc, 3\rangle_p$

Furthermore, the analysis may be extended to capture PCC effects involving other ϕ-features, such as gender, animacy and number. In Italian both a masculine and a feminine 3rd person dative clitic exist. However, only the masculine one is grammatical in a clitic cluster (✓<3[–fem, +obl], 3[–obl]>; *<3[+fem, +obl], 3[–obl]>); in the Leísta dialects in Spanish the combination <loc, 3> is generally grammatical, unless the DO is animate (*<loc, 3[+anim]>, where 3[+anim, –obl] is syncretic with 3[+obl]).

5. Conclusion

In this paper I have shown that the scarcity-of-resources approaches can be extended to capture not only the super-strong version but the full typology of the PCC by splitting Agree into Copy and Check and letting them interact with scale-driven Impoverishment. This has the consequences that (i) the Person Case Constraint can be linked to the Hale/Silverstein scales as they are the trigger of Impoverishment; (ii) no asymmetry between the representations of 3rd and local person is needed, meaning that 3rd person is always specified in syntax, which in turn avoids complications for morphology; (iii) Agree, split into Copy and Check, can interact freely with other operations, resulting e.g. in Impoverishment of probes as has been shown. Finally, it should be noted that PCC effects involving other ϕ-features might also be accounted for with the same mechanism. I leave this issue to further research.

References

Adger, David (2003): *Core Syntax: A Minimalist Approach*. Oxford University Press, New York.

Adger, David and Daniel Harbour (2007): 'Syntax and Syncretisms of the Person Case Constraint', *Syntax* **10**, 2–37.

Aissen, Judith (1999): 'Markedness and Subject Choice in Optimality Theory', *Natural Language and Linguistic Theory* **17**, 673–711.

Aissen, Judith (2003): 'Differential Object Marking: Iconicity vs. Economy', *Natural Language and Linguistic Theory* **21**, 435–483.

Anagnostopoulou, Elena (2005): Strong and Weak Person Restrictions: A Feature Checking Analysis. *In:* H. Ordoñez and F. Ordoñez, eds, *Clitic and Affix Combinations*. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 199–235.

Anagnostopoulou, Elena (2008): Notes on the Person Case Constraint in Germanic (with special reference to German). *In:* R. D'Alessandro, S. Fischer and G. H. Hrafnbjarganson, eds, *Agreement Restrictions*. Mouton de Gruyter, Berlin/New York, pp. 15–47.

Arregi, Karlos and Andrew Ira Nevins (to appear): Contextual Neutralization and the Elsewhere Principle. *In:* O. Matushansky and A. Marantz, eds, *Distributed Morphology Today: Morphemes for Morris Halle*. MIT Press, Cambridge, Massachusetts.

Bank, Sebastian, Michael Sappir and Jochen Trommer (2012): A User's Guide to Distributed Morphology. Ms., Universität Leipzig.

Bhatt, Rajesh and Martin Walkow (to appear): 'Locating Agreement in Grammar: An Argument from Agreement in Conjunctions'.

Bierwisch, Manfred (1967): Syntactic Features in Morphology: General Problems of So-Called Pronominal Inflection in German. In: *To honour Roman Jakobson*. Mouton, The Hague, pp. 239–270.

Blix, Hagen (2012): 'The Shallowness of Being Third: Beyond Not Participating'. University of Leipzig (B.A. thesis).

Bobaljik, Jonathan David (2008): Where's Phi? Agreement as a Post-Syntactic Operation. *In:* D. Harbour, D. Adger and S. Béjar, eds, *Phi-Theory*. Oxford University Press, pp. 295–328.

Bonet, Eulália (1991): Morphology after Syntax: Pronominal clitics in Romance languages. PhD thesis, MIT.

Bonet, Eulália (1994): The Person-Case Constraint: A Morphological Approach. *In:* H. Harley and C. Philipps, eds, *MIT Working Papers in Linguistics 22. The Morphology-Syntax Connection*. MIT Press, pp. 33–52.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and J. Uriagereka, eds, *Step by step*. MIT Press, Cambridge/Mass, pp. 89–155.

Di Sciullo, Anna Maria and Daniela Isac (2003): Ways of Agreeing, Phrases, and Interfaces. *Lisbon Workshop on Agreement*. New University of Lisbon and Portugal.

Georgi, Doreen (2008): 'A Distributed Morphology Approach to Argument Encoding in Kambera', *Linguistische Berichte* 213, 45–63.

Graf, Thomas (2012): An Algebraic Perspective on the Person Case Constraint. *In:* T. Graf, D. Paperno, A. Szabolcsi and J. Tellings, eds, *Theories of Everything. In Honor of Ed Keenan*. Vol. 17 of *UCLA Working Papers in Linguistics*, pp. 85–90.

Hale, Ken (1972): A New Perspective on American Indian Linguistics. *In:* A. Ortiz, ed., *New Perspectives on the Pueblos*. University of New Mexico Press, Albuquerque, pp. 87–103.

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

Haspelmath, Martin (2004): 'Explaining the Ditransitive Person-Role Constraint: A Usage-Based Account', *Constructions* 2. Available at http://www.constructions-online.de.

Haspelmath, Martin (2011): Occurrence of Nominal Plurality. *In:* M. Dryer and M. Haspelmath, eds, *The World Atlas of Language Structures Online*. Max Planck Digital Library, München.

Heck, Fabian and Gereon Müller (2007): Extremely Local Optimization. *In:* E. Brainbridge and B. Agbayani, eds, *Proceedings of the 34th Western Conference on Linguistics (WECOL 34)*. California State University, Fresno, pp. 170–183.

Hornstein, Norbert (2009): *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press, Cambridge and UK and New York.

Keine, Stefan (2010): *Case and Agreement from Fringe to Core: A Minimalist Approach*. de Gruyter, Berlin.

Keine, Stefan and Gereon Müller (2008): Differential Argument Encoding by Impoverishment. *In:* M. Richards and A. Malchukov, eds, *Scales*. Vol. 86 of *Linguistische Arbeitsberichte*, Universität Leipzig, Leipzig, pp. 83–136.

Keine, Stefan and Gereon Müller (2009): 'Non-Zero/Non-Zero Alternations in Differential Object Marking'. Universität Leipzig.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Kiparsky, Paul (1999): 'Analogy and OT: Morphological Change as Emergence of the Unmarked'. Stanford University.

Klamer, Marian (1997): 'Spelling out Clitics in Kambera', *Linguistics* 35, 895–927.

Lee, Hanjung (2003): 'Parallel Optimization in Case Systems'. University of Minnesota.

Müller, Gereon (2004): 'Argument Encoding and the Order of Elementary Operations'. IDS Mannheim.

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. *In:* K. Grohmann, ed., *Explorations of Phase Theory*. Mouton de Gruyter, Berlin, pp. 269–308.

Nevins, Andrew Ira (2007): 'The Representation of Third Person and its Consequences for Person-Case Effects', *Natural Language & Linguistic Theory* 25(2), 273–313.

Noyer, Rolf (1998): Impoverishment Theory and Morphosyntactic Markedness. *In:* S. Lapointe, D. Brentari and P. Farrell, eds, *Morphology and its Relation to Phonology and Syntax*. CSLI, Palo Alto, pp. 264–285.

Perlmutter, David (1971): *Deep and Surface Structure Constraints in Syntax*. Holt, Rinehart and Winston, New York.

Prince, Alan and Paul Smolensky (1993): ‘Optimality Theory: Constraint Interaction in Generative Grammar’. Rutgers University.

Prince, Alan and Paul Smolensky (2004): ‘Optimality Theory: Constraint Interaction in Generative Grammar’. Oxford.

Pylkkänen, Lina (2002): Introducing Arguments. PhD thesis, MIT, Cambridge Mass.

Richards, Marc (2008): Defective Agree, Case Alternations, and the Prominence of Person. *In:* M. Richards and A. Malchukov, eds, *Scales*. Vol. 86 of *Linguistische Arbeitsberichte*, Universität Leipzig, Leipzig, pp. 137–161.

Shlonsky, Ur (1997): *Clause Structure and Word Order in Hebrew and Arabic*. Oxford University Press, Oxford.

Silverstein, Michael (1976): Hierarchy of Features and Ergativity. *In:* R. M. W. Dixon, ed., *Grammatical Categories in Australian Languages*. Australian Institutes of Aboriginal Studies, Canberra, pp. 112–171.

Smolensky, Paul (1993): ‘Harmony, Markedness, and Phonological Activity’. University of Colorado.

Smolensky, Paul (1995): ‘On the Internal Structure of the Constraint Component *Con* of UG’. John Hopkins University.

Smolensky, Paul (2006): Harmonic Completeness, Local Constraint Conjunction, and Feature Domain Markedness. *In:* P. Smolensky and G. Legendre, eds, *The Harmonic Mind*. MIT Press, Cambridge/Mass, pp. 27–160.

Stiebels, Barbara (2002): *Typologie des Argumentlinkings: Ökonomie und Expressivität*. Akademie-Verlag, Berlin.

Stiebels, Barbara and Dieter Wunderlich, eds (2000): *Lexicon in Focus*. Akademie-Verlag, Berlin.

Sturgeon, Anne, Boris Harizanov, Maria Polinsky, Ekaterina Kravtchenko, Carlos Gómez Gallo, Lucie Medová and Václav Koula (2011): ‘Revisiting the Person Case Constraint in Czech’. Ms., Harvard University.

Wunderlich, Dieter (2000): ‘Optimal Case in Hindi’. Ms., Universität Düsseldorf.

Three Stages in the Derivation of Free Relatives

Anke Assmann[*]

Abstract

In this paper, I argue for a Comp account of free relatives that is based on the idea that the covert external head of the relative clause is part of the *wh*-phrase. Thereby, the approach derives the close link between the covert head and the *wh*-phrase and is able to account for the puzzling behavior of free relatives with respect to the position of the *wh*-phrase. Concretely, I assume that the derivation of a free relative clause proceeds in three stages: in the first stage, the covert head is created out of the *wh*-item in the lexical array via pre-syntactic copying. In the second stage, the syntactic structure is built, whereby the *wh*-item is merged inside the embedded clause and the covert head is merged as the CP-external head of the relative clause. Finally in the third stage, the additional features created by pre-syntactic copying are deleted post-syntactically.

1. Introduction

Free relatives (FR) – together with parasitic gap and ATB constructions – belong to a class of phenomena where the position of certain items that have undergone movement is not clear. In FR constructions, the item in question is the *wh*-phrase, which appears to stand for the nominal head of the relative clause as well as the relative pronoun. Since it fulfills the function of both an item that is outside the embedded clause and an item that is inside the embedded clause, determining the position of this phrase presents an interesting task for syntacticians. Looking at various properties of FRs leads to a paradox though: the *wh*-phrase seems to be part of the embedded clause as well as the matrix clause.

---

[*] This research is supported by the DFG (Deutsche Forschungsgemeinschaft) within the project "Lokale Modellierung nicht-lokaler Abhängigkeiten in der Syntax (local modeling of non-local dependencies in syntax). For helpful comments and discussion I would like to thank Petr Biskup, Doreen Georgi, Fabian Heck, Timo Klein, Gereon Müller, Martin Salzmann, Philipp Weisser, as well as the participants of the Kolloquium *Neuere Arbeiten zur Grammatiktheorie* at the University of Leipzig and the audience at ConSOLE XXI in Potsdam, January 2013.

Rule Interaction in Grammar, 203–245
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

An example of an FR in German is given in (1).[1]

(1) Ich werde niemandem zeigen was ich gefunden habe.
I will nobody.DAT show what I found have
'I won't show to anybody what I found.' *(Ott 2011: 184)*

(1) shows the two most prominent properties of FRs that distinguish them from normal headed relative clauses: the lack of an external nominal head and the form of the relative pronoun, which appears in the form of a *wh*-phrase. FRs can be found in various languages. In what follows, the data I am concerned with are mainly drawn from German, but the observations made for them hold cross-linguistically.

The aim of this paper is to develop a new approach to FRs that is capable of deriving their puzzling properties, which seem to result from the ambivalent function of the *wh*-phrase. In order to analyze FRs, I will make use of a very traditional means of syntactic analysis: timing of operations. Concretely, I will assume that there are three stages in the derivation that roughly correspond to the traditional notions of the pre-cycle, the cycle, and the post-cycle (cf. McCawley 1970, Postal 1972, Pullum 1979): the lexical array (pre-cycle), the syntactic derivation (the cycle) and a post-syntactic component (post-cycle). Crucially, I will suggest that the copy operation, which applies in the syntax as part of the complex process of movement (Internal Merge), may also apply pre-syntactically in the lexical array.

Under this assumption, a copy of certain features of the *wh*-item, namely what I will call the argument licensing features (categorial features, ϕ-features, case features, etc.) can be created. The copied features remain with the *wh*-item while the original features create a new lexical item. During the derivation the *wh*-item containing the copied features and the new item containing the original features can undergo Merge separately; the *wh*-item is merged in the FR where it serves as the relative pronoun while the new item is merged in the matrix clause as the (covert) head of the FR. Thus, the *wh*-phrase is indeed part of two clauses at the same time which explains its

---

[1] Note that there are two types of FRs in German. One type uses a *wh*-phrase as a relative pronoun as exemplified in (1), while the other type uses a normal *d*-relative pronoun. Fuß and Grewendorf (2012), however, argue that this *d*-type of FRs behaves differently and should therefore not be classified as a genuine FR. In what follows, I will confine myself to the discussion of FRs with *wh*-relative pronouns.

puzzling behavior. Finally, in the post-syntactic component the copied features are deleted under a checking relation with the original features.

This account does not only provide a good basis for the analysis of the properties of FRs discussed below, but also presents a development of the covert head approach of Groos and van Riemsdijk (1981), among others, which assumes that the *wh*-phrase is part of the embedded FR and that the FR is headed by a covert element. The main problem with this approach is, however, that the special link between the covert head and the *wh*-phrase has to be explained (cf. Hirschbühler 1978). Within the present account the tight connection between these two items comes for free since the covert head is actually a part of the *wh*-item.

The paper is structured as follows: In section 2, I show that the status of the *wh*-phrase is indeed ambivalent with respect to its position. Afterwards, in section 3 I develop an account of FRs that forms the basis for the concrete analyses of the properties in section 4. Finally, in section 5 I discuss the consequences of the account with respect to the interaction of operations and compare it to previous approaches to FRs. Section 6 concludes.

2. Data

The aim of this section is to show that the *wh*-phrase seems to be part of two clauses at the same time. This is done by outlining two kinds of data: the first kind suggests that the *wh*-phrase is part of the matrix clause while the second kind suggests that it is part of the embedded clause.

In the data below I have not indicated the exact position of the *wh*-phrase. Instead, the category containing the *wh*-phrase is labeled "FR". This is done for reasons of readability should not suggest whether the *wh*-phrase is the head of this constituent or not.

2.1. The *wh*-Phrase is Part of the Matrix Clause

2.1.1. Number Agreement

The first fact that suggests that the *wh*-phrase is part of the matrix clause involves number agreement. Bresnan and Grimshaw (1978) observed for English FRs that a plural *wh*-phrase in an FR induces plural number agreement in the matrix clause. This observation also holds for other languages, e.g., German.

The German data in (2) show that a plural *wh*-subject induces plural number agreement in the clause where it occurs.

(2) *Agreement: matrix clause*
[Welche Bücher] haben/ *hat dir gefallen?
which book.PL have.PL/ have.SG you liked
'Which books did you like?'

Now, usually plural *wh*-phrases that occur in embedded clauses do not affect the number agreement in the matrix clause. This is exemplified in (3) for indirect interrogative clauses in subject position.

(3) *Agreement: embedded clause*
[$_{\text{SClause}}$ Welche Bücher ihm gefallen], ist/ *sind unklar.
which book.PL him like, be.SG/ be.PL unclear
'It is unclear which books he likes.'

However, in case of FRs, the *wh*-phrase induces plural agreement, in contrast to indirect interrogative clauses. Note that in the sentences in (4) the entire FR is the subject of the matrix clause.

(4) *Agreement: free relative*
?[$_{\text{FR}}$ [$_{\text{Rel}}$ welche Bücher] ich auch immer gelesen habe], haben/
which book.PL I ever read have have.PL/
*hat mir gefallen.
have.SG me liked
'I liked whatever books I read.'

If it is true that number agreement in a matrix clause cannot be induced by a constituent in the embedded clause, the *wh*-phrase in (4) must be part of the matrix clause.

Note that this property of FRs can only be shown with complex *wh*-phrases in German since simple *wh*-phrases always induce singular number agreement. But the use of complex *wh*-phrases in FRs in German is often considered to be marginal and it requires the use of the particle *auch immer* 'ever', the status of which is yet to be clarified. However, the number agreement property of FRs also occurs in languages with simple plural *wh*-phrases, as shown in (5) for Spanish.

(5) [$_{\text{FR}}$ Quienes son del sur] son en gran parte bajos.
who.PL be.PL of.the south be.PL in great part short.PL
'Most people from the South are short.' *(Caponigro 2003: 169)*

Again, the plural *wh*-pronoun *quienes* 'who' induces plural agreement in the matrix clause.

2.1.2. Extraction

The second fact that suggests that the *wh*-phrase is part of the matrix clause comes from extraction data. Rooryck (1994) observed for English that nothing can be extracted out of an FR unless it is part of the *wh*-phrase. This observation also holds for German (see Ott 2011). The extraction property is illustrated for topicalization in German.

In general, topicalization in German may cross a *wh*-island (Fanselow 1987, Müller and Sternefeld 1993), see (6-b).

(6) *Extraction: wh-complement clause*

a. Ich weiß [$_{\text{OClause}}$ welche Bücher *Der Spiegel* diesen Leuten
I know which books *Der Spiegel* these people
empfiehlt]
recommends
'I know which books *Der Spiegel* recommends to these people.'

b. ?Diesen Leuten$_i$ weiß ich [$_{\text{OClause}}$ welche Bücher *Der Spiegel* t_i
these people know I which books *Der Spiegel*
empfiehlt].
recommends
'As for these people, I know which books *Der Spiegel* recommends to them.' *(based on Ott 2011: 188f)*

However, topicalization out of an FR results in strong ungrammaticality if the category is extracted from within the FR, see (7-b). In contrast, if the category is a part of the *wh*-phrase, topicalization is possible, see (7-c).

(7) *Extraction: free relative*

a. Ich lese [$_{\text{FR}}$ welche Bücher von Jostein Gaarder auch immer
I read which books by Jostein Gaarder ever
Der Spiegel diesen Leuten empfiehlt]
Der Spiegel these people recommends
'I read whatever books by Jostein Gaarder *Der Spiegel* recommends to these people.'

b. *Diesen Leuten$_i$ lese ich [$_{\text{FR}}$ welche Bücher von Jostein Gaarder
these people read I which books by Jostein Gaarder
auch immer *Der Spiegel* t_i empfiehlt].
ever *Der Spiegel* recommends
'As for these people, I read whatever books by Jostein Gaarder *Der Spiegel* recommends to them.' *based on (Ott 2011: 188f)*

c. ?Von Jostein Gaarder$_i$ lese ich [$_{\text{FR}}$ [welche Bücher t_i]
by Jostein Gaarder read I welche books
auch immer *Der Spiegel* empfiehlt].
ever *Der Spiegel* recommends
'As for Jostein Gaarder, I read whatever books by him *Der Spiegel* recommends.'

Again, these data suggest that the *wh*-phrase is actually part of the matrix clause, because then, the contrast between (7-b) and (7-c) follows immediately: in (7-b), topicalization is excluded since it involves extraction out of a relative clause island. In (7-c), however, the topicalized PP is extracted out of an object DP in the matrix clause which is possible in German as can be seen from the data in (8).

(8) Von Jostein Gaarder$_i$ habe ich schon [viele Bücher t_i] gelesen.
by Jostein Gaarder have I already many books read
'I have already read many books by Jostein Gaarder.'

Originally, the observation about extraction was made for English by Rooryck (1994: 197). The data are given in (9).

(9) a. I will eat [$_{\text{FR}}$ whatever the chef recommends to that person]
b. *This is the person [$_{\text{RClause}}$ to whom$_i$ I will eat [$_{\text{FR}}$ whatever the chef recommends t_i]]

c. This is the author [$_{\text{RClause}}$ of whom$_i$ I buy [$_{\text{FR}}$ [whatever books t_i]] the NYT recommends to its readers]

Here, the type of extraction is relativization. Movement of a relative pronoun out of an FR is impossible as shown in (9-b). However, movement becomes better if the relative pronoun is extracted from the *wh*-phrase, as shown in (9-c).

To sum up, we have seen that there are indeed facts that indicate that the *wh*-phrase is located outside the embedded clause. The next subsection will deal with data that suggest the opposite, namely that the *wh*-phrase belongs to the embedded clause.

2.2. The *wh*-Phrase is Part of the Embedded Clause

2.2.1. Extraposition

The first data set shows the behavior of FRs with respect to extraposition. The observation made by Groos and van Riemsdijk (1981) is that if the FR is extraposed, the *wh*-phrase is extraposed as well, see (10). This is unexpected under the assumption that the *wh*-phrase is outside the embedded clause (as in the analysis of Bresnan and Grimshaw 1978).

In German, there is a ban on extraposing DPs. (Finite) CPs, on the other hand, are preferably extraposed. The data in (10-b) and (10-c) suggest that the *wh*-phrase *was* 'what' is part of a CP and is not the external D head of the FR. Otherwise, (10-c) should be grammatical and (10-b) ungrammatical.

(10) *Extraposition: free relative*

a. Ich denke, dass ich [$_{\text{FR}}$ was ich mag] essen kann.
I think that I what I like eat can
'I think that I can eat what I like.'

b. Ich denke, dass ich t_{FR} essen kann, [$_{\text{FR}}$ was ich mag].
I think that I eat can what I like
'I think that I can eat what I like.'

c. *Ich denke, dass ich [was t_{FR}] essen kann, [$_{\text{FR}}$ ich mag].
I think that I what eat can I like
'I think that I can eat what I like.'

In (11), it can be seen that indirect interrogative clauses in subject position, where the *wh*-phrase clearly belongs to the embedded CP, show the same behavior as FRs when it comes to extraposition.[2]

(11) *Extraposition: indirect interrogative clause*

a. [$_{\text{SClause}}$ Was ich mag] ist unklar.
what I like is unclear
'It is unclear what I like.'

b. Es ist unklar [$_{\text{SClause}}$ was ich mag].
it is unclear what I like
'It is unclear what I like.'

c. *[$_{\text{SClause}}$ Was t_i] ist unklar [ich mag]$_i$.
what is unclear I like
'It is unclear what I like.'

On the other hand, as shown in (12), if the *wh*-phrase is the external head of a relative clause CP, it cannot be extraposed along with the relative clause.[3]

(12) *Extraposition:* what-*headed relative clause*

a. ?Wer hat [$_{\text{DP}}$ was [$_{\text{RClause}}$ das ich gern gegessen hätte]] weggeworfen?
who has what that I gladly eaten would have thrown away

b. *Wer hat weggeworfen [$_{\text{DP}}$ was [$_{\text{RClause}}$ das ich gern gegessen hätte]]?
who has thrown away what that I gladly eaten would have

c. ?Wer hat [$_{\text{DP}}$ was t_i] weggeworfen [$_{\text{RClause}}$ das ich gern gegessen hätte]$_i$?
who has what thrown away that I gladly eaten would have

---

[2] Note that the clauses in (11) can also be understood as FRs. Under this reading, the matrix predicate *be unclear* would predicate over the things I like. The intended reading in (11) is, however, the reading of the indirect interrogative clause, i.e., the question of what I like cannot be answered clearly.

[3] Note that relative clauses headed by *wh*-phrases are generally considered to be a bit quirky in German. Nevertheless, (12-a) and (12-c) are grammatical and stand in sharp contrast to (12-b).

Thus, this argument strongly suggests that the *wh*-phrase in FR is actually part of the embedded clause.

2.2.2. *Case Matching*

The final facts that argue for a position of the *wh*-phrase in the embedded clause concern a restricted set of possible case mismatches between the embedded and the matrix clause. In general, FRs exhibit a case matching property (Bresnan and Grimshaw 1978, Groos and van Riemsdijk 1981): the *wh*-phrase must bear a case marker that fits the case assigning properties of both the matrix clause and the FR, see (13).

(13) a. Ich folge [$_{\text{FR}}$ wem ich vertraue]
I follow→DAT who.DAT I trust→DAT
'I follow who I trust.' *(Vogel 2001: 902)*
b. *Ich folge [$_{\text{FR}}$ wem/wen ich bewundere]
I follow→DAT who.DAT/who.ACC I adore→ACC
'I follow who I adore.' *(Vogel 2001: 902)*

Only in (13-a) do both the matrix clause and the embedded clause assign dative case to the *wh*-phrase. Hence, (13-a) is grammatical. In (13-b), on the other hand, the *wh*-phrase receives accusative case from the embedded verb *bewundern* 'adore', but it receives dative case from the matrix verb *folgen* 'follow'. This case mismatch cannot be resolved and the sentence is ungrammatical, no matter which case marker the *wh*-phrase actually bears.

However, based on the case hierarchy in (14) (cf. Pittner 1991, 1995, Vogel 2001, Grosu 2003), certain case mismatches are allowed: if the case assigned by the matrix clause is higher on the hierarchy than the case assigned within the FR, the *wh*-phrase may bear the case of the FR, violating the matching condition. Importantly, if a case mismatch is allowed, the case marker on the *wh*-phrase *must* be the case assigned within the embedded clause, see (15-a) and (15-c).

(14) *Case Hierarchy*
NOM ≫ ACC ≫ DAT (≫ GEN)

(15) a. [$_{\text{FR}}$ Wem/*Wer Maria vertraut] wird
who.DAT/who.NOM Maria.NOM trusts→DAT is→NOM
eingeladen
invited
'Who Maria trusts gets invited.' *(Vogel 2001: 903)*

b. *Er zerstört [$_{\text{FR}}$ wer ihm begegnet]
he destroys→ACC who.NOM him meets→NOM
'He destroys who meets him.' *(Vogel 2001: 904)*

c. [$_{\text{FR}}$ Wen/*Wer Maria.NOM mag] wird
who.ACC/who.NOM Maria likes→ACC is→NOM
eingeladen
invited
'Who Maria trusts gets invited.' *(Vogel 2001: 903)*

That the *wh*-phrase must bear the case marker of the embedded clause in case of a mismatch strongly suggests that it is actually part of the embedded clause and not the matrix clause. If the *wh*-phrase was part of the matrix clause, it would remain unclear why it should appear with the morphological case marker of the embedded clause.

Before continuing, a few remarks are in order. First, there is a second exception to the matching condition that concerns morphologically syncretic forms. Thus a sentence like (16-a) is acceptable whereas (16-b) is not.

(16) a. Er tut immer [$_{\text{FR}}$ was mich ärgert].
he does→ACC always what.NOM/ACC me annoys→NOM
'He always does something annoying to me.'

b. *Er liebt [$_{\text{FR}}$ wer mich ärgert].
he loves→ACC who.NOM me annoys→NOM.
'He loves who annoys me.'

(16-a) is grammatical even though the matching condition is violated and the mismatch does not obey the case hierarchy. (16-b), on the other hand, is ungrammatical under the same condition. The reason for the grammaticality of (16-a) is that the morphological form *was* 'what' can mark both nominative and accusative case. In contrast, *wer* 'who' in (16-b) unambiguously encodes nominative case.

Second, for a certain group of speakers, (15-b) is actually acceptable (Pittner 1991, 1995). These speakers seem to have a slightly different case hierarchy, shown in (17).

(17) NOM, ACC ≫ GEN, DAT.

Assuming the hierarchy in (17), the case mismatch in (15-b) should be allowed.[4]

2.3. Interim Summary

So far, we have seen that there are arguments coming from number agreement and extraction that speak in favor of a position of the *wh*-phrase in the matrix clause, where it acts as the external head of the FR. On the other hand, there are also arguments coming from extraposition and case matching that suggest an analysis where the *wh*-phrase is part of the embedded clause.

3. An Analysis of Free Relatives

In this section, I will develop an analysis of free relatives within a minimalist framework that accounts for the data introduced in section 2. The account makes use of the idea that there are three stages in a derivation: the lexical array (a pre-syntactic component), the syntax, and a post-syntactic component. The stages are discussed in more detail in section 5.1.

Concretely, I will assume that the copy operation that is part of Internal Merge may also apply pre-syntactically, thereby creating an additional item in the lexical array. During the syntactic derivation the structure is built. Afterwards, the copied features created pre-syntactically will be deleted in a post-syntactic component.

---

[4] It should be noted that the possibility for case mismatches in FRs has been called into question before. For example, van Riemsdijk (2006: 355) argues that syncretisms of the type shown in (16-a) are the only exception to the matching condition. He claims that other cases of mismatching, such as the grammatical versions of (15-a) and (15-c), can be traced back to the fact that German is in a state where it loses its morphological case system. Thus, speakers do not actually conceive these cases as mismatches but treat them as another instance of (16-a). Still, this explanation does not capture the hierarchy effects found in data such as (15), as it would predict (15-c) to be on a par with (15-b), which does not seem to be the case for many speakers of German.

I will start this section by outlining the main assumptions. Afterwards, I will go through a derivation of an FR and add a few remarks on the semantics of FRs within the present approach.

3.1. Assumptions

The first assumption concerns lexical items (LIs). Following standard assumptions, they are sets of features consisting of syntactic (formal), phonological and semantic features. However, I assume that these features are not part of *one* set but belong to different sets. Furthermore, I assume that there are two kinds of syntactic features that belong to different sets: features involved in argument licensing (categorial features, ϕ-features, case features etc.) and operator features (e.g. *wh*-features). An abstract schema is given in (18).

(18) $\text{LI} = \{\{\text{arg}_1, \text{arg}_2, \text{arg}_3, \ldots\}, \{\text{op}_1, \text{op}_2, \text{op}_3, \ldots\}, \{\text{phon}_1, \text{phon}_2, \text{phon}_3, \ldots\}, \{\text{sem}_1, \text{sem}_2, \text{sem}_3, \ldots\}\}$

The LIs that are relevant for a derivation are gathered in a lexical array (LA). This LA must be empty by the end of the syntactic derivation which is ensured by the application of three operations: Copy, Merge and Agree (Chomsky 1995 et seq.).[5]

Agree is a checking operation that affects features directly: a probe feature that needs a value looks for a matching goal feature that has a value and the two enter into an agreement relation whereby the probe feature is valued by the goal feature (see Chomsky 2000, 2001).

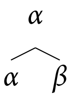
Merge is a set-building operation that acts upon sets of features: two sets α and β become the elements of a new set. The operation is formalized in (19).

(19) $\text{Merge}\ (\alpha,\beta) = \{\alpha,\{\alpha,\beta\}\}$

---

[5] Note that there is an ongoing trend to dismiss the copy operation and only have a simple Merge operation, see Chomsky (2013). The consequence of this assumption would be that there is a complex Internal Merge operation which comprises the copy operation. This would, however, result in a loss of the identity of External Merge and Internal Merge because different Merge operations – one with copying, the other without copying – must be defined. The simpler assumption is therefore to separate Merge and Copy. (Note that assuming Remerge (Epstein et al. 1998) instead of Copy+Merge does not help either: first of all, Remerge provides no possibility to unify Internal and External Merge since it is a separate operation by definition. Furthermore, Remerge is incompatible with the set-theoretic syntactic structure assumed in the present approach.)

For the sake of simplicity, the set structures created in (19) are represented as trees or labeled bracketing.

(20) $\{\alpha,\{\alpha,\beta\}\} = [_{\alpha}\ \alpha\ \beta\ ] =$ 

Following Chomsky (2001 et seq.), there are two possibilities for Merge: *External Merge* and *Internal Merge.* External Merge combines two undominated sets, that is, sets that have no supersets, i.e., lexical items or undominated complex structures. Internal Merge, on the other hand, involves Merge of a subset α of a complex structure β with β. Following Chomsky (1995 et seq.), Internal Merge involves a copy operation.

Copy, cf. (21), is an operation that precedes Internal Merge. (For the independency of Copy and Merge, see also Nunes 1995, 2004 and the discussion in footnote 5.) It creates a copy of a syntactic object. Departing from the standard assumption, I assume that the copy actually replaces the original item and that it is the original item that is merged in a new position.

(21) Copy (α) = < α,α☢>

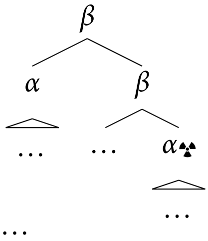
Copies must be deleted by the end of the derivation. Deletion of features applies post-syntactically for reasons of interface (especially PF) interpretability (Chomsky 1995, Nunes 2004).

In order to distinguish copied and original items, the copied item is marked by a diacritic ☢, which is supposed to suggest that this object is unstable and has to be deleted. Deletion of a copy α☢ requires that it is c-commanded by the original item or another copied item. Conrete assumptions about the locality of post-syntactic deletion are presented at the end of section 3.2.

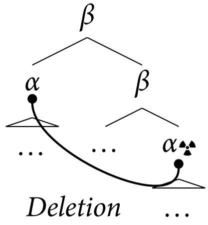
An abstract example of an Internal Merge operation that involves Copy is shown in (22).

(22) Let $\beta = [_{\beta} \ldots \alpha]$.
a. (i) Merge(__,β)
(ii) = Merge(__, [$_{\beta}$ … Copy(α)])
(iii) = Merge(__, [$_{\beta}$ … < α, α☢>])
(iv) = Merge(α, [$_{\beta}$ … α☢])
(v) = [$_{\beta}$ α [$_{\beta}$ … α☢]]

b.



c.



(22-a-i) shows the starting point for Internal Merge. The first argument of the operation Merge that involves a structure β still needs to be created out of β. Thus, Copy must be invoked and applies to the constituent α in (22-a-ii). In (22-a-iii), Copy has created an ordered pair consisting of the original α and the copied α. Now, in (22-a-iv) and (22-a-v), the original α can be the target for Merge while the copied α remains inside β. (22-b) shows a tree diagram of the labeled bracketing in (22-a-v). Afterwards in a post-syntactic component, the unstable occurrence of α must be deleted. This is illustrated in (22-c).

One very important assumption I make here is that the possibility for Copy is given at any time, also *before* the derivation actually starts, namely pre-syntactically in the lexical array.[6] Again, Copy may affect only sets that have supersets, i.e., parts of LIs may be copied pre-syntactically.

---

[6] Whether Merge and Agree may also operate pre-syntactically will not be discussed in what follows. Note, however, that the possibility of pre-syntactic operations is an interesting advancement to standard minimalism and facilitates the analysis of different phenomena. See, e.g., Heck (2010) for an analysis of the ban on direct recursion that is based on the possibility of pre-syntactic Agree; Assmann et al. (2012*a*) for an analysis of possessor advancement that builds upon the idea of pre-syntactic Agree; Georgi (2012) for a pre-syntactic deletion operation that is involved in analyzing global case splits; and Assmann (2012*b*) as well as Agbayani and Ochi (2007) for an approach to parasitic gaps that makes use of a pre-syntactic fission/splitting operation. See also Thomas (2013) for an approach to split ergativity that is based on pre-syntactic deletion of case features.

After an LI has entered the derivation, its internal structure is no longer accessible to Copy, i.e, lexical integrity is preserved. (This assumption might be dismissed, if theories like Chomsky 1995, Agbayani 1998, Brosziewski 2003 turn out to be correct, which assume that parts of lexical items can be subject to Internal Merge, and hence Copy, also in the syntax.)

In most cases, pre-syntactic Copy of parts of LIs will lead to a crash of the derivation since there is no position available in the structure where the additional items can be merged. This is due to syntactic constraints like e.g. the Θ-Criterion (Chomsky 1981), the case filter (Rouveret and Vergnaud 1980, Chomsky 1980) or the assumption that Merge is feature-driven (e.g. Müller 2011). Hence, the additional items will remain in the LA, which causes a violation of the constraint that the LA has to be empty. However, if the LA lacks an item to begin with, Copy may create the missing item out of an existing one.

Assuming a structure of LIs as in (18), syntactic features do not form a constituent either with phonological or semantic features. Hence Copy may only affect the syntactic features of a lexical item. (If only phonological or only semantic features are copied, an element is created that is not viable in the syntax, since it has no syntactic features. Note however that Copy of phonological and/or semantic features might be of use for deriving certain PF and LF phenomena.[7])

3.2. Syntax of Free Relatives

With these assumptions in place, we can now start to derive FRs. Another example of an FR in German is given in (23).

(23) dass alle [$_{\text{FR}}$ was ich tue] gut finden
that everyone what I do good find
'that everyone likes what I do'

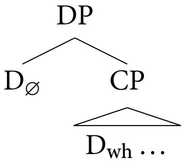
Basically, the analysis that follows is a development of the analysis first proposed by Groos and van Riemsdijk (1981). (Other versions of this analysis can be found in Grosu 1996, 2003, Citko 2004.) In this type of analysis, FRs are as-

---

[7] Furthermore, there have been proposals that Copy of syntactic features may apply post-syntactically. In order to analyze complementizer agreement, Fuß (2008) suggests that the morpheme bearing the subject agreement features is copied post-syntactically. The copy is then inserted on C which ensures that both the agreeing complementizer and the agreeing verb realize the subject's ϕ-features.

sumed to have a structure as in (24), where a CP is adjoined to a D element.[8] This guarantees that the entire category is a DP which can be merged in an argument position.[9]

(24)



The main question here concerns the covert D head. It is often assumed that this head comes directly from the lexicon. Furthermore, the data presented in section 2 suggest a special, very close, link between the covert D head and the *wh*-phrase, a link not found in headed relative clauses. This tight relation between the *wh*-phrase and the covert head basically leaves two possibilities within this type of approach: first, one could assume that the covert head is massively underspecified and needs to Agree with the *wh*-phrase in all its features (even the categorial feature, see footnote 9). This begs the question as to why the lexicon should provide such a massively underspecified covert head to begin with (see also the discussion in section 5.2.2.1).

Therefore a second possibility is pursued here: the covert head is a copy of the *wh*-phrase, more exactly, of a part of it.

The main assumption that the derivation of FRs is based on is that the lexical array (LA) of a sentence containing an FR contains only one *wh*-phrase. This *wh*-phrase can only be merged either inside the FR, where it fills an argument position and satisfies the *wh*-feature of the embedded C, or in the matrix clause, where it fills an argument position.

---

[8] Assuming bare phrase structure (Chomsky 1995), the D element is a head and a phrase at the same time.

[9] Note that there are also FRs of other categories, e.g., AP-FRs (cf. mainly Grosu 1996, 2003). A German example is given in (i).

(i) Ich zahle [$_{FR}$ wie viel du bezahlst hast].
I pay how much you have paid.

In the present account, the external head of the FR results from pre-syntactic Copy that applies to the *wh*-phrase. Essentially, the categorial feature of the *wh*-phrase is also copied. Thus, the external head of an AP-FR is expected to be of category A, just like the *wh*-phrase.

(25) *Working hypothesis*[10]
The lexical array of a sentence containing an FR is deficient and does not provide enough LIs to guarantee a converging derivation.

Assuming that one DP can only fill one argument position (presumably for case and theta-role reasons), one possibility to ensure that both clauses have enough arguments is by Copy applying in the LA. In this way an additional item is created that may be merged in another argument position. For reasons outlined below, Copy has to affect the *wh*-item. The feature structure of the *wh*-item *was* ('what') is given in (26).

(26) *was* = {{D,ϕ:3sg,case:__},{wh}, PHON, SEM}

An important point about the structure in (26) is that certain features, namely the argument licensing features {D, ϕ, case, …} form a constituent to the exclusion of the operator features, e.g., the *wh*-feature. The intuition behind this structure is that there are certain features that are necessary for a constituent in order to fill an argument position, e.g., the categorial feature to satisfy the selectional needs of the verbal projection, ϕ-features to satisfy the needs of the functional heads v and T, and case features to pass the case filter (Rouveret and Vergnaud 1980, Chomsky 1980). The *wh*-feature, on the other hand, is an operator feature that is not needed to fill an argument position but rather to satisfy the needs of the C head of the clause. In that sense it does not classify as an argument licensing feature and is therefore part of a different feature set.

By definition, Copy can only affect subsets. Now, given the structure in (26), there are various possibilities as to how Copy can apply to *was*. There is, however, only one possibility that will lead to a converging derivation.

The sets comprising the semantic and the phonological features cannot be copied for reasons outlined above in section 3.1: copying would create an LI that is not viable in the syntax since it does not have any syntactic features.

The next option consists of copying the set containing the operator features. In this case, the item created would not have features that allow it to be merged in an argument position since it would lack the necessary argument licensing features.

---

[10] See also Nunes (1995, 2004) for the idea that the structure of the lexical array is not subject to any constraints (as originally proposed in Chomsky 1995). Rather lexical arrays are more or less arbitrary collections of lexical items, which gives rise to the possibility of deficient lexical arrays.

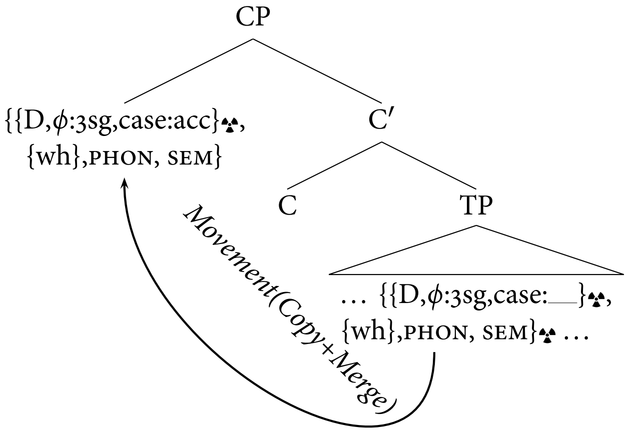
The only option that leads to a converging derivation is therefore to copy only the set containing the argument licensing features.

Thus the operation applying in the first stage of the derivation – the pre-syntactic component – is Copy of the argument licensing features of the *wh*-item. This is illustrated in (27).

(27) *Pre-syntactic Copy*
  {Copy({D,ϕ:3sg,case:__}), {wh}, PHON, SEM}
  = {<{D,ϕ:3sg,case:__},{D,ϕ:3sg,case:__}☢>, {wh}, PHON, SEM}
  = {D,ϕ:3sg,case:__}, {{D,ϕ:3sg,case:__}☢, {wh}, PHON, SEM}

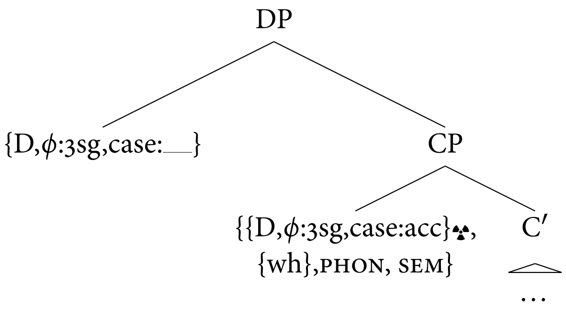
When the syntactic derivation starts, the embedded clause is built first. The *was*-item that contains the *wh*-feature must be merged in this clause in order for C to check its *wh*-feature. In its base position, *was* receives accusative case from the embedded v. Movement of *was* to Spec-C leaves a copy behind that must be deleted later on.

(28)



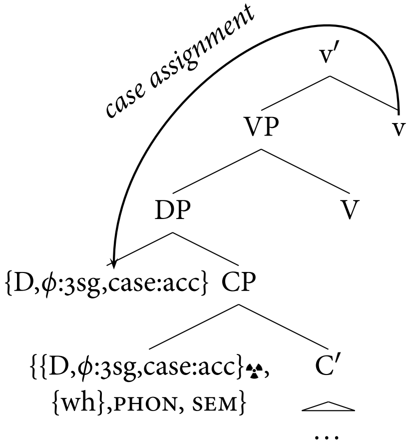
The set {D, ϕ, case} created by pre-syntactic Copy can now be merged with this CP, obtaining the structure in (24), see (29). Note that the D head does not contain any phonological features and is thus covert.

(29)



Afterwards, the entire DP can be merged as an argument of the matrix clause and the covert D head receives case by the matrix v, see (30).

(30)

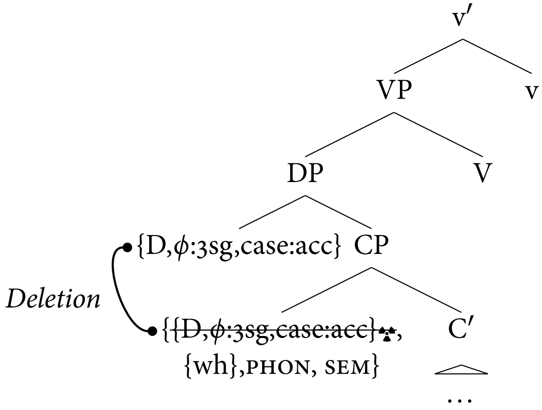


Finally, after the syntactic derivation is terminated and the covert head of the FR has received case, post-syntactic deletion of the copied feature set on the *wh*-phrase applies under c-command and identity with the set containing the original features.[11]

---

[11] In the context of FRs, it becomes clear why deletion of copies cannot apply in the syntax already: assuming that the copied feature set on the *wh*-phrase can only be deleted if all the features on the original feature set are valued, case assignment must precede deletion. If deletion was a syntactic operation, its application in the configuration in (30) would be counter-cyclic because deletion would affect only the DP-cycle although the DP is already dominated by another cycle. Assuming that Strict Cyclicity (Chomsky 1973) holds for all syntactic operations, deletion of copies must then apply post-syntactically.

(31)



Up to now, we have achieved the following result: pre-syntactic Copy of the argument licensing features of the *wh*-item creates a new LI in the lexical array that is needed to enable a converging syntactic derivation. Post-syntactically, the copied features are deleted again.

The account of FRs proposed here can provide answers to the three main questions raised by this construction. First, the question why a clause can occur in a position reserved for non-clausal arguments is answered by assuming the structure in (24) where the clause is the complement of a covert D head. The entire structure is a DP and can therefore occur in DP positions.

Second, the head of the FR must be covert because it is an LI created by pre-syntactic Copy. Since the phonological features cannot undergo pre-syntactic Copy for theory-internal reasons, the new LI, which serves as the external head later on, is covert.

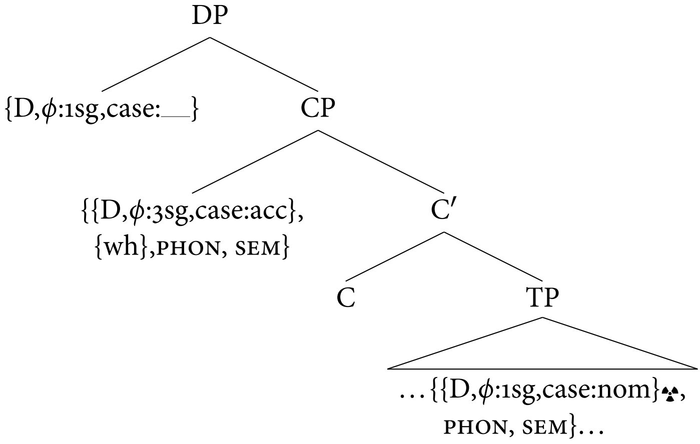
Third, the close link between the *wh*-phrase and the covert head arises from the fact, that the covert head is made up of features from the *wh*-head. Thus, the features and their values have to coincide.

Before I proceed with the analysis of the properties of FRs, one question remains to be answered: why must pre-syntactic Copy affect the *wh*-phrase? The reasons are the following: First, since it is a D item that is actually missing, Copy of any other category but D would not bring the required result. Thus, we have already narrowed down the possibilities for pre-syntactic Copy to the D items in the lexical array.

Furthermore, Copy of some D item that appears in the matrix clause would also lead to a crashing derivation because in this case, the copied features would be higher in the structure than the original features, i.e., exactly the opposite of the structures above. Under the assumption that the original features have to c-command the copied features in order for deletion to apply, the derivation would crash due to undeleted copied features.

Now, there are not many D elements left. In a transitive FR-clause like (23) it boils down to two D items: *was* 'what' and the pronoun *ich* 'I'.[12] Now, what would exclude a structure as in (32) where Copy applied to *ich*?

(32)



A way to exclude (32) would be to assume that spell-out applies cyclically and that copies are tolerated only for a short time in the post-syntactic component. Concretely, (32) can be ruled out as follows. Assume, following Chomsky (2000, 2001), Heck and Zimmermann (2004), Svenonius (2004), that phases are vP, CP and DP. As soon as a phase is built, its complement is sent to Transfer. In order to escape spell-out, categories must move to the edge of the current phase. These assumptions ensure that at most one phase boundary may intervene between a copy and its antecedent (another copy or the original category). In other words, the antecedent of a copy enters the post-syntactic component always on the next spell-out cycle. Assuming further, that this is a necessary condition for deletion to apply, the structure in (32) can be excluded. In (32), the subject inside the TP contains a copied feature set. Deletion of this copy may only apply if the CP contains a potential antecedent for it. This, however, is not the case in (32). Put differently, the problem with (32) is that

---

[12] For the sake of concreteness I assume that pronouns are simple D heads.

it looks as if movement would have skipped the CP phase (even though no actual syntactic movement took place).[13]

Having excluded all other options, the only converging derivation results from Copy of the *wh*-item that moves to Spec-C of the embedded clause for independent reasons. In sum, the options for pre-syntactic Copy are highly limited by the conditions for converging derivations.

In the next part of this section, I will show that the syntactic structure created so far is compatible with a standard semantics for free relatives.

3.3. Semantics of Free Relatives

The syntactic structure of FRs derived above is compatible with standard semantic analyses of FRs. Following Caponigro (2003), Jacobson (1995) (cf. also Grosu and Landman 1998, Grosu 2003), I assume that FRs are semantically like DPs in that both denote the maximal entity (Link 1983) described by a predicate.

Furthermore, I assume that the semantic type of a lexical item depends on its features: *wh*-phrases, e.g., have a *wh*-feature and are therefore of type <<e,t>,<e,t>> (Caponigro 2003).[14]

The λ-representation of the *wh*-item *was* is given in (33) (cf. Caponigro 2003).

(33) {{D,ϕ:3sg,case:___}☢,{wh},PHON, SEM}
<<e,t>,<e,t>>: λPλx[−anim′(x) ∧ P(x)]

Following Caponigro (2003), questions are predicate abstracts. That is, the *wh*-item *was* "applies to a set of entities to give back all and only the entities of that set that are inanimate" (Caponigro 2003: 58).

---

[13] Of course, this situation can only arise if Copy has applied pre-syntactically. Then, the copy and the original object can enter the derivation at any point and no locality conditions can be imposed on them in the syntax, in contrast to applications of Copy that precede syntactic movement.

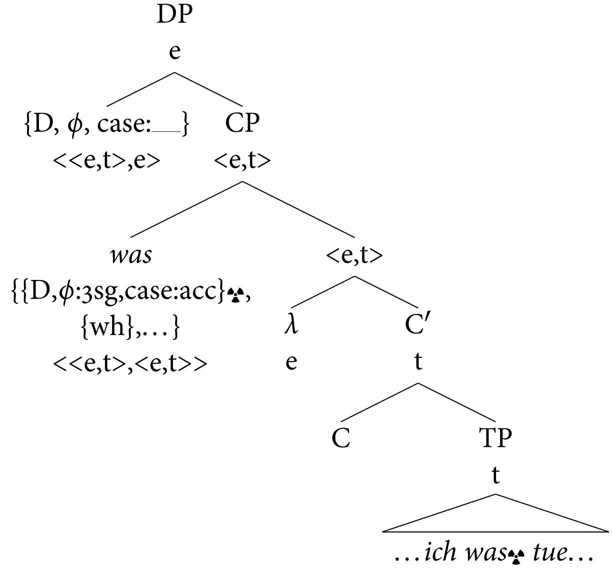
[14] At this point, the question arises as to what extent the semantics of a lexical item depends on its semantic features and what semantic features actually are. Following Chomsky (1995: 230), semantic features are to be understood as descriptive features that constitute the result of a semantic decomposition of the meaning of the lexical item. (Chomsky gives the example *airplane* that has a semantic feature [artifact].) Thus, only non-functional elements have true semantic features. The semantic type and the formal parts of the λ-representation of an LI, on the other hand, depend on the formal, i.e, the syntactic features.

The semantic operator σ that returns the maximal entity of this set is of type <<e,t>,e>, i.e., basically the same type as a simple D head. Such a simple D head is indeed available in the theory above due to Copy: it is the head of the DP dominating the embedded CP. This head has the feature specification of a simple D head, a semantics for which is given in (34).

(34) {D,ϕ:3sg,case:__}
<<e,t>,e>:$\lambda P \sigma x[P(x)]$

In (34), the σ-operator takes a set and returns its maximal entity. With these assumptions in mind, the meaning of FRs can be computed compositionally on the basis of the structure in (35).

(35)



The crucial point is what happens in the CP and the DP. Movement of *was* to Spec-C has created a λ-abstract $\lambda x_1[do'(x_1)(\textsc{speaker})]$. Now, the *wh*-item *was* applies to the set described by this λ-abstract and we end up with a form $\lambda x_1[-anim'(x_1) \wedge do'(x_1)(\textsc{speaker})]$. Finally the D head applies to this set and returns its maximal entity: $\sigma x_1[\lambda x_1[-anim'(x_1) \wedge do'(x_1)(\textsc{speaker})]]$.

As can be seen, the analysis of Caponigro (2003) is fully compatible with the syntactic structure of FRs as proposed here. Furthermore, the emergence of the σ-operator – a point not explained in Caponigro (2003) – is captured: it is the result of pre-syntactic Copy.[15]

4. Deriving the Behavior of the *wh*-Phrase

In this section, I will show how the properties of FRs introduced in section 2 can be analyzed within the account outlined in section 3.

4.1. The *wh*-Phrase is Part of the Matrix Clause

4.1.1. Number Agreement

The first property described in section 2.1.1 concerned number agreement. The observation was that a plural *wh*-phrase in an FR induces plural number agreement in the matrix clause. Within the present account the data can be derived as follows. Let's first have a look at a simple *wh*-phrase, as illustrated by the Spanish data in (36), repeated from (5).

(36)	[FR Quienes	son	del	sur	]	son	en	gran	parte	bajos.
	who.PL	be.PL	of.the	south		be.PL	in	great	part	short.PL

'Most people from the South are short.'

A plural number feature is part of the ϕ-features of *quienes* 'who'. Hence, after pre-syntactic Copy there are two plural number features, one on the item *quienes* and one on the newly created D head. The pre-syntactic Copy operation is shown in (37).

(37) a. *quienes* = {{D,ϕ:3pl,…}, { wh}, PHON, SEM}
b. {Copy({D,ϕ:3pl,…}), {wh}, PHON, SEM}
= {D,ϕ:3pl,…}, {{D,ϕ:3pl,…}☢, {wh}, PHON, SEM}

Thus, plural number agreement is supposed to be possible in the FR as well as in the matrix clause. The Agree relations are illustrated in (38).

---

[15] Note that the analysis of Caponigro (2003) was assumed here for the sake of concreteness. Since the resulting syntactic structure of FRs is quite a standard, very widespread one, any other semantic analysis that is based upon the structure in (24) should be compatible with the syntactic analysis of the present approach.

(38) [$_{\text{vP}}$ [$_{\text{DP}}$ {ϕ:3pl,…} [$_{\text{CP}}$ {{ϕ:3pl,…}☢,{wh}, …} …son…]] …son…]

number agreement

Things are more complicated with the complex *wh*-phrase in German in (39), repeated from (4).

(39) ?[$_{\text{FR}}$ [$_{\text{Rel}}$ welche Bücher] ich auch immer gelesen habe], haben mir
which book.PL I ever read have have.PL me
gefallen.
liked
'I liked whatever books I read.'

In general, there are two possibilities to extend the analysis in order to derive the German data. First, there are two forms of *welch* ('which'): a singular form selecting for a singular NP and a plural form selecting for a plural NP. Under these assumptions, the same analysis as sketched in (38) can account for (39).

The second, perhaps more attractive, possibility is that *welch* agrees with the NP in its number feature. Then, both the original and the copied number feature must agree with the number feature of the NP which is merged in the FR. The analysis is sketched in (40).

(40) [$_{\text{DP}}$ {D,ϕ:3pl,…} [$_{\text{CP}}$ [$_{\text{DP}}$ {{D,ϕ:3pl,…}☢,…} [$_{\text{NP}}$ Bücher$_{pl}$]]]]

number agreement

4.1.2. Extraction

The second set of data involved extraction. It was shown in section 2.1.2 that topicalization out of an FR is impossible if the category is a part of the FR. On the other hand, if the category is a part of the *wh*-phrase, topicalization is possible. The crucial data are repeated in (41).

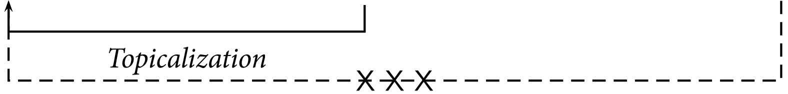
(41) a. *Diesen Leuten$_i$ lese ich [$_{\text{FR}}$ welche Bücher von Jostein Gaarder
these people read I which books by Jostein Gaarder
auch immer *Der Spiegel* t_i empfiehlt].
ever *Der Spiegel* recommends
'As for these people, I read whatever books by Jostein Gaarder *Der Spiegel* recommends to them.'

b. ?Von Jostein Gaarder$_i$ lese ich [$_{\text{FR}}$ [welche Bücher t_i]
by Jostein Gaarder read I welche books
auch immer *Der Spiegel* diesen Leuten empfiehlt].
ever *Der Spiegel* these people recommends
'As for Jostein Gaarder, I read whatever books by him *Der Spiegel* recommends to these people.'

Now, whatever prohibits extraction out of a relative clause, rules out extraction from FRs since in this theory, they are relative clauses. Hence, (41-a) is expected to be ungrammatical.

In those cases where extraction out of an FR seems to be possible, I claim that extraction actually proceeds from outside the FR. The PP complement can choose to be merged either to the covert head outside the FR or inside the overt *wh*-phrase. If it is merged outside the FR, it can be topicalized.[16]

(42) …[$_{\text{DP}}$ [$_{\text{DP}}$ {D,ϕ:3pl,…} [$_{\text{PP}}$ von J.G.]] [$_{\text{CP}}$ {{D ϕ:3pl☢,…},…},…}.. α…]]



A technical way to capture this idea would be to assume that the feature selecting the PP is not an intrinsic feature but an optional feature that is added to the item in the lexical array.[17] Then, a feature [uP] that is responsible for taking a PP complement can end up either on an item to be merged in the *wh*-phrase inside the FR or it can end up on the newly created D head.[18]

---

[16] Note that if the PP is merged with the covert head, the relative clause CP cannot be directly merged with the D head. Hence, the D head does not c-command the *wh*-phrase, which is a necessary condition for post-syntactic deletion to apply. It thus seems necessary to relax the conditions for deletion and assume that m-command is sufficient. As far as I can see, this assumption does not lead to any false predictions.

[17] This is a slight divergence from Chomsky (1995) who claims that optional features are "added as the LI enters the numeration"[p.231].

[18] That D heads may take PP complements is independently verified by sentences such as (i) where the overt D head *das* is combined with a *von*-PP.

(i) Ich habe [$_{\text{DP}}$ das [$_{\text{PP}}$ von ihm]] gelesen.
I have that by him read.
'I have read that one by him.'

4.2. The *wh*-Phrase is Part of the Embedded Clause

4.2.1. Extraposition

The extraposition data in section 2.2.1 showed that if the FR is extraposed, the *wh*-phrase is extraposed as well (Groos and van Riemsdijk 1981), see (43).

(43) a. Ich denke, dass ich t_{FR} essen kann, [$_{\mathrm{FR}}$ was ich mag].
I think that I eat can what I like
'I think that I can eat what I like.'
b. *Ich denke, dass ich [was t_{FR}] essen kann, [$_{\mathrm{FR}}$ ich mag].
I think that I what eat can I like
'I think that I can eat what I like.'

In the present account, the grammaticality of (43-a) and the ungrammaticality of (43-b) are derived because of the restrictions for pre-syntactic Copy of phonological features. Remember that only syntactic features may undergo pre-syntactic Copy. The result of this restriction is that the newly created D head does not possess any phonological features. The phonological features are only available on the *wh*-item *was*, that is, the phonological features are part of the embedded CP. This in turn means that the *wh*-phrase must be pronounced within the CP, i.e., if the CP is extraposed, the *wh*-phrase must be extraposed as well.[19]

4.2.2. Case Matching

Finally, it was shown in section 2.2.2 that certain case mismatches with FRs are allowed: if the case assigned by the matrix clause is higher on the case hierarchy than the case of the FR, the *wh*-phrase may bear the case of the FR, violating the matching condition. The case hierarchy is repeated in (44).

---

[19] The fact that FRs can be extraposed seems to be incompatible with the condition for post-syntactic feature deletion as outlined in section 3.1 and 3.2. In the extraposed position the unstable copied feature of the *wh*-phrase are not c-commanded by the covert D head. Then, deletion should not be possible, which leads to a crash of the derivation. It thus seems necessary that feature deletion applies before extraposition. This could be indeed the case if extraposition is a post-syntactic operation itself (cf. Chomsky 1995, Truckenbrodt 1995, Göbbel 2007, Wurmbrand 2007, Chomsky 2008; see Inaba 2005 for German), which would also explain why it does not behave like typical syntactic movement (see e.g. Fox and Nissenbaum 1999).

(44) *Case Hierarchy*
NOM ≫ ACC ≫ DAT (≫ GEN)

These data can be analyzed within the present account of FRs under the following assumptions: First, case features are rather case slots to which case features are added (cf. also Assmann et al. 2013, where the same assumption for analyzing the case split on possessors in Udmurt is made).

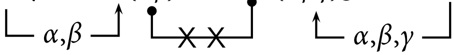
Second, cases are decomposed in a way that a case higher on the case hierarchy is a superset of a case lower on the hierarchy (cf. Béjar and Řezáč 2009; for similar ideas, see Trommer 2006, 2008). The case decomposition is given in (45).[20]

(45) *Case decomposition*
NOM $[\alpha,\beta,\gamma] \supset$ ACC $[\alpha,\beta] \supset$ DAT $[\alpha]$

Furthermore, the identity condition of deletion will be revised slightly: in order for a copied feature set to delete, it must be a subset of the original feature set, that is, all features in the copied feature set must have a matching feature in the original feature set.

Now, if the case assigned in the FR is higher on the hierarchy, i.e., if it consists of more features than the case assigned in the matrix clause, the copied feature set contains more case features than the original feature set. In this case deletion cannot apply, as shown in (46).

(46) *Er zerstört [$_{\text{DP}}$ D [$_{\text{CP}}$ wer ihm begegnet]]
he destroys Ø$_{\{\alpha,\beta\}}$ who$_{\{\alpha,\beta,\gamma\}}$☢ him meets



In (46), the *wh*-item *wer* receives nominative from the functional head T in the embedded clause. Thus the copied case slot contains the features $[\alpha,\beta,\gamma]$☢. The original case slot on the external D head, on the other hand, receives accusative, i.e. $[\alpha,\beta]$. Then, however, there are features in the copied feature set on the *wh*-phrase for which there are no matching features on the original feature set. Thus, deletion cannot apply, which leads to a crash of the derivation.

---

[20] Note that the decomposed case features are abstract in (45). In what follows, I will refrain from exchanging the abstract case features by concrete features since it is not important for the idea of the analysis.

In contrast, if the case assigned in the FR is lower on the hierarchy, i.e., consists of fewer features than the case assigned in the matrix clause (or if both cases are identical), all case features on the copy inside the FR have a matching feature on the original and deletion can apply, see (47) and (48).

(47) Ich folge [$_{\text{DP}}$ D [$_{\text{CP}}$ wem ich vertraue]]
I follow $\varnothing_{\{\alpha\}}$ who$_{\cancel{\{\alpha\}}}$☢ I trust
α α

(48) [DP D [$_{\text{CP}}$ Wen/ *Wer Maria mag]] wird eingeladen
$\varnothing_{\{\alpha,\beta,\gamma\}}$ who$_{\cancel{\{\alpha,\beta\}}}$☢/ $_{\{\alpha,\beta,\gamma\}}$☢ Maria like is invited
α,β
α,β,γ

The trivial case is illustrated in (47). Here the case features assigned within the embedded clause and the case features assigned within the matrix clause are identical. Thus the original and the copied case bear an identical feature set. Consequently, deletion of the copied feature set is possible.

In (48), the copied case slot in the embedded clause bears fewer features than the original case slot on the external D head. More concretely, in (48), the *wh*-phrase *wen* bears the case features [α,β]☢ while the covert external D head bears the case features [α,β,γ]. Since both [α] and [β] find a matching feature on the D head, deletion of the copied feature set is possible and the derivation will converge. The case hierarchy involved in the mismatches is therefore correctly derived.[21]

---

[21] The analysis developed so far has nothing to say about the syncretism exception to the hierarchy effects, illustrated in (16) in section 2.2.2. A possible extension of the account that could derive the facts would consist in assuming that the morphological operation Impoverishment can apply in the syntax already (Keine 2010, Doliana 2013). The relevant rule with respect to FRs would be that in the context of inanimate nominative *was*, the nominative feature [γ] deletes and *was* bears only the accusative features [α,β].

(i) [γ] → $\varnothing$/__[+wh–anim]

Consequently, in a context like (16-a), repeated below in (ii), [γ] deletes (marked as ($\cancel{\gamma}$) in (ii)) and a fatal case mismatch in FRs is circumvented, since the covert head and overt *was* possess an identical case feature set.

(ii) Er tut immer [$_{\text{DP}}$ D [$_{\text{CP}}$ was mich ärgert]].
he does always $\varnothing_{\{\alpha,\beta\}}$ what$_{\cancel{\{\alpha,\beta,(\cancel{\gamma})\}}}$☢ me annoys
α,β α,β,γ

Finally, note that the *wh*-phrase must always be pronounced with the case of the FR. Since the phonological features are a part of the *wh*-phrase in the embedded CP, they must realize the case assigned in this clause. Hence, the ungrammatical version of (48) is correctly excluded.

4.3. Interim Conclusion

Free relatives are puzzling since their relative pronouns (*wh*-phrases) seem to be simultaneously part of two sentences. The special behaviour of the *wh*-phrase is derived by making use of timing of operations: a part of the *wh*-phrase, namely the part containing the argument licensing features, is copied pre-syntactically in the lexical array. Then, the *wh*-phrase containing only a copy of its argument licensing features is merged inside a relative clause CP, where it is moved to Spec-C. The original argument licensing features are merged as the external head of the relative clause. The copied features on the *wh*-phrase must be deleted post-syntactically under c-command and feature identity.

Under these assumptions the puzzle concerning the ambivalent status of the *wh*-phrase is resolved: due to pre-syntactic Copy, certain features of the *wh*-phrase are available not only in the embedded clause but also in the matrix clause, namely exactly those features that are involved in the properties that suggest that the *wh*-phrase is located outside the FR, which are, e.g., ϕ-features and subcategorization features. However, all the other features of the *wh*-phrase are only available inside the FR, e.g., phonological features. Exactly those properties which suggest that the *wh*-phrase is inside the FR can be traced back to these features.

The restricted occurrences of case mismatches were derived by decomposing the case features in a way that models the case hierarchy: a case x which is lower on the case hierarchy than a case y is represented by a subset of the case features of y. Under this assumption, the hierarchy effects follow because the case features of the copied case slot must be in a subset relation $\subseteq$ with the features of the original case slot for deletion to apply, which is only given if the case assigned by the matrix clause is higher than or equal to the case assigned in the embedded clause.

---

5. Discussion

5.1. The Three Stages in More Detail

The main clue of the present account of FRs is that it distributes the operations which are necessary for the derivation to converge among three different components; thus, the derivation proceeds in three different stages. The first stage subsumes operations that apply in the lexical array, before the actual syntactic derivation starts. The second stage comprises the operations in the syntactic derivation. Finally, in the third stage, post-syntactic operations apply.

Due to this division of labor into three components which are strictly ordered, the operations that apply in these components are ordered as well. The order of the operations that are relevant for the present theory of FRs is given in (49).

(49)

Pre-syntactic	*syntactic*	*post-syntactic*
Copy	Copy, Merge, Agree	Deletion (of Copies)

The order of the application of these operations that result from the distribution among different components leads to certain patterns of interaction. The main schema is the following (cf. Pullum 1979): operations that apply earlier in the derivation can affect the application of operations that apply later (feeding, bleeding, cf. Kiparsky 1971, 1976), but not vice versa (counter-bleeding, counterfeeding, ibid.).

In terms of the present approach this means that pre-syntactic Copy can bleed or feed operations in the syntax (Agree, Merge), and operations in the syntax can feed or bleed post-syntactic deletion. On the other hand, operations that apply later cannot affect operations that apply earlier. This leads to potential instances of counter-bleeding and counter-feeding. In what follows, I will summarize the most interesting interactions of operations that arise from the present theory.

Let us first have a look at the transparent interactions. The most obvious example for such an interaction concerns the effects of pre-syntactic applications of Copy. Since pre-syntactic Copy creates a new lexical item, it feeds Merge of this new lexical item as the covert head of the relative clause as well as Agree relations that involve this covert D head (e.g. number agreement in the matrix clause, cf. section 2.1.1 and 4.1.1). Furthermore – indeed a more trivial interaction – pre-syntactic Copy, which creates unstable features, feeds post-syntactic deletion of these features.

But there are also opaque interactions of operations in the present account of FRs. A very interesting kind of counter-feeding interaction involves the case matching property of FRs. In section 2.2.2 it was shown that certain mismatches are allowed while others are not. Leaving the decomposition of case features as pursued in section 4.2.2 aside for a moment, we have an instance of counter-feeding: usually, valuation of the case feature on the covert head feeds deletion of the case features (by assumption unvalued features cannot be deleted; cf. section 3.1). However, in contexts where a case mismatch is not allowed (e.g., (13-b)) this feeding relation is not given; hence, we have a counter-feeding relation. The special nature of this counter-feeding relation is that it cannot be resolved by re-ordering the operations because syntactic case assignment must in any case precede post-syntactic deletion. Instead, counter-feeding is resolved by decomposing the case values as in section 4.2.2 where a case higher on the hierarchy in (14) consists of more features than a lower case. If the *wh*-phrase in the embedded CP receives a high case, while the covert head receives a low case, the unstable copy of the argument licensing features on the *wh*-phrase cannot be deleted since there are not enough matching case features in the original feature set on the covert head.

This survey of interaction types ends the discussion of the three stages in the derivation of free relatives. The discussion has shown that the order of operations arises from a standard architecture of grammar, where the pre-syntactic component precedes the syntactic one, which in turn precedes the post-syntactic component. Since this ordering of grammatical components is strict, the order of operations is strict as well. Thus, re-ordering is not an option for solving all cases of opaque interactions. Instead, reanalyzing the data in a way that a counter-feeding relation does not arise in the first place is in line with the present account of FRs.

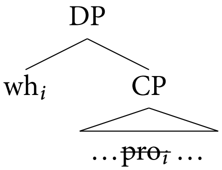
5.2. Previous Analyses

In this section I will summarize previous approaches to FRs. All of them are deficient in the sense that they cannot derive some of the properties of FRs.

5.2.1. *External-Head Accounts*

I start this overview with a discussion of one of the oldest formal approaches to free relatives. Due to facts like the ones presented in section 2.1, Bresnan and Grimshaw (1978) claimed that the *wh*-phrase is the external head of a relative clause which contains a pronoun that is bound by the *wh*-phrase and undergoes a process of *Controlled Pro-Deletion*. The rough structure is depicted in (50).[22]

(50)



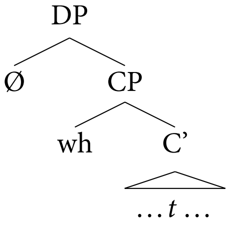
The main problem with this approach is certainly its conflict with the extraposition data from languages like German, as first pointed out by Groos and van Riemsdijk (1981) and discussed in section 2.2.1 and 4.2.1. Within these approaches, it must be possible to extrapose DPs in order to account for the data. However, DP extraposition is not attested otherwise in German. Thus, the approach cannot derive the empirical facts correctly.

5.2.2. *Comp Accounts*

5.2.2.1. *Groos and van Riemsdijk (1981)*

In order to derive the extraposition data, Groos and van Riemsdijk (1981) propose that the *wh*-phrase is inside a relative clause that is headed by an empty category.[23] The analysis is sketched in (51).

(51)



---

[22] See also Bresnan (1973), Daalder (1977), Larson (1987) for other head-external approaches.

[23] The earliest version of this approach was pursued by Kuroda (1968). Other variants can be found in Hirschbühler and Rivero (1981), Harbert (1983), Suñer (1984), Grosu and Landman (1998), Grosu (2003), Caponigro (2002), Gračanin-Yuksek (2008) among others.

Even though the structure in (51) is identical to the structure of FRs in the present approach, former proposals as to how this structure arises lack an explanation for the special relationship between the covert head and the *wh*-phrase. In principle, there are two ways to model this relation: (i) excessive Agree which basically copies all the necessary features from the *wh*-phrase onto the covert head (Grosu 2003 notes that the covert head must even agree in categorial features, since there are also adjectival free relatives; see also footnote 9) and (ii) the approach developed above where the covert head emerges directly from the *wh*-phrase and has the necessary features to begin with.

The potential problem of the first option is the following: it remains unclear why the covert head in free relatives behaves different from the overt head in light-headed relatives, especially when it comes to case matching. In contrast to free relatives, light-headed relatives freely allow case mismatches. This can be seen in (52).

(52) a. *Ich folge Ø [$_{FR}$ wem/wen ich bewundere]
I follow→DAT who.DAT/who.ACC I adore→ACC
'I follow who I adore.'
b. Ich folge dem [$_{Rel}$ den ich
I follow→DAT that.one who.DAT/who.ACC I
bewundere]
adore→ACC
'I follow the one who I adore.'

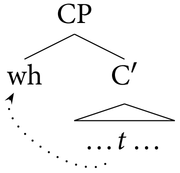
In the empty-head Comp account, the difference between the two structures in (52) is that the external head of the relatives clause is covert in (52-a), but overt in (52-b). However, as (52) suggests, this cannot be the only difference. There must be a special construction-specific constraint on FRs that force the covert head to check also its case features with the *wh*-phrase. It is far from clear how such a specific constraint could be justified.

If, on the other hand, the covert head in (52-a) does not come from the lexicon like the overt head in (52-b), but results from copying (part of) the *wh*-phrase, the close link between the covert head and the *wh*-phrase in (52-a) arises naturally. At the same time, the difference between free relatives and light-headed becomes clear immediately: in light-headed relatives, the external head does not emerge from the relative pronoun but is a separate lexical item. Thus, the relation between the overt head and the relative pronoun is supposed to be different.

5.2.2.2. *Rooryck (1994)*

The next approach, I will discuss was first proposed by Rooryck (1994).[24] Similarly to Groos and van Riemsdijk's (1981) account, the *wh*-phrase is located inside the embedded CP. In contrast to that approach, however, the CP is directly merged in an argument position in the matrix clause. The structure is shown in (53).

(53)

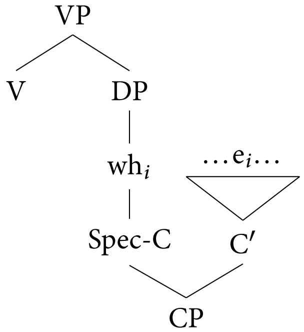


Even though, this approach may capture the extraction facts due to a lack of a higher DP shell, it suffers from an obvious and severe problem: it is unclear why the CP can occur in positions that are entirely reserved for DPs. On these grounds, this theory does not constitute an ideal approach to FRs.

5.2.3. *Multidominance*

In the grafting approach by van Riemsdijk (2006), the *wh*-phrase is simultaneously part of both the embedded and the matrix clause. This comes about by grafting, which is a form of multidominance. The structure is shown in (54).

(54)



---

[24] See also Caponigro (2003).

Even though van Riemsdijk's (2006) analysis is intuitively close to the present approach in that both assume that the *wh*-phrase is part of two clauses, the grafting approach faces at least two kinds of problems: First, there are the conceptual problems related to grafting, and to multidominance more generally (it is, e.g, unclear what predictions with respect to extraction such approaches make). Furthermore, the account faces an empirical problem when it comes to case matching: it is in fact not clear how the *wh*-phrase can fulfill two conflicting case requirements as illustrated by the data discussed in section 2.2.2 and section 4.2.2.[25] In the present approach, case mismatches are in principle allowed because there are two case features, one in the embedded clause on the *wh*-phrase and one in the matrix clause on the covert head. In the grafting approach, on the other hand, there is only one case feature which cannot bear two case values at the same time.

5.2.4. *Derived-Head Accounts*

The following summarizes and discusses two recent derived-head accounts: Ott (2011) and Donati and Cecchetto (2011).[26] In this type of theory, the *wh*-phrase is base-generated in the embedded CP and moves to a position where it becomes the head of the clause.

5.2.4.1. *Ott (2011)*

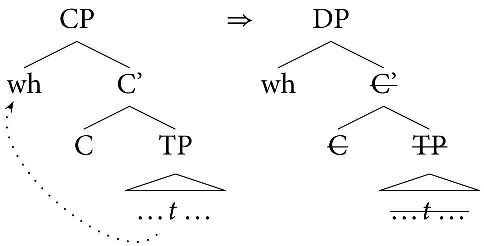
In Ott's (2011) analysis, the FR starts out as a normal CP where the *wh*-phrase is moved to Spec-C. Then, spell-out applies not only to the complement of C but to the C head as well because it does not bear any interpretable features (in contrast to, e.g., embedded questions). Since the head of the CP has been sent to Transfer, only the *wh*-phrase remains and becomes the head of the phrase. The important steps of the derivation are shown in (55).

---

[25] For this reason, van Riemsdijk (2006) argues, that there are in fact no case mismatches. See also the discussion in footnote 4.

[26] Other approaches of this kind include, e.g., Hirschbühler (1976), Bury and Neeleman (1999), Iatridou et al. (2002)

(55)

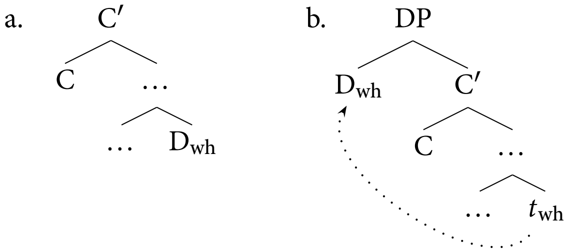


Similarly to the account by Bresnan and Grimshaw (1978), Ott's account has a hard time dealing with the extraposition data, because in order to account for the facts, the DP should be able to extrapose, contrary to fact. To this end, Ott (2011: fn.5,p.186) proposes two solutions: (i) extraposition applies post-syntactically after a process of intonational phrasing that assigns a clausal intonation to the FR-DP; (ii) CPs are base-generated in the right periphery and can undergo leftward-movement. Still, both solutions do not capture the data correctly. If only constituents with a clausal intonation could undergo extraposition, PPs are not supposed to be able to extrapose, contrary to fact. The second solution does not tell us anything about the situation in (55). At the point when the FR is constructed, it is already a DP and should thus not be able to remain in the right periphery. As far as I can see, these problems cannot be overcome without referring to the syntactic labels 'DP' and 'CP'.

5.2.4.2. *Donati and Cecchetto (2011)*

Finally, there is the reprojection approach by Donati and Cecchetto (2011). The main idea here is that the *wh*-phrase is merged inside a CP and moved to Spec-C. If it is a simple *wh*-phrase like *was* ('what') it may reproject as a head and turn the CP into a DP. The derivation is sketched in (56).

(56)



Similarly to Bresnan and Grimshaw (1978) and Ott (2011), the category dominating the *wh*-phrase is a DP. Therefore, the extraposition data do not follow straightforwardly from this analysis. Furthermore, the analysis bans complex *wh*-phrases from occuring in FRs. Donati and Cecchetto (2011) explicitly discuss this issue, claiming that FRs that contain a complex *wh*-phrase followed by *ever* are no real FRs. However, it remains unclear how, e.g., German FRs are derived where the *wh*-phrase is contained in a PP or where it is a possessor (see also Grosu 1996, 2003).

(57) a. Ich lade ein [$_{\text{FR}}$ [$_{\text{PP}}$ auf wen] sich auch Maria freuen würde]
I invite on who self also Maria be.happy would
'I invite whoever Maria would also be happy to meet'
(Vogel 2001: 904)

b. [$_{\text{FR}}$ Wessen Birne noch halbwegs in der Fassung steckt] pflegt
whose bulb still halfway in the socket sticks uses
solcherlei Erloschene zu vermeiden
such extinct to avoid
'Whoever still has half of his wits tends to avoid such vacant characters' *(Müller 1999: 78)*

In sum, the discussion of the previous approaches has shown that all of them lack an explanation for one or the other property of FRs discussed in section 2. Most of them make empirically wrong predictions. The only empirically adequate approach by Groos and van Riemsdijk (1981) was pursued and refined above in section 3. The present version of this approach is able to give an answer to the question about the special link between the covert head and the *wh*-phrase.

6. Conclusion

The paper has investigated free relative constructions, which pose a puzzle for syntactic theories because the *wh*-phrase that occurs in this construction seems to be simultaneously part of two sentences. While certain properties (number agreement, extraction) suggested that the *wh*-phrase is part of the matrix clause, other properties (extraposition, restricted instances of case mismatches) showed that it must be a part of the embedded clause. The ambivalent behavior of the *wh*-phrase with respect to its position is derived by distributing operations among three different components: the pre-syntactic, the syntactic, and the post-syntactic component. Since the order of these components is strict, the derivation of FRs proceeds in three different stages.

First, a part of the *wh*-phrase, namely the part containing the argument licensing features, is copied pre-syntactically in the lexical array. Second in the syntactic component, the *wh*-phrase containing only a copy of its argument licensing features is merged inside an embedded CP, where it is *wh*-moved to Spec-C. The original argument licensing features are merged as the head of the relative clause outside the FR. Third, the copied features must be deleted post-syntactically under c-command and matching with the original features.

The present approach can be seen as a development of the theory by Groos and van Riemsdijk (1981) that assumed that the CP is headed by a covert D head. The close link between this covert head and the *wh*-phrase can now be explained: the covert head is created out of the *wh*-item; i.e., it is a part of it.

References

Agbayani, Brian (1998): Feature Attraction and Category Movement. PhD thesis, University of California, Irvine.

Agbayani, Brian and Masao Ochi (2007): Split Lexical Insertion in Parasitic Gap Constructions. Ms. California State University, Fresno and Osaka University.

Assmann, Anke, Doreen Georgi and Philipp Weisser (2012*a*): A Minimalist Analysis of Possessor Advancement. *In:* E. Boone, K. Linke and M. Schulpen, eds, *Proceedings of ConSOLE XIX (2011, Groningen)*. pp. 77–103.

Assmann, Anke (2012*b*): Deriving Parasitic Gaps by Fission and Fusion. *In:* E. Boone, K. Linke and M. Schulpen, eds, *Proceedings of ConSOLE XIX (2011, Groningen)*. pp. 49–75.

Assmann, Anke, Svetlana Edygarova, Doreen Georgi, Timo Klein and Philipp Weisser (2013): Possessor Case in Udmurt: Multiple Case Assignment Feeds Postsyntactic Fusion. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 23–63.

Béjar, Susana and Milan Řezáč (2009): 'Cyclic Agree', *Linguistic Inquiry* 40(1), 35–73.

Bresnan, Joan (1973): 'Headless' Relatives. Ms., MIT.

Bresnan, Joan and Jane Grimshaw (1978): 'The Syntax of Free Relatives in English', *Linguistic Inquiry* 9(3), 331–391.

Brosziewski, Ulf (2003): *Syntactic Derivations - A Nontransformational View*. Vol. 470 of *Linguistische Arbeiten*, Niemeyer, Tübingen.

Bury, Dirk and Ad Neeleman (1999): Projection of Free Relatives. Ms., University College London.

Caponigro, Ivano (2002): Free Relatives as DPs with Silent D and a CP Complement. *In:* V. Samiian, ed., *Proceedings of WECOL 2000*. California State University, Department of Linguistics, Fresno, pp. 140–150.

Caponigro, Ivano (2003): Free Not to Ask: On the Semantics of Free Relatives and Wh-Words Cross-linguistically. PhD thesis, University of California Los Angeles, Los Angeles, Californien.

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

Chomsky, Noam (1980): 'On Binding', *Linguistic Inquiry* 11(1), 1–46.

Chomsky, Noam (1981): *Lectures on Government and Binding: The Pisa Lectures*. 7 edn, Mouton de Gruyter, Berlin.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and J. Uriagereka, eds, *Step by Step. Papers in Minimalist Syntax in Honor of Howard Lasnik*. MIT Press, Cambridge, chapter 3, pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. *In:* M. Kenstowicz, ed., *Ken Hale: A Life in Language*. MIT Press, Cambridge, Massachusetts, pp. 1–52.

Chomsky, Noam (2008): On Phases. *In:* R. Freidin, C. P. Otero and M. L. Zubizarreta, eds, *Foundational issues in linguistic theory: essays in honor of Jean-Roger Vergnaud*. MIT Press, Cambridge, chapter 6, pp. 133–166.

Chomsky, Noam (2013): 'Problems of Projection', *Lingua* . available online.

Citko, Barbara (2004): 'On Headed, Headless, and Light-Headed Relatives', *Natural Language and Linguistic Theory* 22, 95–126.

Daalder, Saskia (1977): 'Over relativie zinnen met ingesloten antecedent', *Spektator* 7/8, 401–407.

Doliana, Aaron (2013): The Super-Strong Person-Case Constraint: Scarcity of Resources by Scale-Driven Impoverishment. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 177–202.

Donati, Caterina and Carlo Cecchetto (2011): 'Relabeling Heads: A Unified Account for Relativization Structures', *Linguistic Inquiry* 42(4), 519–560.

Epstein, Sam, Erich Groat, Ruriko Kawashima and Hiratsugu Kitahara (1998): *A Derivational Approach to Syntactic Relations*. Oxford University Press, New York and Oxford.

Fanselow, Gisbert (1987): *Konfigurationalität*. Narr, Tübingen.

Fox, Danny and Jon Nissenbaum (1999): Extraposition and Scope: A Case for Overt QR. *In:* S. Bird, A. Carnie, J. D. Haugen and P. Norquest, eds, *Proceedings of the*

18th West Coast Conference on Formal Linguistics. Cascadilla Press, Somerville, Massachusetts, pp. 132–144.

Fuß, Eric (2008): 'Multiple Agreement and the Representation of Inflection in the C-Domain', *Linguistische Berichte* **213**, 77–106.

Fuß, Eric and Günther Grewendorf (2012): Freie Relativsätze mit d-Pronomen. Handout. Talk given at GGS 38, Freiburg im Breisgau.

Georgi, Doreen (2012): A Local Derivation of Global Case Splits. *In:* A. Alexiadou, T. Kiss and G. Müller, eds, *Local Modelling of Nonlocal Dependencies in Syntax*. Niemeyer, pp. 305–336.

Göbbel, Edward (2007): Extraposition and PF Movement. *In:* E. Brainbridge and B. Agbayani, eds, *Proceedings of WECOL 2006*. California State University, Fresno, Fresno, California, pp. 132–145.

Gračanin-Yuksek, Martina (2008): 'Free Relatives in Croatian: An Argument for the Comp Account', *Linguistic Inquiry* **39**, 275–294.

Groos, Anneke and Henk van Riemsdijk (1981): Matching Effects with Free Relatives: A Parameter of Core Grammar. *In:* A. Belletti, L. Brandi and L. Rizzi, eds, *Theory of Markedness in Generative Grammar: Proceedings of the 1979 GLOW Conference*. Scuola Normale Superiore, Pisa, pp. 171–216.

Grosu, Alexander (1996): 'The Proper Analysis of "Missing-P" Free Relative Constructions', *Linguistic Inquiry* **27**(2), 257–293.

Grosu, Alexander (2003): 'A Unified Theory of 'Standard' and 'Transparent' Free Relatives', *Natural Language and Linguistic Theory* **21**(2), 247–331.

Grosu, Alexander and Fred Landman (1998): 'Strange Relatives of the Third Kind', *Natural Language Semantics* **6**, 125–170.

Harbert, Wayne (1983): 'On the Nature of the Matching Parameter', *The Linguistic Review* **2**, 237–284.

Heck, Fabian (2010): Categories, Recursion and Bare Phrase Structure. Talk at GGS, Freie Universität Berlin.

Heck, Fabian and Malte Zimmermann (2004): Remarks on the Phasal Status of DP. Ms., Universität Tübingen, Humboldt-Universität Berlin.

Hirschbühler, Paul (1976): Two Analyses of Free Relatives in French. *In:* A. Ford, J. Reighard and R. Singh, eds, *North East Linguistic Society (NELS) 6*. Montreal Working Papers in Linguistics, Montreal, pp. 137–152.

Hirschbühler, Paul (1978): The Semantics and Syntax of wh-Constructions. PhD thesis, University of Massachusetts, Amherst.

Hirschbühler, Paul and María-Luisa Rivero (1981): 'Catalan Restrictive Relatives: Core and Periphery', *Language* **57**, 591–625.

Iatridou, Sabine, Elena Anagnostopoulou and Roumyana Izvorski (2002): Some Observations about the Form and Meaning of the Perfect. *In:* M. Kenstowicz, ed., *Ken Hale: A Life in Language*. MIT Press, Cambridge, Massachusetts, pp. 189–238.

Inaba, Jiro (2005): Extraposition and the Directionality of Movement. *In:* S. Blaho, L. Vicente and M. de Vos, eds, *Proceedings of ConSOLE XIII*. University of Leiden, Leiden, pp. 157–169.

Jacobson, Pauline (1995): On the Quantificational Force of English Free Relatives. *In:* E. Bach, E. Jelinek, A. Kratzer and B. H. Partee, eds, *Quantification in Natural Languages*. Kluwer, Dordrecht, pp. 451–486.

Keine, Stefan (2010): *Case and Agreement from Fringe to Core: A Minimalist Approach*. Vol. 536 of *Linguistische Arbeiten*, De Gruyter, Berlin.

Kiparsky, Paul (1971): Historical Linguistics. *In:* W. O. Dingwall, ed., *A Survey of Linguistic Science*. University of Maryland Linguistics Program, College Park, pp. 576–642.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Kuroda, Sige-Yuki (1968): 'English Relativization and Certain Related Problems', *Language* 44, 244–266.

Larson, Richard K. (1987): '"Missing Prepositions" and the Analysis of English Free Relative Clauses', *Linguistic Inquiry* 18(2), 239–266.

Link, Godehard (1983): The Logical Analysis of Plural and Mass Terms: A Lattice-Theoretical Approach. *In:* R. Bäuerle, C. Schwarze and A. von Stechow, eds, *Meaning, Use and Interpretation of Language*. de Gruyter, Berlin, pp. 302–323.

McCawley, James (1970): 'English as a VSO Language', *Language* 46, 286–299.

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

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

Müller, Stefan (1999): 'An HPSG-Analysis for Free Relatives in German', *Grammars* 2, 53–105.

Nunes, Jairo (1995): The Copy Theory of Movement and Linearization of Chains in the Minimalist Program. PhD thesis, University of Maryland, College Park.

Nunes, Jairo (2004): *Linearization of Chains and Sideward Movement.*. MIT Press, Cambridge, Massachusetts.

Ott, Dennis (2011): 'A Note on Free Relative Clauses in the Theory of Phases', *Linguistic Inquiry* 42(1), 183–192.

Pittner, Karin (1991): Freie Relativsätze und die Kasushierarchie. *In:* E. Feldbusch, ed., *Neue Fragen der Linguistik Vol. 1*. Niemeyer, Tübingen, pp. 341–347.

Pittner, Karin (1995): 'Regeln für die Bildung von freien Relativsätzen', *Deutsch als Fremdsprache* 32(4), 195–200.

Postal, Paul (1972): 'On Some Rules That Are Not Successive Cyclic', *Linguistic Inquiry* 3, 211–222.

Pullum, Geoffrey K. (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.

Riemsdijk, Henk van (2006): Free Relatives. *In:* M. Everaert and H. v. Riemsdijk, eds, *The Blackwell Companion to Syntax*. Blackwell, Oxford, pp. 338–382.

Rooryck, Johan (1994): 'Generalized Transformations and the wh-Cycle: Free Relatives as Bare wh-CPs', *Groninger Arbeiten zur germanistischen Linguistik* 37, 195–208.

Rouveret, Alain and Jean-Roger Vergnaud (1980): 'Specifying Reference to the Subject: French Causatives and Conditions on Representations', *Linguistic Inquiry* 11(1), 97–202.

Suñer, Margarita (1984): 'Free Relatives and the Matching Parameter', *The Linguistic Review* 3, 363–387.

Svenonius, Peter (2004): On the Edge. *In:* D. Adger, C. de Cat and G. Tsoulas, eds, *Peripheries: Syntactic Edges and their Effects*. Kluwer, Dordrecht, pp. 261–287.

Thomas, Daniela (2013): Split Ergativity in Subordination. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 1–21.

Trommer, Jochen (2006): Plural Insertion is Constructed Plural. *In:* G. Müller and J. Trommer, eds, *Subanalysis of Argument Encoding in Distributed Morphology. Volume 84 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 197–228.

Trommer, Jochen (2008): A Feature-Geometric Approach to Amharic Verb Classes. *In:* A. Bachrach and A. Nevins, eds, *Inflectional Identity*. Oxford University Press, pp. 206–236.

Truckenbrodt, Hubert (1995): Extraposition from NP and Prosodic Structure. *In:* J. Beckman, ed., *Proceedings of the North Eastern Linguistics Society 25*. GLSA Publications, Amherst, Massachusetts, pp. 503–517.

Vogel, Ralf (2001): Case Conflict in German Free-Relative Constructions: An Optimality-Theoretic Treatment. *In:* G. Müller and W. Sternefeld, eds, *Competition in Syntax*. Mouton de Gruyter, Berlin, pp. 341–375.

Wurmbrand, Susi (2007): 'How Complex Are Complex Predicates?', *Syntax* 10(3), 243–288.

A Local Reformulation of the Williams Cycle

Gereon Müller[*]

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

---

[*] For comments and discussion, I would like to thank Klaus Abels, David Adger, Josef Bayer, Petr Biskup, Fabian Heck, Antje Lahne, Martin Salzmann, Halldor Sigurðsson, Jochen Trommer, and audiences at the workshop on structure building at Konstanz University (April 2012) and at Leipzig University (April 2012). The research documented here was supported by DFG grant MU 144/8-1 ('Local Modelling of Non-Local Dependencies in Syntax'), principal researchers: Fabian Heck & Gereon Müller.

Rule Interaction in Grammar, 247–299
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

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) typically 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' 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 movement 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 and 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$_2$
which radio know you not how one
repariert]]?
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 and 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

---

[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 and 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,

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), Neeleman and van de Koot (2010), and Bader (2011), among others. The basic idea is that movement to (or, more generally,

---

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.

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 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]

---

[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' (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,

(12) *Generalized Ban on Improper Movement* (GBOIM; Williams 2003):
Given a Pollock/Cinque-style clausal structure $X_1 > \ldots > X_n$ (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 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).

---

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 calls the constraint *Principle of Hierarchy-Compliant Movement*.

(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 it 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 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

---

[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 on a single notion of affectedness of constituents.

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*b*) 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 to 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 [$_{vP}$ t_1''' Karl gemeint [$_{CP}$ t_1'' dass [$_{vP}$ t_1'
which book$_{acc}$ has Karl meant that
jeder t_1 lesen möge]]]?
everyone read should

b. *dass Karl [$_{vP}$ das Buch$_1$ glaubt [$_{CP}$ t_1'' dass [$_{vP}$ t_1' keiner
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$
t_1 liest]]]
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 and 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' 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 and 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 and Epstein (2011), respectively, are confined

---

[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".

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

2.6.2. Locality

Except for, possibly, the Activity Condition analysis and the Feature Splitting analysis, all the above accounts of improper movement require scanning large amounts of syntactic structure. Thus, the Principle C account must simultaneously take into account the base position of the moved item (which contains the trace that will ultimately give rise to a violation of Principle C); the position of the intermediate trace in the embedded SpecC position (which is relevant for determining whether the trace in base position obeys Principle C or not); and the position of the moved item in the final landing site that induces the constraint violation.[7]

---

[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 conditions that complies with this requirement (but cannot be extended to improper movement in any obvious way); also see Reuland (2001).

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" 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).[9]

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 up to Chomsky 1986).[10] Given the PIC and the assumption that CP, vP, and DP are phases, intermediate movement steps to Specv, SpecC,

---

[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.

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

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

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 and Grohmann (2007), and Müller (2011), among many others.[11] 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 move-

---

[11] 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).

ment. 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.[12]

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 and 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 un-

---

[12] Abels (2012*b*) 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*a*,*b*) adduces potentially interesting evidence from Tagalog to support such an approach. However, I will not consider this approach in any more detail in what follows because it denies that there is any inherent systematicity to the effects in (1-b)–(6-b); i.e., on this view, the generalization formulated at the beginning of the present 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*b*) 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 and 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.

der 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 and Sternefeld (1988: ch. 11.7), Sternefeld (1992), Müller and 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.[13]

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

---

[13] 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.

for improper movement effects by a reformulated Williams Cycle to be developed below, some clarifications about edge features and the role that they play in derivations are called for at this point.

The basic question is whether edge feature insertion is assumed to be freely available or severely constrained. A version of the first option is pursued in Chomsky (2008), 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.[14] 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.[15]

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

---

[14] Unless, that is, one assumes edge features to be flavoured in the sense of Abels (2012*b*); recall the discussion in footnote 12.

[15] 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 specific choice of frame of reference does matter; I will discuss these issues when they arise.

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 and 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 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 phase 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.[16]

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 gen-

---

[16] 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.

eral case" (see Halle and Marantz 1993), which then implies that reflexes of successive-cyclic movement will always qualify as morphological default exponents. 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 [•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).

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.

---

[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.

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) $[V] \rightarrow [V]\ [\bullet X_{V} \bullet]$
        (ii) $[v] \rightarrow [v], [\bullet X_{v} \bullet]$
        (iii) $[T] \rightarrow [T], [\bullet X_{T} \bullet]$
        (iv) $[C] \rightarrow [C], [\bullet X_{C} \bullet]$

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.[18] 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)).

---

[18] 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 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, pace Ott 2012, Chomsky 2013), or has been projected before edge feature generation applies.

(23) *Edge Feature Condition (EFC, revised)*:
An edge feature [•X_γ•] 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 [•β•]) 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.[19]

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

Third, edge feature discharge involves valuation of the movement-related feature of the moved item by the (defective) categorial information on the phase head, so as to ensure complete matching of the two items. Categorial information is successively added on top of the list. This is shown in (25) for an abstract derivation in which a wh-phrase undergoes successive-cyclic movement via all intervening phase edges to the embedded SpecC position (which

---

[19] 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 a feature like [(•)Σ(•)] for scrambling, see Grewendorf and 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.

is not its ultimate target position because, by assumption, C is declarative here and lacks [$\bullet$wh$\bullet$]).

(25) a. Merge(V:[$\bullet X_V \bullet$], DP:[wh:$\square$]) → V DP:[wh:$\boxed{\text{V}}$]
b. Merge(v:[$\bullet X_v \bullet$], DP:[wh:$\boxed{\text{V}}$]) → v DP:[wh:$\boxed{\text{vV}}$]
c. Merge(T:[$\bullet X_T \bullet$], DP:[wh:$\boxed{\text{vV}}$]) → T DP:[wh:$\boxed{\text{TvV}}$]
d. Merge(C:[$\bullet X_C \bullet$], DP:[wh:$\boxed{\text{TvV}}$]) → C DP:[wh:$\boxed{\text{CTvV}}$]

Fourth, when identical categorial information is added at the top, the original information is deleted at the bottom (more precisely, in a non-top position).[20] 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:[$\bullet X_C \bullet$], DP:[wh:$\boxed{\text{TvV}}$]) → C DP:[wh:$\boxed{\text{CTvV}}$]
b. Merge(V:[$\bullet X_V \bullet$], DP:[wh:$\boxed{\text{CTvV}}$]) → V DP:[wh:$\boxed{\text{VCTv}\cancel{\text{V}}}$]
c. Merge(v:[$\bullet X_v \bullet$], DP:[wh:$\boxed{\text{VCTv}\cancel{\text{V}}}$]) → 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) and (28).

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

(28) *Deletion*:
Y Z:[F:$\boxed{\gamma(\delta_1)\ldots(\delta_i)\gamma(\delta_j)\ldots(\delta_n)}$] → Y Z:[F:$\boxed{\gamma(\delta_1)\ldots(\delta_i)(\delta_j)\ldots(\delta_n)}$], where F is a movement-related feature, and $\delta_1 \ldots \delta_n$ and γ are pieces of categorial information.[21]

---

[20] 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) and Arsenijević and Hinzen (2012), where it is argued that direct recursion is not available in syntax. The conceptual underpinning for this latter view (a kind of ambiguity avoidance) looks very similar to the present restriction, but I will refrain from trying to suggest a unified approach here.

[21] This formulation presupposes that deletion of categorial information in feature lists is to be

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., [•wh$_{\mathrm{C}}$•]; this ensures deletion of the earlier $\boxed{\mathrm{C}}$ information in the case at hand; see (29) (again, the last relevant step of the earlier derivation is repeated here for convenience).

(29) a. Merge(v:[•X_{v}•], DP:[wh:$\boxed{\mathrm{VCTv}\cancel{\mathrm{V}}}$]) → v DP:[wh:$\boxed{\mathrm{vVCT}\cancel{\mathrm{vV}}}$]
b. Merge(T:[•X_{T}•], DP:[wh:$\boxed{\mathrm{vVCT}\cancel{\mathrm{vV}}}$]) → T DP:[wh:$\boxed{\mathrm{TvVC}\cancel{\mathrm{TvV}}}$]
c. Merge(C:[•wh$_{\mathrm{C}}$•], DP:[wh:$\boxed{\mathrm{TvVC}\cancel{\mathrm{TV}}}$]) → C DP:[wh:$\boxed{\mathrm{CTvV}\cancel{\mathrm{CTvV}}}$]

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>T>v>V); cf. (30).

(30) a. Merge(C:[•wh$_{\mathrm{C}}$•], DP:[wh:$\boxed{\mathrm{TvVC}\cancel{\mathrm{TV}}}$]) → C DP:[wh:$\boxed{\mathrm{CTvV}\cancel{\mathrm{CTvV}}}$]
b. Check C DP:[wh:$\boxed{\mathrm{CTvV}\cancel{\mathrm{CTvV}}}$] → √

This is the new version of the Williams Cycle, which can be formulated as in (31).

(31) *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 [•β_π•] 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 (32) in English.

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

---

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.

The derivation is shown in (33). 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 $[\bullet X_V \bullet]$ can be generated by defective copying here in accordance with the Edge Feature Condition in (23), and edge feature discharge triggers movement of $\text{DP}_{\text{wh}:\square}$ to SpecV, valuing the movement-related feature by adding categorial V information to the (initially empty) list: $\text{DP}_{\text{wh}:\boxed{\text{V}}}$; see (33-b).[22] Next, on the vP level, a new edge feature $[\bullet X_v \bullet]$ is generated and discharged by movement of $\text{DP}_{\text{wh}:\boxed{\text{V}}}$ to Specv, which further values the wh-feature on the moved item: $\text{DP}_{\text{wh}:\boxed{\text{vV}}}$. 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 (33-de). The pattern is repeated on the TP level, where $[\bullet X_T \bullet]$ is first generated and then discharged by movement of DP to SpecT, thereby valuing its wh-feature with the newly encountered syntactic context; see $\text{DP}_{\text{wh}:\boxed{\text{TvV}}}$ in (33-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 (33-h)). Finally, on the CP level, the inherent movement-inducing feature $[\bullet \text{wh}_C \bullet]$ on the interrogative C head triggers movement of $\text{DP}_{\text{wh}:\boxed{\text{TvV}}}$ to SpecC, valuing the wh-feature by adding the categorial information and thereby producing $\text{DP}_{\text{wh}:\boxed{\text{CTvV}}}$; cf. (33-j). Since an inherent structure-building feature has been checked at this point, the Williams Cycle in (31) 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.

---

[22] 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.

(33) a. $[_{\text{V}'}$ $[_{\text{V}}$ said $]_{[\bullet\text{X}_{V}\bullet]}$ what$_{[\text{wh}:\square]}]$
b. $[_{\text{VP}}$ what$_{[\text{wh}:\boxed{\text{V}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]$
c. $[_{\text{v}'}$ v$_{[\bullet\text{X}_{v}\bullet]\succ[\bullet\text{D}\bullet]}$ $[_{\text{VP}}$ what$_{[\text{wh}:\boxed{\text{V}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]$
d. $[_{\text{v}'}$ what$_{[\text{wh}:\boxed{\text{vV}}]}$ $[_{\text{v}'}$ v$_{[\bullet\text{D}\bullet]}$ $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]]$
e. $[_{\text{vP}}$ she $[_{\text{v}'}$ what$_{[\text{wh}:\boxed{\text{vV}}]}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]]]$
f. $[_{\text{T}'}$ T$_{[\bullet\text{X}_{T}\bullet]\succ[\bullet\text{EPP}\bullet]}$ $[_{\text{vP}}$ she $[_{\text{v}'}$ what$_{[\text{wh}:\boxed{\text{vV}}]}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]]]]$
g. $[_{\text{T}'}$ what$_{[\text{wh}:\boxed{\text{TvV}}]}$ T$_{[\bullet\text{EPP}\bullet]}$ $[_{\text{vP}}$ she $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]]]]$
h. $[_{\text{TP}}$ she $[_{\text{T}'}$ what$_{[\text{wh}:\boxed{\text{TvV}}]}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]]]]]$
i. $[_{\text{C}'}$ C$_{[\bullet\text{wh}\bullet]}$ $[_{\text{TP}}$ she $[_{\text{T}'}$ what$_{[\text{wh}:\boxed{\text{TvV}}]}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ said $]]]]]]]]]$
j. $[_{\text{CP}}$ what$_{[\text{wh}:\boxed{\text{CTvV}}]}$ $[_{\text{C}'}$ C $[_{\text{TP}}$ she $[_{\text{T}'}$ T $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{v}'}$ v $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ 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 (34).

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

The derivational steps in the embedded CP are almost exactly as in (32); see (35). 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 (32) 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 $[\bullet\text{X}_{\text{C}}\bullet]$ 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 (35-j) does not activate the Williams Cycle: Movement has not yet targetted a criterial position.

(35) *Embedded clause*:

a. [$_{\text{V}'}$ [$_{\text{V}}$ said]$_{[\bullet\text{X}_V\bullet]}$ what$_{[\text{wh:}\square]}$]
b. [$_{\text{VP}}$ what$_{[\text{wh:}\boxed{\text{V}}]}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]
c. [$_{\text{v}'}$ v$_{[\bullet\text{X}_v\bullet]\succ[\bullet\text{D}\bullet]}$ [$_{\text{VP}}$ what$_{[\text{wh:}\boxed{\text{V}}]}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]
d. [$_{\text{v}'}$ what$_{[\text{wh:}\boxed{\text{vV}}]}$ [$_{\text{v}'}$ v$_{[\bullet\text{D}\bullet]}$ [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]
e. [$_{\text{vP}}$ she [$_{\text{v}'}$ what$_{[\text{wh:}\boxed{\text{vV}}]}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]]
f. [$_{\text{T}'}$ T$_{[\bullet\text{X}_T\bullet]\succ[\bullet\text{EPP}\bullet]}$ [$_{\text{vP}}$ she [$_{\text{v}'}$ what$_{[\text{wh:}\boxed{\text{vV}}]}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]]]
g. [$_{\text{T}'}$ what$_{[\text{wh:}\boxed{\text{TvV}}]}$ T$_{[\bullet\text{EPP}\bullet]}$ [$_{\text{vP}}$ she [$_{\text{v}'}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]]]
h. [$_{\text{TP}}$ she [$_{\text{T}'}$ what$_{[\text{wh:}\boxed{\text{TvV}}]}$ T [$_{\text{vP}}$ [$_{\text{v}'}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]]]]
i. [$_{\text{C}'}$ C$_{[\bullet\text{X}_C\bullet]}$ [$_{\text{TP}}$ she [$_{\text{T}'}$ what$_{[\text{wh:}\boxed{\text{TvV}}]}$ T [$_{\text{vP}}$ [$_{\text{v}'}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]]]]]
j. [$_{\text{CP}}$ what$_{[\text{wh:}\boxed{\text{CTvV}}]}$ [$_{\text{C}'}$ C [$_{\text{TP}}$ she [$_{\text{T}'}$ T [$_{\text{vP}}$ [$_{\text{v}'}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ said]]]]]]]]]]

On the matrix VP, vP, and TP levels, edge feature generation, edge feature discharge, and wh-valuation on the moved item proceed as in the embedded domain, but there is an interesting difference: Movement to matrix SpecV in (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 (28)); 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., [•wh$_{\text{C}}$•], so the Williams Cycle will spring into action and demand a correspondence of f-seq and the wh-feature list present on the moved DP. However, movement to SpecC has resulted in adding C to the feature list, and the lower C symbol is then deleted. Therefore, the Williams Cycle is (non-vacuously) satisfied by the final movement step.

(35) *Matrix clause*:

k. [$_{\text{V}'}$ [$_{\text{V}}$ think]$_{[\bullet\text{X}_{V}\bullet]}$ [$_{\text{CP}}$ 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$_{[\bullet\text{X}_{v}\bullet]\succ[\bullet\text{D}\bullet]}$ [$_{\text{VP}}$ 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$_{[\bullet\text{X}_{T}\bullet]\succ[\bullet\text{EPP}\bullet]}$ [$_{\text{vP}}$ 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]]] ($\surd$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 ([Σ:□]) 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 [•X$_{\text{V}}$•] 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 [•Σ•], 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. [$_{\text{V}'}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\square]}$ [$_{\text{V}}$ liest]$_{[\bullet \text{X}_{V} \bullet]}$]
b. [$_{\text{VP}}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{V}}]}$ [$_{\text{V}'}$ [$_{\text{V}}$ liest]]]
c. [$_{\text{v}'}$ [$_{\text{VP}}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{V}}]}$ [$_{\text{V}'}$ [$_{\text{V}}$ liest]]] v$_{[\bullet\text{D}\bullet]\succ[\bullet\Sigma\bullet]}$]
d. [$_{\text{v}'}$ keiner [$_{\text{v}'}$ [$_{\text{VP}}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{V}}]}$ [$_{\text{V}'}$ [$_{\text{V}}$ liest]]] v$_{[\bullet\Sigma\bullet]}$]]
e. [$_{\text{vP}}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{vV}}]}$ [$_{\text{v}'}$ keiner [$_{\text{v}'}$ [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ liest]]] v]]] (✓f-seq)
f. [$_{\text{TP}}$ [$_{\text{vP}}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{vV}}]}$ [$_{\text{v}'}$ keiner [$_{\text{v}'}$ [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ liest]]] v]]] T]
g. [$_{\text{CP}}$ [$_{\text{C}}$ dass] [$_{\text{TP}}$ [$_{\text{vP}}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{vV}}]}$ [$_{\text{v}'}$ keiner [$_{\text{v}'}$ [$_{\text{VP}}$ [$_{\text{V}'}$ [$_{\text{V}}$ 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 [$_{\text{CP}}$ dass keiner t_1 liest]
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$ reads

In the embedded clause, the derivation proceeds in basically the same way as the derivation of well-formed long-distance wh-movement in steps a.–j. of (35). 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 $\boxed{\text{CTvV}}$ resulting at the CP level conforms to f-seq, but this is immaterial since the embedded SpecC is not yet a criterial position (movement to SpecC is triggered by an edge feature generated in the derivation, rather than by an inherent feature of C).

(39) *Embedded clause*:

a. $[_{\text{V}'}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\square]}$ $[_{\text{V}}$ liest $]_{[\bullet\text{X}_{V}\bullet]}$]
b. $[_{\text{VP}}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{V}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ liest]]]
c. $[_{\text{v}'}$ $[_{\text{VP}}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{V}}]}$ $[_{\text{V}'}$ $[_{\text{V}}$ liest]]] $\text{v}_{[\bullet\text{X}_{v}\bullet]\succ[\bullet\text{D}\bullet]}$]
d. $[_{\text{v}'}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{vV}}]}$ $[_{\text{v}'}$ $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ liest]]] $\text{v}_{[\bullet\text{D}\bullet]}$]]
e. $[_{\text{vP}}$ keiner $[_{\text{v}'}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{vV}}]}$ $[_{\text{v}'}$ $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ liest]]] v]]]
f. $[_{\text{T}'}$ $[_{\text{vP}}$ keiner $[_{\text{v}'}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{vV}}]}$ $[_{\text{v}'}$ $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ liest]]] v]]] $\text{T}_{[\bullet\text{X}_{T}\bullet]}$]
g. $[_{\text{TP}}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{TvV}}]}$ $[_{\text{T}'}$ $[_{\text{vP}}$ keiner $[_{\text{v}'}$ $[_{\text{v}'}$ $[_{\text{VP}}$ $[_{\text{V}'}$ $[_{\text{V}}$ liest]]] v]]] T]]
h. $[_{\text{C}'}$ $[_{\text{C}}$ dass $]_{[\bullet\text{X}_{C}\bullet]}$ $[_{\text{TP}}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{TvV}}]}$ $[_{\text{T}'}$ $[_{\text{vP}}$ keiner liest v] T]]]
i. $[_{\text{CP}}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{CTvV}}]}$ $[_{\text{C}'}$ $[_{\text{C}}$ dass] $[_{\text{TP}}$ $[_{\text{T}'}$ $[_{\text{vP}}$ 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 (39-k), which gives rise to a Σ-feature list $\boxed{\text{VCTv}}$, 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 (39-n): Movement to Specv gives rise to a list $\boxed{\text{vVCT}}$ on the long-distance scrambled DP's [Σ] feature, and since this last movement step is triggered by an inherent (albeit optional) structure-building feature $[\bullet\Sigma_{\text{v}}\bullet]$ on v, rather than by an edge feature $[\bullet\text{X}_{\text{v}}\bullet]$, the mismatch between f-seq and the feature list on the moved DP is 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 (39-o)–(39-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.

(39) *Matrix clause*:

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

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

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

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

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

(*f-seq → *Williams Cycle → crash)

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

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

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

r. $[_{\text{CP}}$ $[_{\text{C}}$ dass] $[_{\text{TP}}$ Karl $[_{\text{T}'}$ $[_{\text{vP}}$ $[_{\text{v}'}$ $[_{\text{DP}}$ das Buch $]_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ $[_{\text{v}'}$ $[_{\text{v}'}$ 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.[23]

Thus, on this view, clause-bound wh-movement in (40-a) does not induce a superiority effect, and clause-bound wh-movement in (40-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 (40-a), and *wen*$_1$ ('whom') in (40-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 $[\Sigma]$-feature on the DP); and not by virtue of $[\bullet X_v\bullet]$ on v.

(40) 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 (40-ab) involves a valuation of *both* movement-related features on the affected DP (viz., $[\Sigma:\square]$ and $[wh:\square]$), the list of $[\Sigma:\square]$ needs to conform to f-seq on the vP level (which it does: $[\Sigma:\boxed{\text{vV}}]$); 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 (41-a)) and the existence of weak crossover effects with long-distance wh-movement (see (41-b)) follow without further ado; see Frey (1993), Büring and Hartmann (1994), Fanselow (1996), Heck and Müller (2000), and Pesetsky (2000) for discussion of this phenomenon. It is the pres-

---

[23] 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 and 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).

ence 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.

(41) 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 (42) (cf. (4-b)).

(42) *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 [•EPP•]. Successive-cyclic movement must take place via the embedded TP and CP domains, and via the matrix VP and vP domains. In the final matrix SpecT position where [•EPP•] is discharged by attracting the moved DP, [EPP] on DP has the value $\boxed{\text{TvVC}}$, which fatally violates f-seq (hence, the Williams Cycle) because 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 (43-a), and accompanied by a slightly different construction instantiating the same problem.

(43) 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.[24] 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 (44).

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$\underline{\text{DNP}}$. 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 (43), 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, [$\bullet\omega_\text{D}\bullet$] on D, and matching [$\omega$:$\square$] on the PP), the list on the movement-related feature

---

[24] 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.

on the moved PP in the criterial SpecD target position respects the Williams Cycle: $[\omega{:}\boxed{\text{DN}}]$. 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 (44), 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 CNPC violations with argument extraction, which are standard cases of weak islands.[25] However, (44) 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.

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

Consider next (45), 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 (44) is striking.[26] This follows from the present version of the Williams Cycle: Movement originates in a CP; hence,

---

[25] 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).

[26] 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*.

given the extended f-seq, it must not end in a DP domain but needs to target the matrix CP domain again. PP$_1$ in (45) 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.

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

Finally, in (46), 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.

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

5.2. Relativization

As noted by Bayer and 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., (47-ab).

(47) 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 f-seq. Under this assumption, the relative pronouns in (47) bear a feature [rel] in the criterial position that violates the Williams Cycle: *[rel:$\boxed{\text{RelTvVC}}$].[27]

5.3. Exceptional Case Marking Constructions

Abels (2008) and Bader (2011) observe 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). Given the present implementation, the reason is that the relevant movement-related feature on the raised object (whatever this ultimately turns out to be; see above on similar issues with EPP-driven movement to SpecT) would then end up having a value $\boxed{\text{vVT}}$ (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.[28] 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 (48-abc) (where (48-b) is well formed only if the adverb *repeatedly* is construed with the embedded clause, an option that does not arise with *sincerely*

---

[27] 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.

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

in (48-c)). This is an argument for exceptional case marking, and against raising to object.

(48) 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 (49) 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.

(49) ?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 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 and Sternefeld 1993 and Bailyn 2001) and Japanese (see Saito 1985 and Grewendorf and 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.

(50) 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 and Soames 1979: ch. 43 and Alexiadou and Anagnostopoulou 2002, among others) and Kilega and other Bantu languages (see Obata and Epstein 2011 and references cited there). A Greek example is given in (51) (see Perlmutter and Soames 1979: 156):

(51) $[_{DP}$ 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.[29]

For concreteness, I would like to suggest that a key to a solution of the problem posed by data such as those in (50) and (51) 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 cat-

---

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

egories 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 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].[30] 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 (28) may refer only to a small (but fundamental) part of the category label, viz., information related to the [±Π] status of the phase head.[31]

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

---

[30] 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 and 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.

[31] 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.

effects are illustrated in (52-abc), for long-distance movement to the VP, vP , and TP domains, respectively.[32]

(52) a. (i) *$\boxed{\mathrm{VC}\cancel{\mathrm{TvV}}}$ *f-seq
(ii) *[$_{\mathrm{VP}}$ DP$_1$ … [$_{\mathrm{CP}}$ t_1'''' … [$_{\mathrm{TP}}$ t_1''' … [$_{\mathrm{vP}}$ t_1'' … [$_{\mathrm{VP}}$
V([−Π]) C([+Π]) ~~T([−Π])~~ ~~v([+Π])~~
t_1' … t_1 …]]]]]
~~V([−Π])~~

b. (i) $\boxed{\mathrm{vV}\cancel{\mathrm{CTvV}}}$ √f-seq
(ii) [$_{\mathrm{vP}}$ DP$_1$ … [$_{\mathrm{VP}}$ t_1''''' … [$_{\mathrm{CP}}$ t_1'''' … [$_{\mathrm{TP}}$ t_1''' … [$_{\mathrm{vP}}$
v([+Π]) V([−Π]) ~~C([+Π])~~ ~~T([−Π])~~
t_1'' … [$_{\mathrm{VP}}$ t_1' … t_1 …]]]]]]
~~v([+Π])~~ ~~V([−Π])~~

c. (i) $\boxed{\mathrm{Tv}\cancel{\mathrm{VCTvV}}}$ √f-seq
(ii) [$_{\mathrm{TP}}$ DP$_1$ … [$_{\mathrm{vP}}$ t_1'''''' … [$_{\mathrm{VP}}$ t_1''''' … [$_{\mathrm{CP}}$ t_1'''' … [$_{\mathrm{TP}}$
T([−Π]) v([+Π]) ~~V([−Π])~~ ~~C([+Π])~~
t_1''' … [$_{\mathrm{vP}}$ t_1'' … [$_{\mathrm{VP}}$ t_1' … t_1 …]]]]]]]
~~T([−Π])~~ ~~v([+Π])~~ ~~V([−Π])~~

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 (53-abc), again for long-distance movement to the VP, vP , and TP domains.

(53) a. (i) *$\boxed{\mathrm{VCT}\cancel{\mathrm{vV}}}$ *f-seq
(ii) *[$_{\mathrm{VP}}$ DP$_1$ … [$_{\mathrm{CP}}$ t_1'''' … [$_{\mathrm{TP}}$ t_1''' … [$_{\mathrm{vP}}$ t_1'' … [$_{\mathrm{VP}}$ t_1' …
V C([+Π]) T ~~v([+Π])~~ ~~V~~
t_1 …]]]]]

---

[32] 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.)

b. (i) *$\boxed{\text{vV}\cancel{\text{C}}\text{T}\cancel{\text{vV}}}$ *f-seq
(ii) *[$_{\text{vP}}$ DP$_1$ … [$_{\text{VP}}$ t_1''''' … [$_{\text{CP}}$ t_1'''' … [$_{\text{TP}}$ t_1''' … [$_{\text{vP}}$
v([+Π]) V ~~C([+Π])~~ T
t_1'' … [$_{\text{VP}}$ t_1' … t_1 …]]]]]]
~~v([+Π])~~ ~~V~~

c. (i) $\boxed{\text{TvV}\cancel{\text{CTvV}}}$ √f-seq
(ii) [$_{\text{TP}}$ DP$_1$ … [$_{\text{vP}}$ t_1'''''' … [$_{\text{VP}}$ t_1''''' … [$_{\text{CP}}$ t_1'''' … [$_{\text{TP}}$ t_1''' …
T v([+Π]) V ~~C([+Π])~~ ~~T~~
[$_{\text{vP}}$ t_1'' … [$_{\text{VP}}$ t_1' … t_1 …]]]]]]]
~~v([+Π])~~ ~~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. (54-abc).

(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'' … [$_{\text{VP}}$
V([−Π]) C ~~T([−Π])~~ v
t_1' … t_1 …]]]]]
~~V([−Π])~~

b. (i) *$\boxed{\text{vVC}\cancel{\text{TvV}}}$ *f-seq
(ii) *[$_{\text{vP}}$ DP$_1$ … [$_{\text{VP}}$ t_1''''' … [$_{\text{CP}}$ t_1'''' … [$_{\text{TP}}$ t_1''' … [$_{\text{vP}}$ t_1'' …
v V([−Π]) C ~~T([−Π])~~ ~~v~~
[$_{\text{VP}}$ t_1' … t_1 …]]]]]]
~~V([−Π])~~

c. (i) *$\boxed{\text{Tv}\cancel{\text{V}}\text{C}\cancel{\text{TvV}}}$ *f-seq
(ii) *[$_{\text{TP}}$ DP$_1$ … [$_{\text{vP}}$ t_1'''''' … [$_{\text{VP}}$ t_1''''' … [$_{\text{CP}}$ t_1'''' … [$_{\text{TP}}$
T([−Π]) v ~~V([−Π])~~ C
t_1''' … [$_{\text{vP}}$ t_1'' … [$_{\text{VP}}$ t_1' … t_1 …]]]]]]]
~~T([−Π])~~ ~~v~~ ~~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 (53) (with shared [+Π] permitting deletion in addition to full identity).[33] Languages like Russian could choose the option in (52) (with shared [+Π] and [−Π] permitting deletion in addition to full identity).[34] However, the option in (53) 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 individual movement-related features (in the case at hand: [Σ:□] vs. [EPP:□] as F in (28)), rather than among languages.

As for option (54), 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 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 (54) 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. (Finally, v deletes the lower v symbol.)

---

[33] Such an analysis does not by itself correlate the availability of super-raising in a language with some other, independently established property. Obata and 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 and 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 and Epstein 2011: 139 note), Greek super-raising is potentially problematic for the feature splitting approach.

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

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 (55) 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.

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

The same conclusion can be drawn for pronoun fronting from ECM complements in German; see (56).

(56) a. dass er es$_1$ [$_{XP}$ den Jungen t$_1$ lesen sah]
that he$_{nom}$ it$_{acc}$ the boy$_{acc}$ read saw
b. dass er es$_1$ [$_{XP}$ den Jungen t$_1$ machen ließ]
that he$_{nom}$ it$_{acc}$ the boy$_{acc}$ make let

This potential problem disappears if the option in (54) 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 (55) and (56); Fanselow and Wurmbrand suggest a vP/VP embedding, which renders the data in (55) and (56) 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.[35]

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

---

[35] 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 $[_{DP}$ dieser Frau $]_1$ $[_{XP}$ den Jungen t$_1$ helfen sah/ließ]
that no-one$_{nom}$ this woman$_{dat}$ the boy$_{acc}$ help saw/let
b. *dass er $[_{DP}$ ihm $]_1$ $[_{XP}$ 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, 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

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), Neeleman and van de Koot (2010) and Bader (2011). In the existing analyses where a version of the Williams Cycle is put to use, it is generally formulated in a non-local way, such that large amounts of syntactic structure must be scanned in order to decide whether a given interaction of movement steps counts as improper or not (though see footnote 10). 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 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.[36] Still, the approach chosen here strikes me as a fairly straightforward one because it is both fine-grained and potentially flexible, and (not least of all) because it structurally assimilates the 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.[37] For reasons of space and coherence, and since my main goal has only been to show that a formulation of the Williams

---

[36] 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.

[37] 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 (Petr 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

Cycle is possible in principle in a strictly derivational approach to syntax, I will not try to pursue them here.

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*a*): *Phases. An Essay on Cyclicity in Syntax*. Vol. 543 of *Linguistische Arbeiten*, De Gruyter, Berlin.

Abels, Klaus (2012*b*): 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 and Elena Anagnostopoulou (2002): Raising Without Infinitives and the Nature of Agreement. In: *Dimensions of Movement: From Remnants to Features*. Benjamins, Amsterdam, pp. 17–30.

Arsenijević, Boban and Wolfram Hinzen (2012): 'On the Absence of X-Within-X Recursion in Human Grammar', *Linguistic Inquiry* 43, 423–440.

Bader, Monika (2011): Constraining Operations. A Phase-Based Account of Improper Movement and Anaphoric Binding. PhD thesis, CASTL, Tromsø University., Tromsø.

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

Bayer, Josef and 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 and 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.

---

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.

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

Chomsky, Noam (1970): Remarks on Nominalization. *In:* R. Jacobs and 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 and 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 and 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 and 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 and 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 and M. L. Zubizarreta, eds, *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, Mass., pp. 133–166.

Chomsky, Noam (2013): 'Problems of Projection', *Lingua* . To appear.

Collins, Chris (2002): Eliminating Labels. *In:* S. D. Epstein and 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 and 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 and Sam Gutmann (2002): Crash-Proof Syntax. *In:* S. D. Epstein and 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 and 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 and Alec Marantz (1993): Distributed Morphology and the Pieces of Inflection. *In:* K. Hale and S. J. Keyser, eds, *The View from Building 20*. MIT Press, Cambridge, Mass., pp. 111–176.

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

Heim, Irene and 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.

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

Neeleman, Ad and 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 and N. Richards, eds, *Proceedings of the EPP/Phase Workshop*. MITWPL, Cambridge, Mass., pp. 287–310.

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

Ott, Dennis (2012): *Local Instability*. Vol. 544 of *Linguistische Arbeiten*, Mouton de Gruyter, Berlin.

Perlmutter, David and 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.

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 and 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 and 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.

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

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

Sternefeld, Wolfgang (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 and M. Everaert, eds, *The Blackwell Companion to Syntax*. Blackwell, Oxford, pp. 392–436.

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.

In Between Subordination and Coordination: A Minimalist Analysis of the Left Subordinating *and*-Construction

Philipp Weisser*

Abstract

In Culicover and Jackendoff (1997), the left-subordinating and-construction is cited as one example of a literal mismatch between syntax and semantics because it seems to exhibit syntactic properties of coordinate structures and semantic properties of subordinate structures at the same time. Hence, as they conclude, this construction cannot be derived by frameworks such as Minimalism where the semantics is derived from some syntactic level of representation such as LF. I want to argue that it is possible to derive the specific properties of this construction under Minimalist assumptions if one adheres to a strictly derivational model of Minimalism and subscribes to the assumption that a clause can be base-generated as an adjunct low in the tree and then be moved to the specifier of a coordination phrase as the regular first conjunct of a coordination.

1. Introduction

Ever since Munn (1987) made the much acclaimed proposal that the structure of coordination is asymmetrical in nature in the sense that the first conjunct is higher in the structure and c-commands the second one, a lot of people have presented arguments in favor of that view (see e.g. Munn 1993, 1999, Kayne 1994, Johannessen 1998, Harbert and Bahloul 2002, Citko 2004, van Koppen 2006 and many others). Some of these arguments were based on theory-internal considerations, for example that all syntactic structures must be binary branching or that syntactic phrases must contain exactly one head.

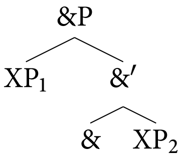
---

* For helpful comments and discussion I would like to thank the participants of the Colloquium "Neuere Arbeiten zur Grammatiktheorie" at the University of Leipzig and in particular Anke Assmann, Doreen Georgi, Martin Salzmann and Gereon Müller. This research was funded by the DFG in the course of the project "Lokale Modellierung nicht-lokaler Abhängigkeiten in der Syntax".

Rule Interaction in Grammar, 301–331
Fabian Heck & Anke Assmann (eds.)
LINGUISTISCHE ARBEITS BERICHTE 90, Universität Leipzig 2013

Other arguments were based on empirical findings. The empirical arguments in favor of the asymmetric coordination structure mainly circled around the topics of binding and partial conjunct agreement. In the light of these arguments it seems justified to say that one has reached a consensus on how the structure of coordination looks like[1]. There is one coordination head &, morphologically realized by the conjunction, which selects the second conjunct as its complement and the first conjunct as its specifier:

(1)



If (1) is in fact the right structure for a coordination and the postulated $\&^0$ is a normal binary branching syntactic head, then we would expect that there should in principal be cases of XP-movement to the specifier of this coordination phrase, at least from a syntactic point of view. However, as far as I know, this possibility has not been explored in the literature despite the fact that it seems an undesirable stipulation to say that a coordination head should be the only head to whose specifier movement is not possible. As is well-known, there are additional syntactic and semantic restrictions on the type of both arguments of a coordination head, so there may not be too many grammatical cases of movement to Spec&P but if these restrictions are met, movement to Spec&P should, in principle, be possible.

In the present paper, I argue that the so-called "left-subordinating and-construction" (Culicover 1970, 1972, Culicover and Jackendoff 1997), which exhibits properties of subordinate but also of coordinate clauses, is an instance of movement to the specifier of a coordination head.

I will proceed as follows: Section 2 illustrates the properties of the aforementioned left-subordinating and-construction. In section 3, I will outline the analysis by Culicover and Jackendoff (1997) and discuss its shortcomings and why it is radically incompatible with a Minimalist framework. Section 4 presents the assumptions I make and shows how, under these assumptions, the properties of the left-subordinating and-construction can be derived. Section 5 discusses some of the consequences of my approach, some of its ad-

---

[1] Although there is definitely no consensus on how this coordination phrase should be labelled. Proposals include &P, CoP, CoordP, andP, BP or BoolP. I will stick to the term &P.

vantages as well as potential problems and revisits the discussion within the broader context of rule interaction. Section 6 concludes.

2. The Left-Subordinating *and*-Construction

The left-subordinating and-construction (in what follows: ${}_{\text{LS}}$and-construction) as well as the related "One more"- or OM-construction have been the subject of linguistic studies for several decades now (cf. Ross 1967, Culicover 1970, 1972, Lakoff 1986, Culicover and Jackendoff 1997, Takahashi 2004 among others). Both constructions are characterized by an apparent mismatch between syntax and semantics because, syntactically, examples like (2) look like cases of coordination but their semantics resembles the semantics of a conditional clause, which is generally assumed to be subordinate.

(2) a. You drink one more can of beer and I'm leaving.
${}_{\text{LS}}$*and-construction*
b. One more can of beer and I'm leaving.
OM-construction
c. If you drink one more can of beer, I'm leaving.
Conditional (Culicover and Jackendoff 1997: 197ff)

Culicover and Jackendoff (1997), who were, to my knowledge, the first to carve out the properties of these constructions in detail, observed that ${}_{\text{LS}}$and-constructions may be used to paraphrase conditional clauses; however their distribution is much more restricted. For example, the conditional reading of these constructions is lost when they appear in perfect tense:

(3) You've drunk another can of beer and I've left.
(Culicover and Jackendoff 1997: 198)

Apparently, ${}_{\text{LS}}$and-constructions cannot paraphrase irrealis conditionals. The same happens, when more than two conjuncts are involved. (4) does not mean the same as the conditional *If you drink another can of beer (and if) Bill eats more pretzels, I'm leaving.*.

(4) You drink another can of beer, Bill eats more pretzels, and I'm leaving.
(Culicover and Jackendoff 1997: 198)

Another important property of $_{\text{LS}}$and-constructions is that they are restricted to TP/IP-coordination. The conditional meaning is lost when you conjoin either two CPs[2] (as in (5-b)) or two VPs (as in (5-c)).

(5) a. You know, of course, [$_{\text{CP}}$ that [$_{\text{TP}}$ you drink one more beer] and [$_{\text{TP}}$ you get kicked out]].
b. You know, of course, [$_{\text{CP}}$ that you drink one more beer] and [$_{\text{CP}}$ that you get kicked out]. (Culicover and Jackendoff 1997: 198)
c. You [$_{\text{VP}}$ drink one more beer] and [$_{\text{VP}}$get kicked out].

Even though $_{\text{LS}}$and-constructions look like coordinate clauses on the surface, they share a lot of properties with the construction they paraphrase, namely conditional clauses. For example, conditionals and $_{\text{LS}}$and-constructions can neither undergo Right Node Raising (cf. (6-b) and (6-c)) nor Gapping (cf. (7-b) and (7-c)). Normal coordinate clauses can (cf. (6-a) and (7-a)).

(6) a. Big Louie found out about _ and Big Louie put a contract on _, that guy who stole some loot from the gang. *Coordination*
b. *Big Louie finds out about _ and Big Louie puts a contract on _, that guy who stole some loot from the gang. $_{\text{LS}}$*and-construction*
c. *If Big Louie finds out about _ , then Big Louie puts a contract on _, that guy who stole some loot from the gang. *Conditional*
(Culicover and Jackendoff 1997: 198f)

(7) a. Big Louie stole another car radio and Little Louie the hubcaps. *Coordination*
b. *Big Louie steals one more car radio and Little Louie the hubcaps. $_{\text{LS}}$*and-construction*
c. *If Big Louie steals one more car radio, then Little Louie the hubcaps. *Conditional*
(Culicover and Jackendoff 1997: 198f)

---

[2] Of note is that in German the first conjunct seems to be a CP as it is verb-first.

(i) Fass mich noch einmal an und du bist tot.
touch me more once PRTC and you are dead.
'Touch me once more and you are dead.'

I will confine myself to the English data here. For an overview of the properties of the equivalent construction in German see te Velde (2005).

Also when it comes to binding, LSand-constructions behave like conditional if-clauses. An anaphor in the left conjunct of an LSand-construction may be bound by an antecedent within the second conjunct (cf. (8-a)). This parallels the behaviour of conditionals (8-b) but not the behaviour of normal coordination (8-c).[3]

(8) a. Another picture of himself$_i$ appears in the newspaper and John$_i$ will definitely go out and get a lawyer. LS*and-construction*
b. If another picture of himself$_i$ appears in the newspaper, then Susan thinks John$_i$ will definitely go out and get a lawyer. *Conditional*
c. *Another picture of himself$_i$ has appeared in the newspaper and Susan thinks John$_i$ will definitely go out and get a lawyer. *Coordination*
(Culicover and Jackendoff 1997: 202)

The same picture can be found with respect to variable binding. An antecedent in the second conjunct can bind a variable in the first conjunct in LSand-constructions and conditionals but not in classical cases of coordination:

(9) a. You give him$_i$ enough opportunity and every senator$_i$, no matter how honest, will succumb to corruption. LS*and-construction*
b. If you give him$_i$ enough opportunity, every senator$_i$, no matter how honest, will succumb to corruption. *Conditional*
c. *We gave him$_i$ enough opportunity and every senator$_i$, no matter how honest, succumbed to corruption. *Coordination*
(Culicover and Jackendoff 1997: 204)

However, when it comes to extraction, LSand-constructions behave neither like coordinate clauses nor like subordinate conditionals. Unlike with coordinate clauses (cf. (10-a)), ATB movement out of an LSand-construction is, according to Culicover and Jackendoff (1997), "decidedly strange" (cf. (10-b)).

(10) a. This is the thief that you just pointed out _ and we arrested _ on the spot.

---

[3] As Culicover and Jackendoff (1997) show, there are several exceptions to the rule that an anaphor in the first conjunct can be bound, but all these exceptions equally hold for if-clause conditionals.

b. ??This is the thief that you just point out _ and we arrest _ on the spot. (Culicover and Jackendoff 1997: 206)

Asymmetrical non-ATB-movement out of only one conjunct, on the other hand, seems to be possible (and is only slightly degraded if at all) with $_{\text{LS}}$and-constructions, both from the left conjunct (as in (11-a)) or from the the right conjunct (as in (11-b)).

(11) a. ?This is the loot that you just identify _ and we arrest the thief on the spot.
b. ?This is the thief that you just identify the loot and we arrest _ on the spot. (Culicover and Jackendoff 1997: 206)

The data in (11) sets the $_{\text{LS}}$and-construction apart from both, if-clause conditionals as well as classical cases of coordination. With if-clause constructions, you can, of course, only extract out of the matrix clause since extraction out of the adjunct violates the Condition on Extraction Domain (CED) (cf. (12-a)). With classical coordination, you can neither extract out of the first nor the second conjunct since these kinds of extraction are forbidden by the Coordinate Structure Constraint (CSC) (cf. (12-b,c)).

(12) a. ??This is the loot that if you identify _ we will arrest the thief on the spot.
b. *This is the loot that you have identified _ and we have arrested the thief on the spot.
c. *This is the thief that you have identified the loot and we have arrested _ on the spot. (Culicover and Jackendoff 1997: 207)

Let us briefly summarize the peculiar properties of the construction: The $_{\text{LS}}$and-construction consists of two TPs connected by the usual conjunction *and*. Even though it looks like a coordination on the surface, the $_{\text{LS}}$and-construction behaves like a subordinate conditional clause with respect to Right Node Raising, Gapping and Binding. It behaves neither like coordination nor like conditional subordination with respect to extraction since it allows extraction out of both of these TPs but only one at a time. ATB-like movement out of both conjuncts is impossible.

3. The Analysis by Culicover and Jackendoff (1997)

The fact that $_{\text{LS}}$and-constructions behave just like conditional clauses with respect to most of the tests in the last section might suggest that they are actually not coordinate at all, and one might think the conjunction *and* may also be used as a subordinator. Culicover and Jackendoff (1997) briefly discuss this possibility but discard it for three reasons.

First, they argue that if *and* was a subordinator, it would be a very strange one, at least for English because it would be the only clause-final subordinator that English has. All other complementizers or C-elements in general always appear in clause-initial position.

Second, they show that unlike other subordinate clauses, the subordinate clause of a $_{\text{LS}}$and-construction must always appear in sentence-initial position. Other kinds of subordinate clauses can appear either sentence-intial, sentence-final or sometimes even within the main clause:

(13) a. *[$_{\text{S}}$ Big Louie puts out a contract on you, [$_{\text{S}}$ [$_{\text{S}}$ he sees you with the loot] and]]. (Culicover and Jackendoff 1997: 200)
b. Big Louie puts out a contract on you if he sees you with the loot.
c. If he sees you with the loot, Big Louie puts out a contract on you.

Third, assuming that *and* could be used as a subordinator as well does not solve all the problems because, as we saw in the last section, with respect to extraction, $_{\text{LS}}$and-constructions neither behave like coordinate nor like subordinate structures. Hence, one would need an additional explanation for the fact that one may extract out of these kinds of adjuncts anyway.

Hence, Culicover and Jackendoff (1997) conclude that the analysis of *and* as a normal subordinator of English cannot be maintained. Instead they propose a completely new analysis. It is based on the assumption that the apparent mismatch between syntax and semantics is to be taken literally here. $_{\text{LS}}$and-constructions are syntactically coordinate but semantically subordinate.

This assumption helps to solve the puzzle of the $_{\text{LS}}$and-construction if one further assumes that all cases where the $_{\text{LS}}$and-construction behaves like a coordinate clause are syntactic phenomena and all the cases where it behaves like a subordinate clause are semantic phenomena. This means, for example, that all kinds of binding (i.e. licensing of anaphors, variable binding) make reference to semantics and ignore syntax completely. However, Culicover and

Jackendoff (1997) are not explicit as to how these purely semantic principles can derive the data in (8) and (9). On the other hand, the fact that the semantically subordinate clause of a $_{\text{LS}}$and-construction must always appear clause-initially is due to the fact that the whole construction is syntactically coordinate and extraposition of a first conjunct violates some syntactic principle.[4]

The question of extraction is a little bit more complicated because there are two constraints which restrict extraction, a syntactic one, namely the Condition on Extraction Domain (CED) and a semantic one, namely the Coordinate Structure Constraint (CSC). Since the $_{\text{LS}}$and-construction is semantically subordinate (hence the CSC does not apply) and syntactically coordinate (hence, the CED does not apply), extraction from either conjunct is allowed. The $_{\text{LS}}$and-construction is, so to speak, specifically designed to circumvent all constraints that restrict extraction. However, as we have seen, extraction out of $_{\text{LS}}$and-constructions is not completely unrestricted since one cannot ATB-move out of a $_{\text{LS}}$and-construction. Culicover and Jackendoff (1997) give no explanation for this fact, at least not an explicit one. As far as I can see, to save their account, they would have to assume that ATB-movement is a semantic process.[5] The following table gives an overview about which processes and constraints apply to which structure (or on the basis of which structure).

(14)

Syntax	Semantics
Extraposition	Gapping
CED	Right-Node Raising
	Licensing of anaphors
	Variable Binding
	CSC
	(ATB-Movement)

Whenever the $_{\text{LS}}$and-construction patterns with conditional if-clauses with respect to some property, this property is a semantic property and whenever some property of the $_{\text{LS}}$and-construction patterns with coordinate clauses, the property must be a syntactic one.

---

[4] In its standard formulation the Coordinate Structure Constraint (Ross 1967) prohibits extraction of the first conjunct. However, since the Coordinate Structure Constraint is, according to Culicover and Jackendoff (1997), a purely semantic principle and does not apply to $_{\text{LS}}$and-constructions, this ban on extraposition cannot be attributed to it.

[5] Another possibility would perhaps be to assume that ATB-movement is syntactic but it may only apply to semantically coordinate structures regardless of how the syntax looks like.

However, the question is whether the classification of operations and constraints into these two distinct classes is empirically justified or merely stipulated to make the system work. Since the main purpose of this paper is to present an alternative analysis of $_{\text{LS}}$and-constructions, I do not want to delve too deeply into that discussion but it seems to me that the classification in (14) is, at least for some of the phenomena, far from uncontroversial. The question of whether binding theory (i.e. licensing of anaphors and variable binding) is a matter of syntax or semantics is still under debate. To my knowledge, it has not been possible to capture the complexity of the whole topic of binding theory by means of one module, either syntax or semantics. Hence, it seems problematic to just say that binding applies within the semantics without reference to syntactic structures at all. A similar point can be made for Gapping which is well-known to underlie a certain number of locality constraints (i.e. Gapping may not cross syntactic barriers. (cf. Hankamer 1973, Neijt 1979, Pesetsky 1982, Chao 1988, Hartmann 2000, Murguia 2004)). And since such locality constraints are usually thought to be syntactic in nature, it seems implausible that Gapping is a purely semantic process. In my opinion, this shows that the approach by Culicover and Jackendoff (1997) has very far-reaching and often undesirable consequences.

Apart from that, a note is in order about the relation of syntactic and semantic structure in the account of Culicover and Jackendoff (1997) in general. As we have seen, they try to show that the $_{\text{LS}}$and-construction is semantically subordinate and syntactically coordinate. And since these two states, namely being subordinate and being coordinate, are, according to Culicover and Jackendoff (1997), incompatible with each other within the same level of representation, they try to construct an argument for their hypothesis that syntax and semantics are two completely distinct levels and "that syntax is therefore autonomous in that it is not reducible to semantic structure, and semantic structure is not isomorphic to any level of syntactic structure such as LF" (Culicover and Jackendoff 1997: 196). This, of course, is incompatible with the standard model of Minimalism, where the semantics is built on the basis of syntactic structure.

I will, in the next section, show that it is possible to develop an analysis of the $_{\text{LS}}$and-construction that is consistent with the fundamental assumptions of Minimalist Theory and nevertheless captures all of its central properties.

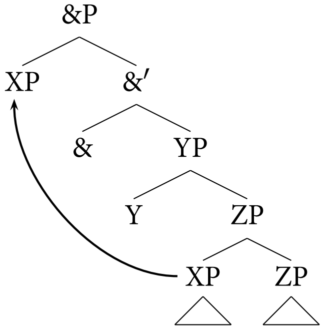
4. A Derivational Analysis of the Left-Subordinating *and*-Construction

4.1. The Main Idea of the Analysis

As we have seen in the previous sections, the $_{\text{LS}}$and-construction seems to exhibit subordinate and coordinate properties at the same time. And since it is generally assumed that a clause cannot be subordinate and coordinate simultaneously at the same level of representation , Culicover and Jackendoff (1997) concluded that the puzzling properties of this construction are due to a mismatch between syntax and semantics which, thus, must be two completely autonomous levels of representation.

I would like to argue instead that a derivational approach is particularly well-suited to derive the dual behavior of $_{\text{LS}}$and-constructions while still maintaining the standard view that the semantics interprets syntactic structures. A clause may start out as an adjunct, which accounts for its subordinate properties; through movement into &P, this clause becomes a conjunct and is therefore expected to also exhibit coordinative properties.

(15)

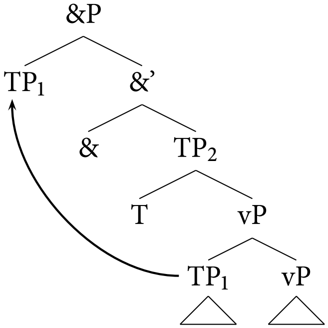


First, XP is merged as an adjunct to ZP. Later on, after ZP became the complement of YP, YP itself is merged with a coordination head & becoming the second conjunct of the coordination. However, unlike in normal cases of coordination, the first conjunct does not come into being via external merge but via internal merge (i.e. movement) of XP. Hence, with respect to processes that (may) apply early in the derivation, it may look like XP has subordinative properties even though it appears to be coordinate on the surface.

Let us make the move from the the abstract to the concrete. We have seen in section 2 that LSand-constructions always involve TP/IPs.The conditional reading is lost with CP- and VP-coordination (recall the examples in (5)). The most natural way to account for fact that none of the conjuncts in an LSand-constructions may contain a complementizer but the whole LSand-construction may itself be embedded directly under a complementizer is to assume that XP and YP are mere TPs.[6]

Since the first clause of an LSand-construction has the interpretation of a conditional clause, I assume that it is base generated in the same position as normal conditionals. According to (Haegeman 2003: 326) event-conditionals[7] are merged "before the IP is completed", hence I take them to be adjuncts to vP. I conclude that the structure of LSand-construction looks as follows:

(16)



---

[6] It has been claimed in the literature that TPs cannot be moved in languages like English or German because one never finds stranded complementizers. However, as Abels (2003) shows, there is no general prohibition against the movement of TPs. Instead, what Abels (2003) argues for is that TPs can be moved if they are not directly embedded under a C-head. And since the TPs we are dealing here are not headed by CPs, I take this kind of TP movement to be unproblematic.

[7] Haegeman (2003) uses the term event-conditionals as opposed to premise-conditionals. Premise-conditionals cannot be paraphrased by LSand-constructions without losing the specific premise-conditional reading completely (if the clause is grammatical at all).

(i) a. If (as you claim) we are so short of teachers, we'd better send our children to Germany to be educated.
b. *We are so short of teachers and we'd better send our kids to Germany to be educated.

This follows under the present analysis if, as argued by Haegeman (2003), premise-conditionals are merged very high up in the structure (i.e. above TP). Hence, the position in the specifier of &P is not available for them.

A TP is merged as an adjunct to vP where clauses may receive conditional interpretation. Then, matrix T is merged building matrix TP (i.e. TP_2). In a next step, the coordination head & is merged taking TP_2 as a complement. Then, TP_1 moves out of TP_2 into the specifier of the coordination head. The whole coordination is then embedded either under a matrix C or a subordinate CP just as if we were dealing with a normal coordinate TP. The result is a complete $_{LS}$and-construction.

4.2. Deriving the Properties

In the previous section, we have seen how the analysis is supposed to work. The idea is simple: The conditional TP is base merged as a subordinate adjunct and then it is moved to the specifier of a coordination phrase yielding a structure which is, linearly practically indistinguishable from a coordination of two independent TPs. But how does this analysis derive the puzzling properties of the $_{LS}$and-construction?

4.2.1. Coordinative Properties

Let us have a look at the coordinative properties first. They are pretty straightforward. In contrast to the approach that Culicover and Jackendoff (1997) rejected, *and* is not a subordinator. Here it is the normal coordinator and thus it must always appear between the two conjuncts. Second, it is clear how we may derive the fixed order of both clauses. We have seen that the conditional clause must always precede the main clause. Here is why: The &-head used with the $_{LS}$and-construction is the exact same &-head we use with normal coordination. Hence, the first conjunct must always precede the second one because the specifier of the &-head is always on the left. Adjuncts may appear on the left as well as on the right of their hosts, specifiers may not. They obligatorily precede their respective heads, at least in English.

In other words: The coordinative properties of the $_{LS}$and-construction derive from the fact that the coordinative head is linearized as follows:

(17) Spec&P (i.e. first conjunct) > & > Comp&P (i.e. second conjunct).

Thus, we correctly predict a sentence like **and I'll leave, you drink one more can of beer* as well as **I'll leave, you drink one more can of beer and* to be ungrammatical. Furthermore, it follows that a sentence like *I'll leave and you* *drink one more can of beer* cannot paraphrase a conditional clause like *If you drink one more can of beer, I'll leave.*

4.2.2. Tense Restrictions

$_{\text{LS}}$and-constructions paraphrase conditional clauses. However, as we have seen in section 2, they cannot paraphrase all kinds of conditionals. They are restricted to present tense realis conditionals. Thus irrealis conditionals like the following cannot be paraphrased by $_{\text{LS}}$and-constructions:

(18) a. If you found her number, you would give her a call.
b. If you had found her number, you would have given her a call.

I take this restriction on $_{\text{LS}}$and-constructions to follow from the likeness condition the coordination head imposes on its arguments. If you try to turn the examples in (18) into $_{\text{LS}}$and-constructions, you get the following:

(19) a. *You found her number and you would give her a call.
b. *You had found her number and you would have given her a call.

In (19) in both cases, the respective conjuncts differ at least in mood and tense. I assume that these differences are grave enough to rule out the whole construction. If you minimize the difference (by using the same mood) and provide a possible context, the result is much better.

(20) [Context: I am about to call Sam and ask him whether he found her number]
He has found her number and you give her a call.

Now I turn to the subordinative properties of the $_{\text{LS}}$and-construction. It is clear that these cannot derive from linearization or other surface-oriented processes but rather must derive from processes which apply early on in the derivation, namely as long as the first TP is still a subordinate adjunct.

4.2.3. Binding

We have seen in section 2 that facts about licensing of anaphors and variable binding suggest that the first clause of $_{\text{LS}}$and-construction is in fact subordinate, at least in some sense because, just as with subordinate conditionals, the first conjunct of an $_{\text{LS}}$and-construction may contain an anaphor or a variable which is bound by an expression in the second conjunct.

In traditional binding theory, anaphors are always subject to the same constraint, namely Principle A of binding theory. Principle A roughly states that an anaphor must be bound within some locality domain by some antecedent. According to Belletti and Rizzi (1988) and much subsequent work, Principle A has the status of an "anywhere condition" which means that it can be satisfied at any step of the derivation.

This essentially derives why the first conjunct of an $_{\text{LS}}$and-construction behaves completely identical to subordinate conditionals with respect to binding of anaphors and variables. Conditional clauses and $_{\text{LS}}$and-constructions are base generated in the same position in the tree. In that position, all anaphors and variables can be bound and hence, Principle A is satisfied for the rest of the derivation, regardless of whether the conditional clause stays in situ (as in if-clauses) or whether it moves higher up in the tree (as in $_{\text{LS}}$and-constructions). Take a look a the example in (8-a), repeated in (21).

(21) Another picture of himself appears in the newspaper and John will definitely go and get a lawyer.

As long as *Another picture of himself appears in the newspaper* is in its base position as an adjunct to the vP of the matrix clause, the anaphor can be bound by the matrix subject *John*:[8]

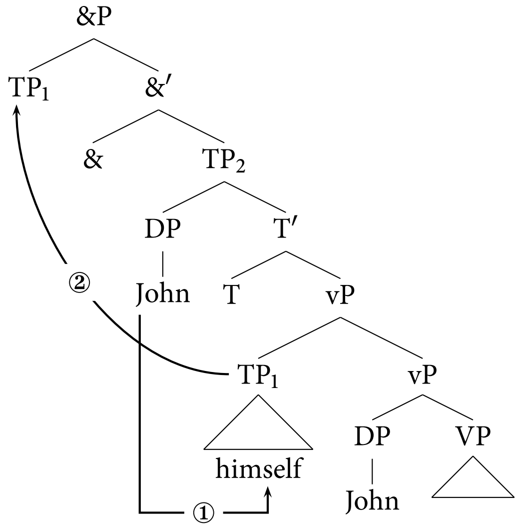
---

[8] In the example in (21) it looks as if mere c-command is sufficient to derive the data which show binding into a vP adjunct. However, it should be noted that this cannot be the whole story as there are examples where an embedded subject may also bind into a vP-adjunct. Also, we also do not encounter a Principle C effect in example (i-c).

(i) a. Another picture of him(self) appears in the newspaper and Susan thinks John will definitely be offended.
b. Another picture of him(self) appears in the newspaper and an early retirement will begin to appeal to John. (Culicover and Jackendoff 1997: 202)
c. Another picture of John appears in the newspaper and he will be offended.

Culicover and Jackendoff (1997) speculate that there is an additional requirement of a logophoric relation that enables to establish a binding relation. They take this as an argument that the whole binding theory should be dealt with in the semantics. However, this is not the only way to go as Landau (2001) and others have shown that logophoric relations can and maybe even should be dealt with in the syntax. I am taking a more moderate stand here in that I am saying that the effects of binding theory are partly syntactic and partly semantic. If,

(22)



The fact that after binding has applied the whole TP₁ is moved to the specifier of the &P does not change anything about the binding relation.

4.2.4. *Gapping*

The next property which suggests that the first conjunct of an LSand-construction is subordinate is Gapping. LSand-constructions and subordinate conditionals cannot undergo Gapping whereas normal coordinate clauses can (cf. (7)). In the literature about Gapping, there is no agreement as to whether Gapping should be derived via ellipsis (Sag 1976, Hankamer 1979, Jayaseelan 1990, Hartmann 2000, Coppock 2001, Lin 2002 etc.) or ATB-movement (cf. Johnson 1996, 2009). However, regardless of the respective type of analysis, the more recent approaches to Gapping (e.g. Coppock 2001, Lin 2002, Johnson 1996, 2009) seem to agree that a prerequisite for Gapping is relatively low coordination, namely vP-coordination. This, among other things, accounts for scope effects where negation and modals take scope over the whole coordination. If, however Gapping can only apply with vP- or VP-coordination, it is not surprising that it is impossible with LSand-constructions which, as we have seen, involve TPs.

---

however, it turned out that we are dealing with a case of semantic binding here, this would still be compatible with the present analysis.

4.2.5. *Right Node Raising*

The same explanation, however, cannot be extended to the ban on Right Node Raising in ${}_{LS}$and-constructions (cf. example (6)). Right Node Raising is generally possible with TPs or even bigger categories:

(23) He said that I saw _ and that Bob didn't see [the man in the long black coat].

Analog to the case of Gapping, it is not clear whether Right Node Raising should be derived via ellipsis or via ATB-movement. Both approaches have been proposed in the literature (see e.g. Wexler and Culicover 1980, Kayne 1994, Wilder 1997, Hartmann 2000 for ellipsis and e.g. Ross 1967, Hudson 1976, Postal 1998, Sabbagh 2007, 2008 for ATB-movement.)[9].

The fact that ${}_{LS}$and-constructions cannot undergo Right Node Raising might speak in favor of the movement approach because if we take Right Node Raising to be an instance of ATB-movement, we have an explanation why it cannot apply to ${}_{LS}$and-constructions. Consider the following situation in (24).

(24) $[_{\&P}\ \mathrm{TP}_1\ [_{\&'}\ \&\ [_{\mathrm{TP}2}\ \mathrm{T}\ [_{vP}\ t_1\ v\mathrm{P}\ ]\ ]\ ]\ ]$

We have a coordinative head & with both its arguments and one of these arguments was merged via internal merge (that is, (24) is an ${}_{LS}$and-construction). Suppose that both TPs do have the same object and hence we want to apply ATB-movement to the right, then we would have to move out of a moved category (i.e. TP_1). However, there are many cases reported in the literature where moved categories are frozen (i.e. no longer transparent for extraction) once they have reached in their final landing site. (On freezing in general see Wexler and Culicover 1980, Browning 1991, Collins 1994, Müller 2010.) There has been an extensive discussion about whether certain structural positions are criterial positions in the sense of Rizzi (2004), which means that movement of an XP into these positions renders that XP intransparent (for a short overview of this discussion see Müller 2010). I will discuss the question of freezing with respect to this analysis of the ${}_{LS}$and-construction in a later sec-

---

[9] It should be noted that there are other approaches to Right Node Raising, e.g. by means of multidominance (Wilder 1999, Abels 2004). However, I will not discuss them here.

tion in more detail but for our purposes here, it is sufficient to note that the specifier of an &-head seems opaque for extraction if it was moved there.[10]

So, in a nutshell, if we adhere to the movement analysis of Right Node Raising, it is to be expected that $_{\text{LS}}$and-constructions cannot undergo Right Node Raising. Before the first conjunct TP is moved to Spec&P, it is in an subordinate position from where Right Node Raising is generally prohibited (cf. (6-c)) After the TP is moved, it is frozen. Hence, further movement out of this TP is prohibited. And since Right Node Raising is, by assumption, movement, it cannot apply to $_{\text{LS}}$and-constructions.

4.2.6. Extraction

In section 2, we saw that $_{\text{LS}}$and-constructions show a very strange behaviour with respect to extraction: They behave neither like subordinate clauses nor like coordinate clauses. First, even though they seem to be coordinate on the surface, they cannot undergo ATB-movement. Second, one may asymmetrically move out of the first conjunct (i.e. the former adjunct clause) and third, one may also move out of the second conjunct (i.e. the underlying main clause). I will provide answers for each of these three properties step by step.

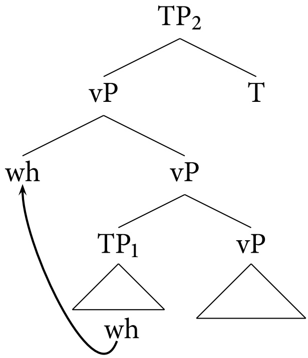
The answer to the question why $_{\text{LS}}$and-constructions cannot undergo ATB-movement has already been given above. It was shown that $_{\text{LS}}$and-constructions cannot undergo Right Node Raising if Right Node Raising is derived via rightward ATB-movement. The reason was that this movement tries to extract out of a frozen category, namely the first conjunct TP. And if rightward ATB-movement results in ungrammaticality, then leftward ATB-movement does so for the exact same reasons.

Second, we must face the question why we can extract out of the first conjunct, i.e. the former subordinate clause. In the paragraphs on Right Node Raising and ATB-movement above, it was shown that movement out of a moved category is ungrammatical. Asymmetrical extraction out of the first conjunct cannot take place after the TP has already moved to Spec&P because it would encounter the same problems as with ATB-movement. Hence, the crucial movement step must have applied earlier, namely as long as the TP is in its base position as an adjunct to vP. The tree in (25) illustrates the situation:

---

[10] Note, however, that the specifier of an &-head is not opaque per se since there are, of course, cases where one can easily move out of an XP which is base generated in Spec&P, for example via ATB-movement.

(25)



In its base position, TP_1 is adjoined to the vP of the matrix clause. It is in that position, that movement of the wh-pronoun out of the adjunct applies. If it applied later in the derivation, this step would encounter a violation of the Coordinate Structure Constraint, but at this point, there is no coordinate structure present (yet). The immediate question that this step raises is, of course, why this movement does not violate the Condition on Extraction Domains (Huang 1982), which usually prohibits movement operations out of (among other things) adjuncts. However, as has been shown by Taylor (2007) for English as well as Etxepare (2002) for Spanish and Yoshida (2006) for Japanese, if-clause conditionals can be transparent for extraction when they precede the matrix clause. This is illustrated in (26) for English:

(26) a. [Which car $]_1$ does Michelle believe if she buys t_1, her insurance premium will increase?

b. ?This is [the kind of car $]_1$ that if Michelle buys t_1, her insurance premium will increase. (Taylor 2007: 189)

This is exactly the same configuration as with the first clause of LSand-constructions in their base position: Left-adjoined event-conditionals. Even though the general question why the sentence-initial if-clause conditionals do not obey the CED (at least in some cases) still lacks a satisfactory answer, the data in (26) strongly support the view that extraction out of the first clause of an LSand-construction may apply as long as it is in its base position. Hence, we may extract an element out of TP_1 as long as TP_1 is in its base position.[11]

---

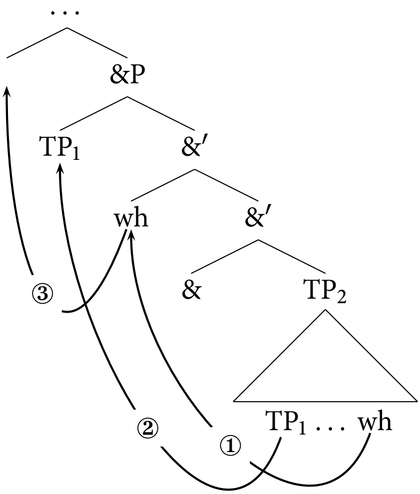
[11] The derivation that involves asymmetric extraction out of the first conjunct of an LSand-construction is highly reminiscent of the "chain interleaving" derivation in Collins (1994). However, in contrast to the case there, we want the chain interleaving derivation to converge

This brings us to the final question, namely how we can extract out of the second conjunct, i.e. TP2. Let us, for explanatory purposes, assume a simplified version of the Coordinate Structure Constraint like the following:

(27) In a structure [A [$_{\&'}$ & B]], movement out of either A or B is prohibited.

If we take (27) as a basis, we might envision a possible derivation which derives an LSand-construction with extraction out of the second conjunct but still does not violate the Coordinate Structure Constraint as formulated in (26). It is given in (28) below:

(28)



After TP2 is merged as the complement of the &-head, we do have two possibilities as how to proceed. The first possibility is to move TP1 out of TP2 into the specifier of &P. Once this is done, a complete coordinate structure is generated and subsequent movement of the wh-phrase out of TP2 would violate the CSC as formulated above (27). Hence, this possibility leads to a crash of the

---

since this asymmetric extraction generally seems to be possible. To exclude the convergence of his derivations as in **Who did a book about cause a scandal?*, Collins (1994) invokes a transderivational economy constraint, a quite controversial concept itself. At this point, I have no satisfying explanation for why the chain interleaving derivation is prohibited in the examples by Collins (1994) but allowed in the case of LSand-constructions.

derivation. There is another possibility which is to move the wh-phrase into an intermediate position[12] (= step 1 in (28)) and subsequent movement of TP_1 (= step 2 in (28)). Since step 1 does not complete the &P (probably because the &-head contains another selectional feature to merge the first conjunct), the CSC does not yet apply. Hence step 2 is still licit. The final step now is to further move the wh-pronoun to a higher position in the tree (= step 3 in (28)). This step also does not violate the CSC because at that point the wh-phrase is no longer part of one of the conjuncts. But since the Coordinate Structure Constraint also prohibits movement of a whole conjunct, it must be ensured that the wh-phrase does not count as a conjunct for the CSC while being in the specifier of the &P. The reason for that might be that the wh-phrase itself contains features which signal that it is not a proper conjunct. Another possibility might be the fact that the wh-phrase has been attracted by a category-neutral edge feature (and not by a c-selectional feature like a proper conjunct) on the &-head shows that the wh-phrase is not a proper conjunct and hence may be moved out without violating the CSC.

A final note about the derivation in (28) is in order. In (28), we successfully circumvented a violation of the CSC but still managed to extract out of only one conjunct. The immediate question that comes to mind is, of course, why this kind of derivation is not possible with "normal" clausal coordination. If this derivation was possible with all kinds of clausal coordination, it would always be possible to circumvent the CSC. This would be equivalent to saying that one could always move out of the second conjunct of a coordination. Hence, we do have to distinguish normal coordination and coordination in case of $_{\text{LS}}$and-constructions. With normal coordination circumventing the CSC is impossible, with $_{\text{LS}}$and-constructions, it is possible. I argue that the difference lies in the fact that the first conjunct of $_{\text{LS}}$and-constructions comes about via Movement. With normal coordination, the first conjunct is the result of (External) Merge. To be more concrete, I assume a well-known principle that regulates the order of operations and distinguishes these two syntactic

---

[12] I assume here that the wh-phrase must cyclically move through the specifier of the coordination phrase. This is basically equivalent to saying that &P is a phase (cf. Reich 2007 for the same assumption).

operations: The Merge-over-Move Principle (Chomsky 1995, 2000, Castillo et al. 2009).[13,14]

(29) Merge over Move (MOM):
If, at some point of the derivation, Merge and Move can both apply, then Merge always applies first.

The exact point of the derivation which is of interest here is when the &-head and TP_2 have been merged. As I illustrated above, with $_{LS}$and-constructions, we do have two possibilities. Either we first move TP_1 and then we move the wh-phrase (which leads to a violation of the CSC) or we move the wh-phrase first and then we move TP_1 (which leads to a circumvention of the CSC). Here, the two operations in competition are two instances of Movement. Hence, the MOM does not regulate their relative order, both orders are possible.

So let us imagine a similar situation with "normal", clausal coordination. We have merged the &-head with the second conjunct. If the second conjunct contains a wh-phrase, we also have two possibilities of how to proceed. Either we first move the wh-phrase and then we merge the first conjunct (which would circumvent the CSC) or we first merge the first conjunct and only then we move the wh-phrase out of the second conjunct (which violates the CSC). Here, the two operations in competition are one instance of Merge and one instance of Move. Hence the MOM forces Merge to apply first. However, if Merge applies first, we encounter a CSC violation in the second step. So, asymmetric extraction out of the second conjunct of normal coordinate clauses either violates the CSC or the MOM and is hence prohibited.

---

[13] I leave the fact aside that in many recent publications, Merge and Move are assumed to be instances of the same operation: Move is to be seen as Internal Merge while Merge is External Merge. I take it that even though these two operations share a common label nowadays, it must be possible to distinguish them at least at a certain point of the derivation.

[14] Note that in the way the MOM is formulated here, it is a transderivational constraint as originally implemented by Chomsky (1995, 2000). However, the version in (29) can be easily implemented without making use of transderivational constraints. In its original version (Chomsky 1995), the MOM was supposed to distinguish between two distinct derivations, one which makes use of Merge and one which makes use of Move. These derivations compete because both Merge and Move may satisfy the crucial feature. In the present case, there is only one derivation in which both operations are applied (because they are induced by two distinct features) and the MOM only restricts the order of their application. Such ordering may be implemented via ordering of operation-inducing features as in Müller (2010), Georgi (2013). Thus, no transderivational constraints are necessary.

4.3. Interim Summary

In the preceeding section, I proposed a new approach to the puzzling $_{\text{LS}}$and-construction, which has led Culicover and Jackendoff (1997) to question one of the fundamental assumptions of the Minimalist Program and related syntactic theories, namely that the semantics of an utterance is calculated on the basis of its syntactic structure. I have argued that it is possible to derive the $_{\text{LS}}$and-construction including all of its puzzling properties and still adhere to Minimalist assumptions.

The analysis I proposed was based on the novel idea that a phrase is base-generated as an adjunct and then moved to the specifier of a coordination phrase. In doing so I was able to derive the fact that $_{\text{LS}}$and-constructions combine subordinative and coordinative properties. In particular, I showed that the analysis is able to account for the properties of $_{\text{LS}}$and-constructions with respect to surface word and clause order, binding of variables and anaphors, Gapping, Right Node Raising and extraction patterns.

In the next section, I want to discuss some open questions and issues of this analysis, but before that, let me just briefly add a note about all the analyses of specific properties above. Some of the specific properties of $_{\text{LS}}$and-constructions such as the binding facts, for example, basically fell out of the analysis I proposed. Others, such as the analyses for Right Node Raising or extraction, needed some additional assumptions. However, it should be emphasized that none of these additional assumptions is crucial for my analysis. If, for example, it turned out that Right Node Raising is better analysed as ellipsis instead of ATB-movement as I assumed, one will have to go one step back and look deeper into existing analyses of RNR as ellipsis and see whether they predict that RNR should be applicable to $_{\text{LS}}$and-constructions or not. The only assumption which is crucial for the present approach is that a movement operation may turn an adjunct into the first conjunct of a coordination phrase.

5. Open Issues and Questions

5.1. Triggering Movement to Spec&P

Up to this point I have been neglecting the question what triggers the movement of the adjunct to Spec&P in the first place. There are certainly several, purely technical solutions as to how this kind of movement can be implemented in a standard Minimalist model. But the solution that I want to sketch here maybe captures one's intuition about the construction as such, namely that this movement of the adjunct might be some kind of Last Resort phenomenon.

$_{\text{LS}}$and-constructions are special with respect to their categorial status. Usually adjuncts are either said to be full finite clauses, i.e. CPs, or they are reduced clauses, which are often assumed to be vPs, at least in English. In both cases, adjuncts are typically phasal categories in the sense of Chomsky (2001) and many others. In the case of $_{\text{LS}}$and-constructions, the adjunct in question is, as we have seen in section 2, a TP, which is generally assumed to be a non-phasal category. Maybe this categorial difference also accounts for the fact that these TP adjuncts must move to Spec&P because they are not licensed as adjuncts. If adjuncts must have phasal status for reasons of interface requirements, then these TP adjuncts are doomed to ungrammaticality unless they move to Spec&P because in that position they no longer count as subordinate adjuncts. This explanation may find support in the fact that the movement in question is not possible with CP adjuncts for example.

(30) I don't think that he has one failed attempt and that he gets kicked out.
≠
I don't think that if he has one failed attempt, then he gets kicked out.

The mechanism I sketched in section 4 is, in principle, not restricted to TPs. So far, there is nothing that precludes CPs or vPs to undergo the same process. However, if the whole movement to Spec&P is modeled as a Last Resort process to avoid the ungrammaticality of TP-adjuncts, we would have an explanation for the non-existence of $_{\text{LS}}$and-constructions with CP-, and vP-coordination.

This process could be implemented into the feature specification of the coordinative &-head as an optional feature which attracts an XP and moves it to its specifier if it cannot be licensed elsewhere, i.e. if it cannot be transfered to the interfaces because of its non-phasal status. Another possibility would be to implement the movement of the adjunct as a repair-driven movement induced by local optimization processes (as proposed by Heck and Müller 2000).[15]

---

[15] Note however that this solution must, like all implementations as Last Resort phenomena,

5.2. Freezing

In this section, I want to briefly adress the notion of “freezing” and the way it is to be understood throughout this paper. In the discussion above the concept of freezing played a role in two unrelated derivations. First, it was of importance in the discussions about ATB-movement and Right Node Raising. $_{LS}$and-constructions cannot undergo ATB-movement or Right Node Raising (under the assumption that RNR is an instance of rightward ATB-movement) because previous movement of the TP into the specifier of the &-head freezes all elements within that TP. Hence, the concept of freezing was needed to exclude examples like (6-b), repeated in (31).

(31) *Big Louie finds out about _ and Big Louie puts a contract on _, that guy who stole some loot from the gang.

Second, the concept of freezing played a role in the derivations of asymmetrical extraction from the left conjunct. On the surface, a sentence like (11-a), repeated in (32), also looks like it violates the freezing principle because the extracted DP *loot* c-commands its trace/copy/base position which itself is contained in a moved category (TP_1).

(32) ?This is the loot that [$_{TP1}$ you just identify _] and [$_{TP2}$we arrest the thief on the spot].

The difference between these two derivations is that in the second derivation, no actual extraction out of a moved category has applied. However, to capture this difference is probably not that easy. A representational version of the freezing principle like the one in (33), for example, cannot distinguish the cases in (31) and (32). It correctly predicts (31) to be ungrammatical but it also rules out (32).

(33) A trace t may not be included in a moved XP (i.e., an XP that binds a trace) if the antecedent of t c-commands XP. (Müller 1998)

Instead of (33), the freezing principle is, at least for my purposes better formulated in a derivational way without reference to traces or other representa-

---

make use of violable constraints and/or transderivational competition, at least to a certain degree.

tional concepts. I informally did so throughout the paper by just stating the following:

(34) Movement out of a moved category is prohibited.

This derivational principle, as it stands, can distinguish between the ATB-case where actual movement out of a moved category has applied and the asymmetrical extraction case where the extraction has applied before the subsequent remnant movement of the TP.

5.3. Rule Interaction

As this volume collects various articles which involve opaque and transparent interactions of rules and constraints, I want to focus on this issue a little bit more. As far as I can see, there are two interesting types of rule interaction which play a role in this paper.

First, there is an interaction between the general movement step which moves the adjunct to Spec&P and several processes that make use of the subordinate properties of the $_{\text{LS}}$and-construction such as binding. Since the construction is, on the surface, practically indistinguishable from a case of normal TP coordination, it may be surprising that an element in the first conjunct can be bound by an element in the second one. It is, however, no longer surprising when looking at the derivational history of the construction. Since Principle A of the Binding Theory is an anywhere principle, it could apply before the movement of the adjunct to Spec&P destroyed its context for application. If this movement step had applied before binding, this would have bled the creation of binding relations. The same holds for other subordinative properties such as the fact that one may asymmetrically extract out of either conjunct. If movement to Spec&P applied before extraction, the context would be destroyed as subsequent extraction would violate the Coordinate Structure Constraint. However, since both extraction as well as binding may apply before movement, this is, in the terms of Kiparsky (1976), an instance of counter-bleeding.

The second kind of rule interaction is found with a special derivation involving asymmetric extraction out of the right conjunct.[16] As I discussed in

---

[16] Since asymmetric extraction out of the left conjunct is accomplished by moving through

the section about extraction in detail, there is one step of the derivation where it faces a dilemma how to proceed. The situation is given in (35).[17]

(35) $[_{\&'}$ & $[_{\mathrm{TP2}}$ T $[_{\mathrm{vP}}$ TP_1 $[_{\mathrm{vP}}$ wh vP]]]]

We have two specifiers to the vP, namely the wh-phrase and the adjunct clause (TP_1). Both are to be promoted to the specifier of the &P. For the wh-element it is another intermediate movement position on its way to SpecCP and for the TP_1 it is the final landing site. Since these two operations are both movement steps, the Merge-Over-Move Principle discussed in section 4 does not require them to be ordered. Hence, in principle, they can apply in either order but, as we have seen, only one order leads to a converging derivation. If TP_1 moves first, the coordination is complete and subsequent movement is bled because it would violate the Coordinate Structure Constraint. If, however, the wh-element moves first, the coordination is not yet complete and subsequent movement of TP_1 may apply. This may be considered another counter-bleeding relation between the two movement steps.

The first counter-bleeding relation between movement of the adjunct and applicance of binding relations follows from derivational principles such as the Strict Cycle Condition (e.g. Chomsky 1973) without further ado and may be thus viewed as intrinsic. The second rule interaction between the two types of movement (intermediate and final) can generally not be ordered by such general principles. Here, I assumed that both orders are, in principal, possible but only one of them (intermediate precedes final movement) results in a grammatical structure.

It should be noted that Georgi (2013), whose account argues for ordering between intermediate and final movement operation, actually predicts that final movement must precede intermediate movement for reasons of specificity, which is the wrong order for my purposes. Whether this means that the gap in

---

some intermediate landing sites contained within the right conjunct, this rule interaction applies to extraction out of the left conjunct as well.

[17] I have not been explicit about whether successive cyclic movement passes through a position above or below the adjunct clause, or whether a principle such as Relativized Minimality regulates the order in which the specifiers are attracted. However, as far as I can see, there is no problem either way. If one adopts Relativized Minimality, one needs a further stipulation, namely that adjunct clauses occupy specfiers below positions that are created by intermediate movement steps. In a system without Relativized Minimality where both options are possible or where both movement types are sufficiently different not to be affected by Relativized Minimality, no additional assumptions are necessary.

the typology of rule interactions she argues for does not exist or whether the analysis of extraction I proposed is on the wrong track remains an interesting question for further research.

6. Conclusion

The present analysis of the $_{\text{LS}}$and-construction can be taken as an argument in favor of two assumptions. First, it presents a new argument for the hypothesis that coordination is in fact an asymmetric construction with the first conjunct asymmetrically c-commanding the second. In a GB model of coordination with ternary branching, an analysis which turns an adjunct into the first conjunct of a coordination by means of movement would probably not be possible. Previous arguments for the asymmetric coordination hypothesis were based either on theory-internal considerations or empirical findings concerning morphological agreement. The argument put forward in this paper was novel in the sense that it has shown that the asymmetric coordination hypothesis may also have syntactic consequences, for example, that the specifier may be a proper landing site for movement processes.

Secondly and maybe more importantly, the present analysis supports the standard Minimalist assumption that the semantics of a sentence is computed on basis of its syntactic structure. It does so by invalidating the claim made by Culicover and Jackendoff (1997) that the so-called left-subordinating-and construction is incompatible with this very assumption and cannot be derived in a standard Minimalist framework. However, as I showed, in a derivational model of syntax being coordinate and being subordinate does not necessarily exclude each other. A subordinate phrase may be turned into a first conjunct by means of movement if certain semantic and syntactic requirements are met. This can account for the fact that the $_{\text{LS}}$and-construction looks like a normal coordinate clause on the surface even though most of its syntactic properties resemble those of a subordinate clause.

On a more abstract level, this analysis can be seen as one possibility to account for syntactic constructions which combine syntactic properties of subordinate and coordinate clauses. In Weisser (2013), it was shown that an analysis that makes use of the same abstract movement step derives the mixed behaviour between subordination and coordination found with a number of clause-chaining constructions in the languages of Papua New-Guinea and other parts of the world.

References

Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Conneticut, Storrs.

Abels, Klaus (2004): Right Node Raising: Ellipsis or Across the Board Movement. *In:* K. Moulton and M. Wolf, eds, *Proceedings of NELS 34*. Amherst, MA: GLSA.

Belletti, Adriana and Luigi Rizzi (1988): 'Psych-Verbs and θ-theory', *Natural Language and Linguistic Theory* 6, 291–352.

Browning, Margareth (1991): 'Bounding Conditions on Representations', *Linguistic Inquiry* 22, 541–562.

Castillo, Juan Carlos, John Drury and Kleanthes Grohmann (2009): 'Merge Over Move and the Extended Projection Principle: MOM and the PP Revisited ', *Iberia: An International Journal of Theoretical Linguistics* 1.1, 53–114.

Chao, Wynn (1988): *On Ellipsis*. Garland, New York.

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

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and 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.

Citko, Barbara (2004): 'Agreement Asymmetries in Coordinate Structures', *Proceedings of FASL 12* .

Collins, Chris (1994): 'Economy of Derivation and the Generalized Proper Binding Condition', *Linguistic Inquiry* 25, 45–61.

Coppock, Elizabeth (2001): Gapping: In Defense of Deletion. *In:* M. Andronis, C. Ball, H. Elston and S. Neuvel, eds, *Proceedings of Chicago Linguistic Society 37*. University of Chicago, Chicago Linguistic Society, pp. 133–148.

Culicover, Peter (1970): 'One More Can of Beer', *Linguistic Inquiry* 1, 366–369.

Culicover, Peter (1972): 'OM-sentences', *Foundations of Language* 8, 199–236.

Culicover, Peter W. and Ray Jackendoff (1997): 'Semantic Subordination despite Syntactic Coordination', *Linguistic Inquiry* 28, 195–217.

Etxepare, Ricardo (2002): 'Null Complementizers in Spanish', *International Journal of Basque Linguistics and Philology* XXX-2, 469–496.

Georgi, Doreen (2013): Opaque Interaction of Internal Merge and Agree. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 413–461.

Haegeman, Liliane (2003): 'Conditional Clauses: External and Internal Syntax', *Mind & Language* 18(4), 317–339.

Hankamer, Jorge (1973): 'Unacceptable Ambiguity', *Linguistic Inquiry* 4, 17–68.

Hankamer, Jorge (1979): *Deletion in Coordinate Structures*. New York, Garland.

Harbert, Wayne and Maher Bahloul (2002): Postverbal Subjects in Arabic and the Theory of Agreement. *In:* J. Ouhalla and U. Shlonsky, eds, *Themes in Arabic and Hebrew Syntax*. Studies in Natural Language and Linguistic Theory, Springer Netherlands, pp. 45–70.

Hartmann, Katharina (2000): *Right Node Raising and Gapping*. Amsterdamn, John Benjamins.

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

Huang, Cheng-Teh James (1982): Logical Relations in Chinese and the Theory of Grammar. PhD thesis, MIT, Cambridge Mass.

Hudson, Richard (1976): 'Conjunction Reduction, Gapping and Right Node Raising', *Language* 52, 535–562.

Jayaseelan, Karattuparambil A. (1990): 'Incomplete VP Deletion and Gapping', *Linguistic Analysis* 20, 64–81.

Johannessen, Janne Bondi (1998): *Coordination*. Oxford University Press, USA.

Johnson, Kyle (1996): *In Search of the English Middle Field*. University of Massachuesetts, Amherst.

Johnson, Kyle (2009): 'Gapping is not (VP-)Ellipsis', *Linguistic Inquiry* 40, 289–328.

Kayne, Richard S. (1994): *The Antisymmetry of Syntax*. MIT Press.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Koppen, Marjo van (2006): One Probe, Multiple Goals: The Case of First Conjunct Agreement. *In:* M. van Koppen, ed., *Special Issue of Leiden Papers in Linguistics 3.2*. Leiden University.

Lakoff, George (1986): 'Frame Semantic Control of the Coordinate Structure Constraint', *CLS 22 - The parasession on Pragmatics and Grammatical Theory* pp. 152–167.

Landau, Idan (2001): 'Control and Extraposition: The Case of Super-Equi', *Natural Language and Linguistic Theory* 19, 109–152.

Lin, Vivian (2002): Coordination and Sharing at the Interfaces. PhD thesis, MIT, Cambridge MA.

Müller, Gereon (2010): 'On Deriving CED effects from the PIC', *Linguistic Inquiry* 41, 35–82.

Müller, Gereon (1998): *Incomplete Category Fronting - A Derivational Approach to Remnant Movement in German*. Kluwer, Dordrecht.

Munn, Alan (1999): 'First Conjunct Agreement: Against a Clausal Analysis', *Linguistic Inquiry* 30(4), 643–668.

Munn, Alan Boag (1987): 'Coordinate Structure and X-bar theory', *McGill Working Papers in Linguistics* 4(1), 121–140.

Munn, Alan Boag (1993): *Topics in the Syntax and Semantics of Coordinate Structures*. Doctoral Dissertation: University of Maryland.

Murguia, Elixabete (2004): Syntactic Identity and Locality Restrictions on Verbal Ellipsis. PhD thesis, University of Maryland.

Neijt, Anneke (1979): *Gapping*. Dordrecht, Foris.

Pesetsky, David (1982): On Paths and Categories. PhD thesis, Massachusetts Institute of Technology.

Postal, Paul (1998): *Three Investigations of Extraction:*. Current Studies in Linguistics Series 29, Cambridge, MA: MIT Press.

Reich, Ingo (2007): From Phases to ATB-Movement. In: *Proceedings of CLS 43*. Chicago Linguistic Society.

Rizzi, Luigi (2004): On the Form of Chains: Criterial Positions and ECP Effects. *In:* L. Cheng and N. Corver, eds, *On WH movement: Moving on*. MIT Press, Cambridge Mass., pp. 97–133.

Ross, John Robert (1967): Constraints on Variables in Syntax. PhD thesis, Massachusetts Institute of Technology. Massachusetts Institute of Technology. Dept. of Modern Languages and Linguistics. Thesis. Ph.D.

Sabbagh, Joseph (2007): 'Ordering and Linearizing Rightward Movement', *Natural Language and Linguistic Theory* 25, 349–401.

Sabbagh, Joseph (2008): 'Right Node Raising and Extraction in Tagalog', *Linguistic Inquiry* 39, 501–511.

Sag, Ivan (1976): Deletion and Logical Form. PhD thesis, Cambridge University.

Takahashi, Hidemitsu (2004): The English Imperative: A Cognitive and Functional Analysis. PhD thesis, Hokkaido University.

Taylor, Heather Lee (2007): Movement from if-Clause Adjuncts. *In:* A. Conroy, C. Jing, C. Nakao and E. Takahashi, eds, *University of Maryland Working Papers in Linguistics 15*. University of Maryland, pp. 189–202.

te Velde, John (2005): *Deriving Coordinate Symmetries: A Phase-Based Approach Integrating Select, Merge, Copy and Match*. John Benjamins: Linguistics Today.

Weisser, Philipp (2013): Dissecting Clause-Chaining Constructions. Ms., Universität Leipzig.

Wexler, Kenneth and Peter Culicover (1980): *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

Wilder, Chris (1997): Some Properties of Ellipsis in Coordination. *In:* A. Alexiadou and T. Hall, eds, *Studies in Universal Grammar and Typological Variation*. Amsterdam, John Benjamins.

Wilder, Chris (1999): Right Node Raising and the LCA. *In:* S. Bird, A. Carnie, J. Haugen and P. Norquest, eds, *WCCFL 18*. Somerville, MA: Cascadilla Proceedings Project, pp. 586–598.

Yoshida, Masaya (2006): Sometimes Smaller is Better: Sluicing, Gapping and Semantic Identity. *In:* C. Davis, A. Deal and Y. Zabbal, eds, *Proceedings of NELS 36*. Amherst:GLSA, pp. 259–287.

Parametrisierter Merkmalsabgleich und das Hauptsondenprinzip

Marie-Luise Popp[*]

Abstract

In this paper, I argue that the analysis on agreement displacement phenomena by Béjar und Řezáč (2009) can only account for agreement in languages such as Basque, but faces a problem with the agreement pattern of the Bolivian language Itonama. In contexts, in which speech-act participants act upon other speech-act participants, Basque agrees with the internal argument, while other languages, such as Itonama, choose agreement with the external argument. I discuss several attempts to extend their theory in order to derive the differences displayed in the data. The final solution I propose includes feature checking under m-command and the parametrisation of Agree. Concretely, I suggest that there are two probes on v: one of them checks its features with the external argument in Spec-v, while the other checks its features with the internal argument under c-command. Since the two Agree operations cannot apply simultaneously, they must be ordered according to language-specific parameters. I show that ordering of Agree operations accounts for the differences between Basque and Itonama concerning agreement.

1. Einführung

In der vorliegenden Arbeit werde ich Daten aus der bolivianischen Sprache Itonama vorstellen, die ein Inverssystem aufweist. Ich werde zeigen, dass die Analyse von Béjar und Řezáč (2009) falsche Vorhersagen für die neuen Daten macht. Bei der Betrachtung von Sätzen, in denen sowohl das interne als auch das externe Argument ein lokales Personenmerkmal tragen, wird deutlich, dass einige Sprachen, wie Itonama, vorzugsweise mit dem externen Argument kongruieren, während andere Sprachen, wie Baskisch oder Georgisch,

---

[*] Die vorliegende Arbeit ist im Rahmen des Bachelor-Seminars *Syntax der Argumentkodierung* entstanden. Ich danke Gereon Müller, Anke Assmann, Daniela Thomas und Daniel Gleim für ihre wertvollen geistigen Anregungen.

Rule Interaction in Grammar, 333–361
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

Kongruenz mit dem internen Argument bevorzugen. Das wird besonders bei dem Unterschied zwischen (1-a) und (1-b) deutlich.[1]

(1) 2>1
a. *Itonama* (Crevels 2010: S. 684)
wase'wa **de'**-kewa-na-he-mo.
yesterday **2PL**-face.see-NEUT-DISTR-1
'Yesterday you all saw us.'
b. *Baskisch* (Béjar und Řezáč 2009: S. 37)
ikusi **n**-ind-u-zu-n.
seen **1**-X-have-2-PST
'You saw me.'

Sowohl im Baskischen als auch im Itonama wird Kongruenz mit dem präferierten Argument mit einem Präfix markiert. In Itonama (1-a) realisiert dieses Präfix die Personenmerkmale des externen Arguments, wohingegen in Baskisch (1-b) die Personenmerkmale des Objekts markiert werden.

Die Theorie von Béjar und Řezáč (2009) macht die Vorhersage, dass jede Sprache im Kontext (1,2)>(1,2) mit dem internen Argument kongruiert. Da die Daten in (1-a) aber deutlich zeigen, dass im Itonama Kongruenz mit dem externen Argument stattfindet, muss die Theorie erweitert werden. Außerdem muss erklärt werden, warum in (1-a) zusätzlich noch mit internen Argument und in (1-b) noch mit dem externen Argument kongruiert wird. Diese sekundäre Kongruenz wird in beiden Sprachen als Suffix realisiert.

Ich werde vorschlagen, dass Merkmalsabgleich unter m-Kommando appliziert und dass es zwei Sonden auf v gibt. Sprachen unterscheiden sich bezüglich ihrer präferierten Sondierungsrichtung: Während im Itonama Merkmalsabgleich mit dem Spezifikator präferiert wird, wird im Baskischen Merkmalsabgleich vorzugsweise unter c-Kommando durchgeführt. Das *Prinzip der Frühzeitigkeit* (Earliness Principle; Pesetsky 1989, Řezáč 2004) verlangt, dass beide Sonden auf v ihre Merkmale so schnell wie möglich abgleichen. Unter der Annahme, dass simultane Applikation von Abgleichsoperationen nicht

---

[1] Die folgenden Abkürzungen werden in diesem Papier verwendet: CAUS – Kausativ, CNT – Kontinuativ, DEM – Demonstrativ, DISTR – distributiv, EXCL – exklusiv, FEM – feminin, HON – Honorativ, INCL – inklusiv, INV – invers, NEUT – neutraler Aspekt, PL – Plural, PST – Past, REP – wiederholte Handlung, SG – Singular, SUB – subordiniert, X>Y – Subjekt Person x, Objekt Person y

möglich ist (siehe Brody 2002, Epstein und Seely 2002), und unter Bezug auf den Unterschied in (1) muss ein Parameter gefunden werden, der entscheidet, welche Sonde ihre Merkmale zuerst abgleichen darf. Aus diesem Grund werde ich das *Hauptsondenprinzip* einführen, welches bestimmt, welche der beiden Sonden die Hauptsonde ist und somit als erste Merkmalsabgleich durchführen darf. Die Hauptsonde kann Abgleich mit dem Argument in ihrem präferierten Suchraum durchführen, während die andere Sonde nur mit Argumenten kongruieren kann, die von der Hauptsonde noch nicht durchsucht wurden. Hauptsondenkongruenz mit dem präferierten Argument *bleedet* also weitere Kongruenzrelationen mit dem präferierten Argument.

Der Artikel ist folgendermaßen gegliedert: In Abschnitt 2 werde ich zunächst die Daten aus Itonama vorstellen. In Abschnitt 3 werde ich die Analyse von Béjar und Řezáč (2009) rekapitulieren und die Probleme der Theorie mit den Daten aus dem Itonama aufzeigen. In Abschnitt 4 werden zunächst verschiedene Lösungsvorschläge unterbreiten und deren Probleme diskutiert. Schließlich werde ich eine Analyse vorschlagen, die auf der Parametrisierung des Merkmalsabgleichs basiert. Abschnitt 5 beinhaltet eine Diskussion über die Regelinteraktionen des hier präsentierten Ansatzes. Es wird gezeigt, dass die sprachspezifische Ordnung der Abgleichsoperationen die Unterschiede zwischen Baskisch und Itonama erklären kann. Abschnitt 6 bietet eine Zusammenfassung der in der Arbeit vorgestellen Probleme und deren Lösungen.

2. Itonama

Itonama ist eine isolierte Sprache, die in Bolivien gesprochen wird. Sie weist einen hohen Synthesegrad und eine verbinitiale Wortstellung auf. Außerdem ist Itonama eine Pro-Drop-Sprache, in der pronominale Affixe am Verb obligatorisch sind, während die vollständige DPs neutral markiert werden (Crevels 2011).

(2)	**si**-yupa-he	warusu	keteno	ma'irï.
	1SG-SOW-DISTR	rice	sugarcane	plantain

‘I sowed rice, sugarcane and plantain.’ (Crevels 2010: S. 684)

In (2) wird die Kongruenz mit dem externen Argument durch den Marker *si-* am Verb markiert, während Kongruenz mit dem internen Argument nicht overt am Verb markiert ist.

Itonama weist ein akkusativisches Kongruenzmuster auf (siehe (3-a) und (3-b)).

(3) a. **si**-yalïs-na.
1SG-be.hungry-NEUT
'I am hungry.'
b. **si**-chudu'-ne-we-'o
1SG-hit-NEUT-2-REP
'I hit you again.' (Crevels 2010: S. 679 & 686)

In (3-a) kongruiert das einzige Argument eines intransitiven Satzes mit dem Verb. Die Kongruenz wird durch den Marker *si-* realisiert, ebenso wie die Kongruenz mit dem externen Argument des transitiven Satzes in (3-b): Die Kongruenz mit dem internen Argument jedoch wird mit dem Suffix *-we* realisiert. Daraus folgt, dass das Kongruenzmuster in Itonama akkusativisch ist.

Des Weiteren weist Itonama in Hauptsätzen ein Inverssystem auf. Direktkontexte (vgl. (4-a)) sind solche, in denen das externe Argument salienter ist, d.h. höher auf der Personenhierarchie (1./2. Person) > 3. Person steht (vgl. auch (Silverstein 1976)).[2] In diesen Kontexten markiert das verbale Präfix Kongruenz mit dem externen Argument. In Inverskontexten hingegen ist das interne Argument salienter. Hier trägt das Präfix am Verb die Personenmerkmale des internen Arguments (siehe (4-b,c)).

(4) a. 2>3
de'-ka-kikïwa'-na yota'-na ubuwa.
2PL-face-know-NEUT DEM-NEUT person
'You all know that person.'
b. 3>1
sih-k'i-ma-doh-ne upa'u.
1PL.EXCL-INV-hand-bite-NEUT dog
'The dog bit us on the hand.'
c. 3>2
kumani **a**'-k'i-pachïhi'-ke kopone.
last.night 2SG-INV-bother-PL rooster
'The rooster was bothering you last night.' (Crevels 2010: S. 685)

---

[2] Die gleiche Hierarchie gilt auch für das Baskische, siehe Abschnitt 3.

In (4-a) wird die Kongruenz mit dem externen Argument durch das Präfix *de'* am Verb realisiert, während das interne Argument nicht overt mit dem Verb kongruiert. In (4-b,c) sind die Präfixe *sih* bzw. *a'* Marker aus dem Possessivparadigma und kodieren Kongruenz mit dem internen Argument. Außerdem tritt in diesen Fällen der Inversmarker *k'i* auf.[3] Das Paradigma der verbalen Präfixe in Direkt- und Inverskontexten ist in den Tabellen 1 und 2 aufgeführt.[4]

Person	**Marker**
1SG	si-
2SG	e'-
2SG.FEM	ke'-
1PL.EXCL	se'-
1PL.INCL	de'-
2PL	de'-

Tabelle 1: Marker in Direktkontexten

Person	**Marker**
1SG	se'-
2SG	a'-k'i-
2SG.FEM	ka'-k'i-
1PL.EXL	sVh-k'i-
1PL.INCL	dVh-k'i-
2PL	dVh-k'i-

Tabelle 2: Marker in Inverskontexten

Bisher wurde aus den Daten deutlich, dass das Präfix am Verb immer Kongruenz mit dem salienteren Argument markiert. Die zentrale Beobachtung für das Itonama betrifft aber Kontexte, in denen beide Argumente gleichermaßen salient sind: Stehen beide Argumente in einer lokalen Person, drückt das Präfix Kongruenz mit dem externen Argument aus. Außerdem tritt in diesen Fällen kein Inversmarker auf.[5]

---

[3] Die einzige Ausnahme bildet der Marker für die 1SG, welcher laut Crevels eine fusionierte Form aus Possessivmarker und Inversmarker ist.

[4] V steht für Vokalharmonie.

[5] Die in (5) aufgeführten Kontexte werden von Crevels sämtlichst als direkte Kontexte bezeichnet. In Abschnitt 4.1 werde ich argumentieren, dass diese Einteilung nicht vollständig korrekt ist.

(5) a. 2>1
wase'wa de'-kewa-na-he-mo.
yesterday 2PL-face.see-NEUT-DISTR-1
'Yesterday you all saw us.'

b. 1>2
si-kamo'-ke-we.
1SG-face.hit-PL-2
'I am going to hit you (in the face).'

c. 3>3
ni-fatima yo-lo'ba-he'-ka t'iyaya.
HON-Fatima CAUS-bathe-DISTR-FEM.SG boy
'Fatima bathed the boy.' (Crevels 2010: S. 684f.)

In (5-a) und (5-b) tragen beide Argumente lokale Personenmerkmale. In diesen Fällen realisiert das Präfix das Personenmerkmal des externen Arguments. Dabei ist es egal, ob das externe Argument in der 1. oder 2. Person steht, d.h. die lokalen Personenmerkmale sind nicht untereinander geordnet. Zusätzlich zur Kongruenz mit dem externen Argument wird in den Fällen (1,2)>(1,2) Kongruenz mit dem internen Argument als Suffix realisiert. Hierbei ist auffällig, dass die Suffixe nur das Merkmal Person kodieren, nicht jedoch den Numerus.

In (5-c) stehen beide Argumente in der dritten Person. Da Personenkongruenz mit der dritten Person nie overt am Verb markiert wird, findet man in (5-c) weder ein verbales Präfix noch ein Suffix, das Person kodiert.

Die unterschiedlichen Kontexte lassen sich folgendermaßen zusammenfassen:

direkt	**invers**
(1,2)>(1,2)	3>(1,2)
(1,2)>3	
Marker aus Tabelle 1	Marker aus Tabelle 2
kein Inversmarker	Inversmarker *k'i*

Tabelle 3: Direkte und inverse Kontexte in Itonama

Über den Kontext 3>3 lässt sich im Moment nichts sagen, da gar kein Argument am Verb markiert wird.

Zusammenfassend lässt sich sagen, dass man aus den Daten in diesem Abschnitt folgendes beobachten konnte: Ist in einem transitiven Satz ein Argument salienter als das andere, realisiert das verbale Präfix die Personenmerkmale des salienteren Arguments. Sind beide Argumente aber gleichermaßen salient, markiert das Präfix Kongruenz mit dem externen Argument. Daraus kann man schließen, dass in Itonama das Verb präferiert mit dem externen Argument kongruiert.

Im folgenden Abschnitt werde ich die Analyse von Béjar und Řezáč (2009) anhand der Daten aus dem Baskischen vorstellen und zeigen, (i) dass Baskisch im Gegensatz zu Itonama bevorzugt mit dem internen Argument kongruiert und (ii) dass die Annahmen von Béjar und Řezáč (2009) nicht ausreichen, um die Daten aus dem Itonama abzuleiten.

3. Zyklische Kongruenz

Um Hierarchieeffekte in verbaler Kongruenz abzuleiten, führen Béjar und Řezáč (2009) die Theorie der *Zyklischen Kongruenz* ein. Diese Hierarchieeffekte zeigen sich, ebenso wie im Itonama, auch im Baskischen in (6) (Béjar und Řezáč 2009: S. 37).

(6) *Baskisch*

a. 1>2
ikusi **z**-in-t-u-da-n.
seen **2**-X-PL-have-1-PST
'I saw you.'

b. 3>1
ikusi **n**-ind-u-en.
seen **1**-X-have-PST
'He saw me.'

c. 2>1
ikusi **n**-ind-u-zu-n.
seen **1**-X-have-2-PST
'You saw me.'

d. 1>3
ikusi **n**-u-en.
seen **1**-have-PAST
'I saw him.'

Aus den Daten in (6) wird deutlich, dass, genau wie in Itonama, die Personenhierarchie bestimmt, mit welchem Argument des Satzes das Präfix des Verbs kongruiert. Es wird stets das Argument als Präfix am Verb markiert, welches in der Hierarchie am höchsten steht (siehe die fettgedruckten Marker in (6)). Die lokalen Personen scheinen untereinander nicht geordnet zu sein, denn sowohl in (6-a) als auch (6-c) kongruiert das Verb mit dem internen Argument.

Béjar und Rezac ordnen den Personen folgende Merkmale zu:

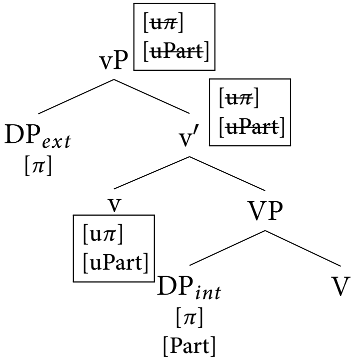
(7) 1. Person: π, Part(izipant), Sprecher
2. Person: π, Part(izipant), Adressat
3. Person: π

Außerdem nehmen sie an, dass auf v eine Sonde mit uninterpretierbaren Personenmerkmalen sitzt. Im Baskischen trägt sie die Merkmale [uπ] und [uPart].[6] Die Sonde kann ihre Merkmale unter c-Kommando abgleichen, siehe (8).

(8) *Merkmalsabgleich unter c-Kommando*
Ein Knoten A kann nicht überprüfte Merkmale [uM] genau dann mit dem Knoten B abgleichen, wenn (i) oder (ii) gelten:
(i) B ist die Schwester von A und trägt ein Merkmal [M].
(ii) B wird von der Schwester von A dominiert und trägt ein Merkmal [M].

Die Sonde sucht zunächst in ihrem c-Kommando-Bereich nach DPs, die ihre Merkmale überprüfen können. Im Baskischen sucht die Sonde also nach 1. oder 2. Person, da diese das Merkmale [Partizipant] und [π] tragen.

(9) 3>(1,2)



---

[6] In Sprachen wie Mohawk, in den 1. und 2. Person untereinander noch hierarchisch geordnet sind, würde die Sonde noch das Merkmal [uSprecher] oder [uAdressat] tragen. Da dies aber weder im Baskischen noch im Itonama der Fall ist, werden diese Merkmale von nun an vernachlässigt.

Ist das interne Argument eine lokale Person, wie in (9), kann DP$_{int}$ beide Merkmale der Sonde überprüfen und das Verb kann mit dem internen Argument kongruieren, wie in (6-b), hier wiederholt als (10):

(10) ikusi n-ind-u-en.
seen 1-x-have-PAST
'He saw me.' (Béjar und Řezáč 2009: S. 37)

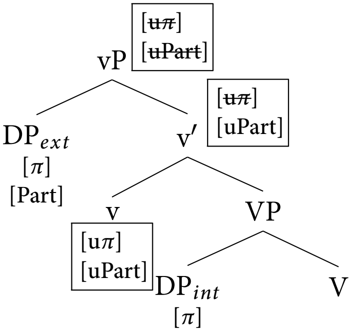
Der in (10) als *x* glossierte Marker tritt immer dann am Verb auf, wenn die Sonde im ersten Suchlauf alle Merkmale abgleichen konnte. Diese Kontexte nennen Béjar und Rezac *invers*, alle anderen Kontexte sind *Direkt*kontexte. Die Kontexte können folgendermaßen zusammengefasst werden:

direkt	**invers**
3>3	3>(1,2)
(1,2)>3	(1,2)>(1,2)
kein Inversmarker x	Inversmarker x

Tabelle 4: Direkt- und Inverskontexte im Baskischen

Ist das interne Argument aber ein Argument der 3. Person, kann im ersten Suchlauf nur das Merkmal [uπ] der Sonde abgeglichen werden. Das noch nicht überprüfte Merkmal [uPart] wird auf den nächsten Knoten weitergegeben, also auf v′, wo die Sonde ihren zweiten Suchlauf starten kann und somit auch in DP$_{ext}$ nach dem Merkmal [Part] suchen kann. Ist das externe Argument eine lokale Person, kann DP$_{ext}$ das Merkmal [uPart] überprüfen, wie in (11) dargestellt. In diesen Fällen kongruiert das Verb mit dem externen Argument, wie in (6-d), hier wiederholt als (12).

(11) (1,2)>3



(12) ikusi n-u-en.
seen 1-have-PAST
'I saw him.' (Béjar und Řezáč 2009: S. 37f)

Wenn beide Argumente in der 3. Person stehen, kann das Merkmal [uPart] gar nicht abgeglichen werden. Béjar und Řezáč legen fest, dass Sondenmerkmale, die auch am Ende der Derivation noch nicht überprüft werden konnten, nicht zum Absturz der Derivation führen.

Die Theorie kann somit ableiten, warum im Baskischen im Kontext (1,2)>3 Kongruenz mit dem externen Argument stattfindet und kein Inversmarker auftritt. Sie kann außerdem erklären, dass in einigen Sprachen, wie bspw. im Georgischen, in Direktkontexten ein anderer Kongruenzmarker gewählt wird als in Inverskontexten (vgl. auch Béjar 2003).

(13) *Georgisch* (Béjar und Řezáč 2009: S. 51)

a. 3>1,
Morphologie des 1. Suchlaufs
m-xedav-s
1SG-see-INV
'He sees me.'

b. 1>3,
Morphologie des 2. Suchlaufs
v-xedav
1SG-see
'I see him.'

Aus (13) wird sichtbar, dass für den ersten Suchlauf ein anderer Marker gewählt wird als für den zweiten Suchlauf. Im Inverskontext in (13-a) tritt das Präfix *m* für die 1. Person sowie der Inversmarker *s* auf. Im Direktkontext in (13-b), in dem die Sonde erst im zweiten Anlauf ihre Merkmale abgleichen kann, tritt ein anderes Präfix, nämlich *v* auf.[7]

Die Theorie macht also die Vorhersage, dass Sprachen zunächst immer versuchen, mit dem internen Argument zu kongruieren. Kongruenz mit dem externen Argument ist nur eine Option, wenn nach Abgleich mit dem internen

---

[7] Béjar und Řezáč (2009) und Béjar (2003) analysieren die Daten aus dem Georgischen aus einer rein syntaktischen Perspektive. Anderson (1992), Halle und Marantz (1993), Trommer (2001, 2003*a*,*b*) und Stump (2001) betrachten die Daten hingegen aus einer morphologischen Sicht. In diesen Analysen können die Merkmale der Affixe bzw. Klitika von den Merkmalen in der Syntax abweichen. Verschmelzungs- und Spaltungsregeln applizieren postsyntaktisch und können so die Merkmale manipulieren. Harris (1981) betrachtet die Präfixe als genuine Inversmarkierung, also als Vertauschung der grammatischen Funktionen der Argumente. Béjar und Řezáč (2009) folgend können die Präfixe jedoch keine reinen Inversmarker sein, da sie nicht die grammatischen Funktionen zweier Argumente wiedergeben, sondern vielmehr den Zyklus markieren, in dem die Sonde fündig geworden ist.

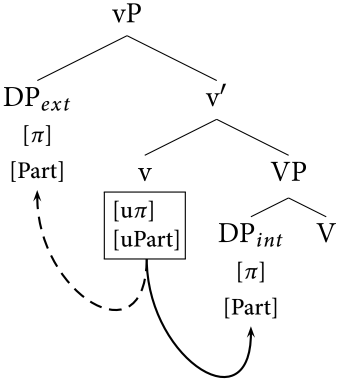
Argument noch Merkmale auf der Sonde verbleiben. Die Daten aus dem Itonama aus Abschnitt 2 zeigen aber deutlich, dass es Sprachen gibt, für die die Theorie von Béjar und Řezáč falsche Vorhersagen macht. Besonders deutlich wird dieser Unterschied, wenn man Sätze aus dem Baskischen und dem Itonama betrachtet, in denen beide Argumente mit lokalen Personenmerkmalen auftreten, wie in (1-a) und (1-b), hier wiederholt als (14-b) und (14-a).

(14) 2>1

a. *Itonama* (Crevels 2010: S. 684)
wase'wa de'-kewa-na-he-mo.
yesterday 2PL-face.see-NEUT-DISTR-1
'Yesterday you all saw us.'

b. *Baskisch* (Béjar und Řezáč 2009: S. 37)
ikusi n-ind-u-zu-n.
seen 1-X-have-2-PST
'You saw me.'

(15) 2>1



In (14-a) und (14-b) steht das externe Argument in der 2. Person und das interne Argument in der 1. Person. Da sowohl Baskisch als auch Itonama lokale Personen hierarchisch nicht untereinander ordnen, trägt auch die Sonde in beiden Sprachen die gleichen Merkmale, nämlich [uπ] und [uPart]. Die Theorie macht die Vorhersage, dass in diesem Fall mit dem internen Argument kongruiert wird, weil es beide Merkmale der Sonde überprüfen kann. In (14-b) ist dies auch der Fall, was in (15) mithilfe der durchgezogenen Linie dargestellt ist.

Im Baskischen wird der Abgleich im ersten Zyklus in diesen Kontexten durch den Marker *x* bzw. die Wahl eines speziellen Präfix markiert. Außerdem kongruiert Baskisch in diesem Fall noch mit dem externen Argument, was als Suffix realisiert wird. Im Itonama in (14-a) hingegen markiert das Präfix in diesen Kontexten Kongruenz mit dem externen Argument. Unter der Annahme, dass die Präfixmarkierung immer den ersten Abgleich kodiert, bedeutet dass, das im Itonama, anders als im Baskischen, zuerst Abgleich mit dem externen Argument eingegangen wird, was von der Theorie von Béjar und Řezáč (2009) jedoch nicht vorhergesagt wird.

Um (14-a) erklären zu können, muss also eine Möglichkeit gefunden werden, die Theorie von Béjar und Řezáč dahingehend zu erweitern, dass zuerst Abgleich mit dem externen Argument stattfinden kann, was in (15) durch die gestrichelte Linie dargestellt ist.

Béjar und Řezáč selbst schlagen für die sekundäre Kongruenz, die sowohl im Baskischen als auch in Itonama als Suffix realisiert wird, vor, dass eine zweite Sonde auf die vP projiziert wird. Von dieser Position aus kann sie dann ihre Merkmale mit dem externen Argument abgleichen, was später als Suffixmarkierung realisiert wird. Diese Erweiterung reicht aber nicht aus, um die Unterschiede zwischen den beiden Sprachen abzuleiten. Nach wie vor sagt die Theorie voraus, dass man immer zuerst mit dem internen Argument kongruiert, weil Merkmalsabgleich nur unter c-Kommando stattfinden kann.

In Abschnitt 4.4 werde ich eine Analyse vorstellen, die das obige Problem löst, indem die Sondierungsrichtung der Abgleichsoperation parametrisiert wird. In einigen Sprachen wird Merkmalsableich präferiert mit dem Spezifikator durchgeführt, in anderen Sprachen erfolgt Abgleich lieber unter c-Kommando. So kann erklärt werden, dass im Itonama, im Gegensatz zum Baskischen, Abgleich mit dem externen Argument stattfindet. Bevor ich jedoch den Ansatz vorstelle, möchte ich in den Abschnitten 4.1–4.3 zunächst drei andere potentielle Erweiterungen der Theorie von Béjar und Řezáč (2009) vorstellen und zeigen, dass diese die Daten nicht korrekt vorhersagen können.

4. Lösungsvorschläge

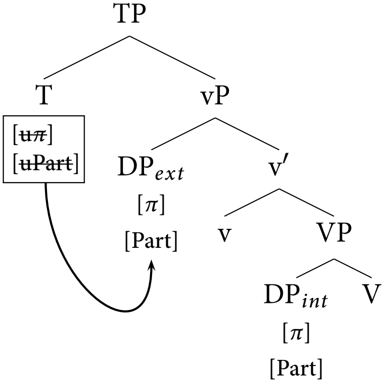
4.1. Sonde auf T

In der Theorie von Béjar und Řezáč (2009) sitzt die Sonde auf v, also zwischen den beiden Argumenten eines transitiven Satzes. Da die Sonde über dem internen, aber unter dem externen Argument steht und somit nur c-Kommando über das interne Argument hat, kann sie im ersten Suchlauf nur die DP des internen Arguments finden.

In Itonama kongruiert das Verb präferiert mit dem externen Argument, also scheint die Sonde zunächst Abgleich mit der DP des externen Arguments durchzuführen. Daher ist es naheliegend, dass die Sonde in Itonama nicht auf v, sondern auf einem höheren funktionalen Kopf, z.B. T sitzt.

Im Kontext (1,2)>(1,2), wie hier in (16) dargestellt, kongruiert in Itonama das Verb zuerst mit dem externen Argument.

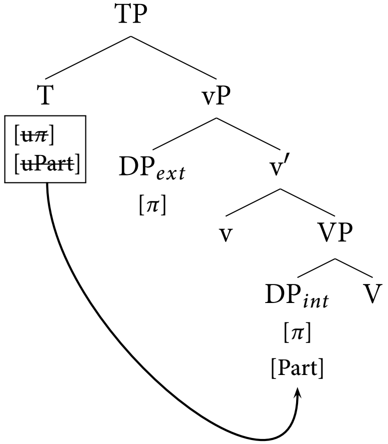
(16) (1,2)>(1,2)



Die Sonde auf T durchsucht ihren c-Kommando-Bereich und findet die nächste DP, nämlich die DP des externen Arguments. Diese kann beide Merkmale der Sonde überprüfen, so dass die Sonde ihre Suche beenden und mit dem externen Argument kongruieren kann.

In (17) ist der Kontext 3>(1,2) grafisch dargestellt.

(17) 3>(1,2)



Die Sonde sieht zunächst wieder die DP des externen Arguments. Da diese aber diesmal in der 3. Person steht, trägt sie das Merkmal [Part] nicht und kann nur das Merkmal [uπ] überprüfen. Die Sonde startet also einen zweiten Suchlauf, um eine DP mit dem Merkmal [Part] zu finden. Sie findet die tiefer stehende DP$_{int}$, welche das Merkmal [uPart] überprüfen kann.Die Sonde gleicht ihre Merkmale somit an zwei Argumenten ab.[8] Aus (4-c), hier wiederholt als (18), wird sichtbar, dass das Verb nun mit dem internen Argument kongruiert und einen Marker aus dem Possessivparadigma sowie den Inversmarker *k'i* wählt, um die unerwünschte Kongruenz mit dem internen Argument morphologisch zu markieren.

(18) 3>2

kumani	a'-k'i-pachïhï'-ke	kopone.
last.night	2SG-INV-bother-PL	rooster

'The rooster was bothering you last night.' (Crevels 2010: S. 685)

Im Kontext 3>3, wie in (5-c), hier wiederholt als (19), wird Kongruenz mit keinem der Argumente durch Affixe am Verb markiert. Außerdem tritt kein Inversmarker auf. Korrekt wäre es also, wenn man den Kontext 3>3 nicht als

---

[8] Ein ähnlicher Vorschlag, in dem ein Kopf seine ϕ-Merkmale mit zwei Argumenten in seinem c-Kommandobereich abgleicht, findet sich in Richards (2008) und Anagnostopoulou (2003, 2005).

Direktkontext aufführt, wie Crevels (2010) es tut, sondern ihn von den Direktkontexten separiert:

(19) 3>3
ni-fatima yo-lo'ba-he'-ka t'iyaya.
HON-Fatima CAUS-bathe-DISTR-FEM.SG boy
'Fatima bathed the boy.' (Crevels 2010: S. 685)

direkt	**invers**	**3. Kontext**
(1,2)>(1,2) (1,2)>3	3>(1,2)	3>3
Sonde im 1. Suchlauf saturiert	Sonde im 2. Suchlauf saturiert	Sonde nicht saturiert
Marker aus Tabelle 1	Marker aus Tabelle 2	gar kein Marker
kein Inversmarker	Inversmarker	kein Inversmarker

Tabelle 5: Morphologische Markierung der verschiedenen Kontexte

Der Inversmarker *k'i* tritt demnach genau dann auf, wenn die Sonde im zweiten Zyklus ihr Merkmal [uPart] abgleichen kann. Damit verhält sich Itonama anders als Baskisch, wo der Inversmarker dann erscheint, wenn die Sonde im ersten Versuch ihre Merkmale abgleichen kann.

In Sätzen, in denen sowohl das interne als auch das externe Argument in einer lokalen Person stehen, wird im Itonama zusätzlich zur Präfixmarkierung, die Kongruenz mit DP$_{ext}$ realisiert, die Kongruenz mit dem internen Argument als Suffix markiert, wie im folgenden Beispiel (siehe auch (5-a) und (5-b)):

(20) a. 2>1
wase'wa de'-kewa-na-he-mo.
yesterday 2PL-face.see-NEUT-DISTR-1
'Yesterday you all saw us.'

b. 1>2
si-kamo'-ke-we.
1SG-face.hit-PL-2
'I am going to hit you (in the face).' (Crevels 2010: S. 684)

Die Suffixe *mo* für 1. Person und *we* für 2. Person sind die einzigen Suffixe, die in dieser Position auftreten können, da 3. Person nicht markiert wird und Numerus nicht unterschieden wird. Die Analyse dieser zweiten Kongruenzposition lässt sich nicht ohne Weiteres erklären. Würde man, der Analyse von Béjar und Řezáč (2009) folgend, eine zweite Sonde nach oben projizieren, würde diese unter c-Kommando dennoch das externe Argument finden und mit diesem Abgleich durchführen müssen. Die Daten in (20-a) und (20-b) zeigen aber, dass jeweils das Personenmerkmal des internen Arguments als Suffix realisiert wird.[9]

4.2. Abgleich der Sonde nach Spec-v

Um die Daten aus Itonama erklären zu können, ist es nicht zwingend notwendig anzunehmen, dass die Sonde auf T anstatt auf v sitzt, wie von Béjar und Řezáč (2009) vorgeschlagen. Die Daten aus Itonama zeigen klar, dass eine Möglichkeit gefunden werden muss, Abgleich zuerst mit der DP des externen Arguments durchzuführen. Dazu könnte die Annahme in (21) dienen (siehe Assmann et al. 2013):

(21) *Präferierter Abgleich mit dem Spezifikator*
Abgleich mit dem Spezifikator wird Merkmalsabgleich unter c-Kommando vorgezogen.

Die Sonde präferiert somit Abgleich mit ihrem Spezifikator. Hierfür wird angenommen, dass Merkmale unter m-Kommando statt unter c-Kommando abgeglichen werden, siehe (22).

---

[9] Dieses Problem könnte potentiell gelöst werden, indem man annimmt, dass Argumente nach Abgleich mit einem funktionalen Kopf deaktiviert werden und so nicht mehr sichtbar sind für weitere Abgleichsrelationen (siehe z.B. Chomsky 2000, 2001). In diesem Fall würde die Kongruenz der ersten Sonde mit dem externen Argument bewirken, dass dieses deaktiviert und somit für eine weitere Sonde nicht mehr zugänglich wäre. Die zweite Sonde könnte dann nur noch mit dem internen Argument kongruieren. Laut Chomsky (2000, 2001) sind deaktivierte Argumente aber trotzdem Intervenienten für Merkmalsabgleich und können so den Abgleich mit dem internen Argument unterbinden (siehe (i)).

(i) *Defektive Intervention* (Chomsky 2000, 2001)
In einer Struktur $\alpha > \beta > \gamma$, in der > für c-Kommando steht und eine Sonde α ihre Merkmale sowohl mit β als auch mit γ abgleichen könnte, unterbindet ein deaktiviertes β den Merkmalsabgleich von α mit γ.

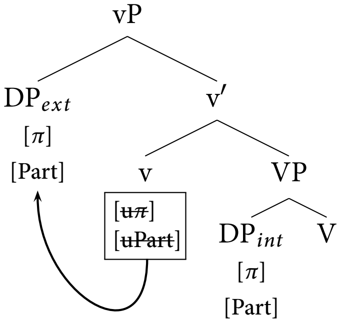
Somit reicht auch diese zusätzliche Annahme nicht aus, um die Daten im Itonama abzuleiten.

(22) *Merkmalsabgleich unter m-Kommando*
Ein Knoten A kann nicht überprüfte Merkmale [uM] genau dann mit dem Knoten B abgleichen, wenn (i), (ii) und (iii) gelten:

(i) B trägt ein Merkmal [M]
(ii) weder dominiert A B noch dominiert B A
(iii) B wird von der maximalen Projektion von A dominiert

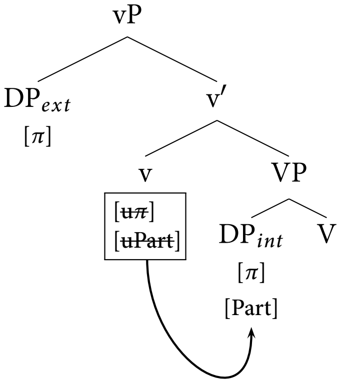
Erst wenn der Spezifikator nicht alle Merkmale überprüfen kann, wird Abgleich mit einem Element im c-Kommandobereich durchgeführt. In (23) ist der Kontext (1,2)>(1,2) dargestellt.

(23) (1,2)>(1,2)



Die Sonde startet ihren ersten Suchlauf und findet in ihrem Spezifikator DP$_{ext}$. Da diese eine lokale Person ist, kann sie beide Merkmale auf der Sonde abgleichen. Das Verb kongruiert aufgrund dessen mit dem externen Argument. Anders verhält es sich jedoch im Kontext 3>(1,2) in (24).

(24) 3>(1,2)



Hier kann im ersten Suchlauf der Sonde nur das Merkmal [uπ] überprüft werden. Dann startet sie einen zweiten Suchlauf, dieses Mal in ihrem c-Kommandobereich, um das Merkmal [uPart] abzugleichen. Sie findet das interne Argument, mit dem das Verb dann kongruiert. Da die Sonde aber erst im zweiten Anlauf ihre Merkmale abgleichen kann, muss wiederum die Morphologie des zweiten Zyklus, also ein Marker aus dem Possessivparadigma, sowie der Inversmarker *k'i* gewählt werden.

Zwar kann diese Analyse die Daten aus Itonama erklären, allerdings macht sie die falschen Vorhersagen für das Baskische: Da das Prinzip in (21) universell ist und somit auch im Baskischen gilt, wird vorausgesagt, dass auch im Baskischen zuerst mit dem externen Argument kongruiert wird. Damit kann diese Analysemöglichkeit nicht erklären, warum im Baskischen vorzugsweise mit dem internen Argument kongruiert wird.

4.3. Sonden auf T und v

Da weder die Positionierung der Sonde auf T noch der präferierte Abgleich mit dem Spezifikator eine empirisch adäquate Analyse für den Unterschied zwischen Baskisch und Itonama bietet, möchte ich noch eine dritte Möglichkeit vorstellen.

Die Daten aus Baskisch und Itonama zeigen deutlich, dass Sprachen unterschiedliche Präferenzen haben können, ob sie Kongruenz mit dem externen oder internen Argument bevorzugen. Außerdem weist in beiden Sprachen das Verb im Kontext (1,2)>(1,2) noch eine zusätzliche Kongruenzposition auf. Im Baskischen wird Kongruenz mit dem internen Argument präferiert und dann zusätzlich in lokalen Kontexten noch mit dem externen Argument kongruiert. In Itonama wird grundsätzlich mit dem externen Argument und in lokalen Kontexten darüber hinaus noch mit dem internen Argument kongruiert. Die Tatsache, dass die beiden Argumente an verschiedenen Positionen am Verb auftreten, legt nahe, dass in beiden Sprachen jeweils zwei verschiedene Sonden für die Kongruenz zuständig sind. Zudem weist Itonama je nach Kongruenzposition unterschiedliche Morphologie auf.

Eine Möglichkeit, dieses Problem zu lösen, wäre, dass jede der Sprachen grundsätzlich zwei Sonden hat, die auf v und T sitzen. Abgleich erfolgt unter c-Kommando. In Itonama und Baskisch tragen beide Sonden die Merkmale [uπ] und [uPart], da 1. und 2. Personen nicht unterschieden werden und 3.

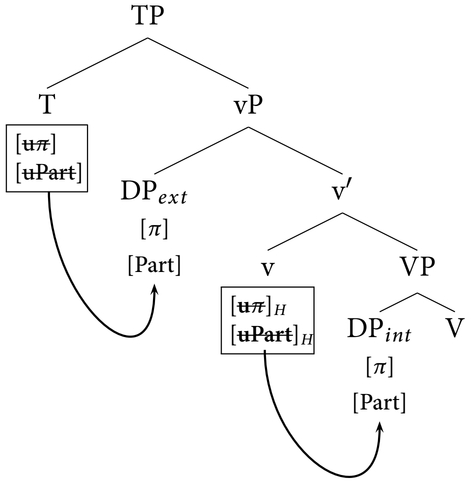
Personen niemals overte Kongruenz auslöst.[10] Eine der Sonden ist die Hauptsonde und für die primäre Kongruenz verantwortlich, die in Baskisch und Itonama als Präfix realisiert wird. Diese Sonde darf als erste ihre Merkmale abgleich. Die zweite Sonde kann ihre Merkmale nur abgleichen, wenn die Merkmale der ersten Sonde bereits überprüft wurden und es noch mindestens eine DP gibt, die von der Hauptsonde noch nicht durchsucht wurde. Das Prinzip der Hauptsonde habe ich in (25) zusammengefasst.

(25) *Hauptsondenprinzip*
Die Hauptsonde trägt nicht überprüfte Merkmale $[\mathbf{uM}]_H$, die zuerst abgeglichen werden. Alle anderen Sonden tragen Merkmale [uM], die erst nach Überprüfung aller Hauptsondenmerkmale abgeglichen werden dürfen.

Sprachen unterscheiden, welche der beiden Sonden die Hauptsonde ist. Sitzt die Hauptsonde auf v, wie im Baskischen, kongruiert die Sprache präferiert mit dem internen Argument, wie in (1-b), hier wiederholt als (26-a).

(26) a. 2>1, *Baskisch*
ikusi n-ind-u-zu-n.
seen **1**-X-have-2-PST
'You saw me.' (Béjar und Řezáč 2009: S. 37)

b.



---

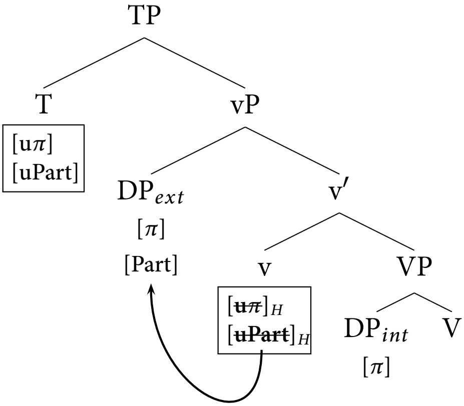
[10] In anderen Sprachen können die Sonden aber durchaus unterschiedliche Merkmale tragen. In Mohawk bspw. trägt eine Sonde [uπ], [uPart] und [uSprecher], die andere aber nur [uπ] (Béjar und Řezáč 2009: S. 59).

In (26-b) sitzt die Hauptsonde auf v. Das interne Argument kann alle Merkmale der Sonde überprüfen, so dass das interne Argument als Präfix erscheint. Das externe Argument steht somit noch der zweiten Sonde auf T zur Verfügung und die zusätzliche Kongruenz wird als Suffix realisiert.

Ein Beispiel für einen Kontext, in dem nicht beide Argumente lokale Personenmerkmale tragen, ist in (27-a) aufgeführt.

(27) a. 1>3, *Baskisch*
ikusi n-u-en.
seen 1-have-PAST
'I saw him.' (Béjar und Řezáč 2009: S. 37)

b.

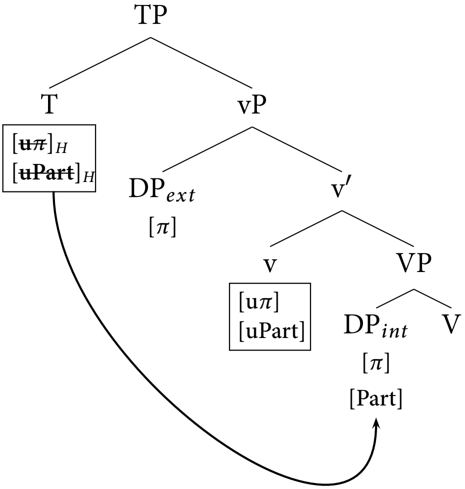


In (27-b) muss die Hauptsonde einen zweiten Suchlauf starten, da das interne Argument nicht alle Merkmale auf der Sonde überprüfen kann. Daher benötigt sie das externe Argument. Es ist somit kein Argument mehr für die zweite Sonde übrig, weswegen nur das externe Argument am Verb markiert wird.

Anders als im Baskischen sitzt die Hauptsonde im Itonama auf T. Wenn das externe Argument die Merkmale der Hauptsonde überprüfen kann, ist die interne DP noch für die sekundäre Kongruenz übrig, die als Suffix realisiert wird. In (4-b), hier wiederholt als (28-a), steht aber das externe Argument in der dritten Person und kann das Merkmal [uPart] der Hauptsonde nicht überprüfen. Diese benötigt also die DP des internen Arguments, womit keine DP mehr für die zweite Sonde übrig ist. Es findet also ausschließlich Kongruenz mit dem internen Argument statt.

(28) a. 3>1, *Itonama*
sih-k'i-ma-doh-ne upa'u.
1PL.EXCL-INV-hand-bite-NEUT dog
'The dog bit us on the hand.' (Crevels 2010: S. 685)

b.



Grundsätzlich wird aber in Fällen wie (28) das Prinzip des Strikten Zyklus verletzt:

(29) *Prinzip des Strikten Zyklus* (basierend auf Chomsky 1973)
a. Keine Operation kann innerhalb einer Domäne applizieren, die von einem zyklischen Knoten α dominiert wird, wenn diese Operation ausschließlich eine echte Subdomäne von α betrifft, die von einem anderen zyklischen Knoten β dominiert wird.
b. Jede Phrase ist ein zyklischer Knoten.

Die Sonde auf v, die zuerst verkettet wird, dürfte in diesem Fall erst ihre Merkmale abgleichen, nachdem die Sonde auf T verkettet wurde und ihre Merkmale abgeglichen hat. Da dies eine Verletzung des Strikten Zyklus in (29) verursacht, wird diese Theorie an dieser Stelle verworfen.

4.4. Zwei Sonden auf v

Wie ich bereits in Abschnitt 4.2 vorgestellt habe, können die Daten aus Itonama erklärt werden, wenn Merkmalsabgleich unter m-Kommando angenommen wird (siehe (22)). Die Theorie in Abschnitt 4.2 konnte allerdings die Daten aus dem Baskischen nicht ableiten. In diesem Abschnitt werde ich zeigen, dass es möglich ist, die Daten aus Itonama und aus dem Baskischen zu erklären, indem man Merkmalsabgleich unter m-Kommando annimmt und den Abgleich dahingehend parametrisiert, dass in einigen Sprachen präferiert mit dem Spezifikator kongruiert wird, andere Sprachen aber Abgleich unter c-Kommando bevorzugen. Außerdem kann das Hauptsondenprinzip in (25) genutzt werden, um die Sprachen voneinander zu unterscheiden.

Ich nehme im Folgenden an, dass es wiederum zwei Sonden gibt, die jeweils die Merkmale [uπ] und [uPart] tragen und diese unter m-Kommando abgleichen möchten (siehe (31)). Beide Sonden sitzen aber auf demselben Kopf, nämlich v, was grundsätzlich zu einem Problem für das Prinzip der Frühzeitigkeit (*Earliness Principle*, Pesetsky 1989, Řezáč 2004) in (30) führt.

(30) *Prinzip der Frühzeitigkeit* (Pesetsky 1989, Řezáč 2004):
Eine Sonde [uF] löst so schnell wie möglich Abgleich aus.

Da beide Sonden so schnell wie möglich Abgleich auslösen wollen, aber nicht zwei Abgleichsoperation gleichzeitig applizieren können, muss es einen Faktor geben, der regelt, welche der beiden Sonden zuerst ihre Merkmale abgleichen darf. Die Ordnung der Abgleichsoperationen wird durch das Hauptsondenprinzip in (25) bestimmt.

Merkmalsabgleich ist wie in (31) definiert. Die Idee ist, dass die präferierte Sondierungsrichtung der Abgleichsoperation nicht universell festgelegt ist, sondern durch einen sprachspezifischen Parameter gesteuert wird.

(31) *Merkmalsabgleich unter m-Kommando*
Ein Knoten A kann nicht überprüfte Merkmale [uM] genau dann mit einem Knoten B abgleichen, wenn (i), (ii), (iii), (iv) und (v) gelten:
(i) B trägt ein Merkmal [M].
(ii) B ist noch aktiv.
(iii) weder dominiert A B noch dominiert B A.
(iv) B wird von der maximalen Projektion von A dominiert.
(v) es gibt keinen Knoten C, der noch aktiv ist und in dem präferierten Suchraum von A liegt.

(32) *Aktivität*
Ein Ziel B ist noch aktiv, wenn es noch nicht von einer Sonde durchsucht worden ist.

(33) *Präferierter Suchraum*
Der präferierte Suchraum eines Kopfes K in einer Sprache L ist (i) oder (ii):
  a. der Spezifikator von K
  b. das Komplement von K.

Im folgenden werde ich zeigen, dass die Annahmen in (25) und (31) – (33) die richtigen Voraussagen machen.

Abstrakt laufen die Derivationen in Baskisch und Itonama gleich ab. Die Sonde auf v, die durch (25) als Hauptsonde ausgezeichnet ist, löst zuerst Abgleich aus. Da zuvor keine Abgleichsoperation appliziert hat, sind sowohl das externe Argument im Spezifikator von v als auch das interne Argument, das im Komplement von v liegt, noch aktiv. Allerdings muss die Hauptsonde Abgleich mit dem Argument α eingehen, das im präferierten Suchlauf liegt, da α gemäß (31) Merkmalsabgleich mit dem anderen Argument blockiert. Durch diese erste Abgleichsoperation wird α somit deaktiviert. Anschließend appliziert die zweite Abgleichsoperation, ausgelöst von der anderen Sonde auf v. An diesem Punkt gibt es aber nur noch ein aktives Argument in der Struktur. Entgegen der sprachspezifisch präferierten Sondierungsrichtung muss die Sonde mit diesem Argument Abgleich eingehen.[11]

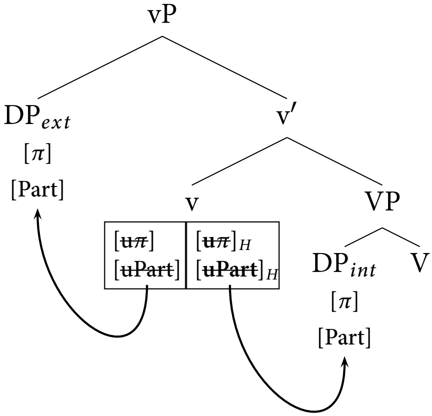
Der Vorteil dieser Analyse gegenüber der Analyse in Abschnitt 4.3 ist, dass die hier getroffenen Annahmen nicht mit dem Prinzip des Strikten Zyklus in (29) konfligieren, da alle Operationen innerhalb der gleichen Phrase applizieren.

In (34) ist der Kontext (1,2)>(1,2) für das Baskische dargestellt.

---

[11] Hierbei ist zu bemerken, dass defektive Intervention in dieser Analyse keine Rolle spielt. Da die Sonden auf v und somit zwischen den beiden Argumenten sitzen, können deaktivierte Argumente den Merkmalsabgleich nicht direkt blockieren. Damit ergibt sich das Problem der Theorie aus Abschnitt 4.1, das in Fußnote 9 besprochen wurde, hier nicht.

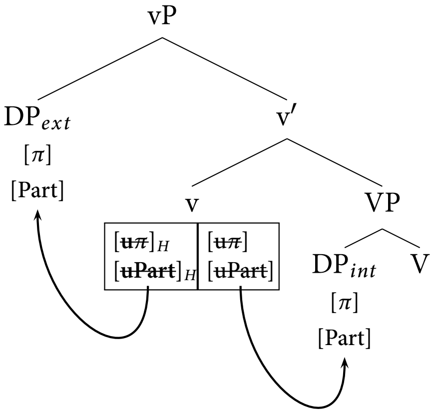
(34) (1,2)>(1,2), *Baskisch*



Im Baskischen ist der präferierte Suchraum der Sonden das Komplement von v. Die Hauptsonde (siehe (25)) darf ihre Merkmale zuerst abgleichen. Da keine Abgleichsoperation stattgefunden hat, sind beide Argumente noch aktiv. Die Hauptsonde kann daher ihre Merkmale mit dem internen Argument im c-Kommandobereich von v abgleichen. Das interne Argument wurde somit durchsucht und wird deaktiviert. Die zweite Sonde kann ihre Merkmale nun nur noch entgegen ihrer präferierten Sondierungsrichtung mit dem externen Argument im Spezifikator von v abgleichen, da das interne Argument nicht mehr zugänglich ist.

In Itonama hingegen ist der präferierte Suchraum der Sonden der Spezifikator von v. Das ist in (35) dargestellt.

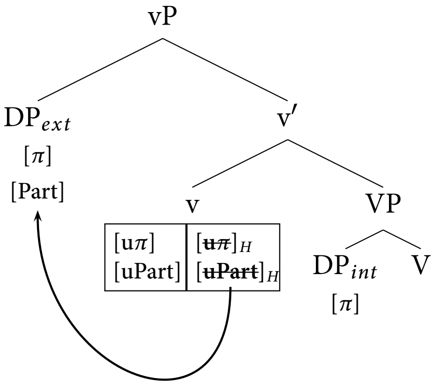
(35) (1,2)>(1,2), *Itonama*



Die Hauptsonde kann in ihrem präferierten Suchraum, also dem Spezifikator von v, ihre Merkmale mit dem externen Argument abgleichen. Dieses Argument wird daher deaktiviert und ist für weitere Abgleichsoperationen nicht mehr zugänglich. Die zweite Sonde kann nur noch entgegen ihrer präferierten Sondierungrichtung Abgleich mit dem internen Argument im Komplement durchführen, da dieses als einziges noch aktiv ist.

In (36) ist nun ein Kontext für das Baskische dargestellt, in dem nicht beide Argumente lokale Personenmerkmale tragen:

(36) (1,2)>3, *Baskisch*



Da der präferierte Suchbereich im Baskischen das Komplement von v ist, versucht die Hauptsonde zunächst ihre Merkmale mit dem internen Argument abzugleichen. Da das interne Argument zwar das Merkmal [uπ], aber nicht das Merkmale [uPart] abgleichen kann, muss die Hauptsonde es mit dem externen Argument abgleichen. Da nun aber sowohl das interne als auch das externe Argument Abgleichsoperationen mit der Hauptsonde eingegangen sind, werden beide Argumente deaktiviert und sind daher nicht mehr für weiteren Merkmalsabgleich zugängig. Die zweite Sonde auf v kann daher keine Abgleichsoperation durchführen und wird per Annahme in diesen Fällen gelöscht.

5. Regelinteraktion

Die Theorie von Béjar und Řezáč (2009) sagt für die Kontexte (1,2)>(1,2) im Baskischen eine *Bleeding*-Relation bezüglich der Abgleichsoperationen voraus. Da die Sonde auf v bereits Abgleich mit dem internen Argument durchgeführt hat, kann kein Merkmalsabgleich mit dem externen Argument mehr stattfinden. Merkmalsabgleich mit dem internen Argument *bleedet* also Merkmalsabgleich mit dem externen Argument.

Wie in Abschnitt 2 gezeigt, gleicht im Itonama in diesen Kontexten aber das externe Argument seine Merkmale mit v ab. Gemäß der Theorie von Béjar und Řezáč (2009) müsste auch im Itonama Abgleich mit dem externen Argument durch Abgleich mit dem internen Argument blockiert werden. Da dies nicht der Fall ist, stellt Itonama in der Theorie von Béjar und Řezáč (2009) einen Fall von *Counter-Bleeding* dar.

In 4.4 wurde diese opake Regelinteraktion aufgelöst, indem angenommen wurde, (i) dass es zwei Sonden auf v gibt und (ii) dass die Ordnung der beiden Abgleichsoperation sprachspezifisch festgelegt ist. Während im Baskischen zuerst mit dem internen Argument kongruiert wird, gleicht im Itonama das externe Argument zuerst seine Merkmale mit v ab. Das hat zur Folge, dass im Itonama Abgleich mit dem internen Argument *zu spät* appliziert, um Abgleich mit dem externen Argument zu bleeden.

Die Ordnung der beiden Abgleichsoperationen ergibt sich dadurch, dass die präferierte Sondierungsrichtung sprachspezifisch festgelegt ist. Im Baskischen werden Merkmale präferiert unter c-Kommando abgeglichen, was zur Folge hat, dass zuerst mit dem internen Argument kongruiert wird, das im c-Kommandobereich von v liegt. Die zweite Abgleichsbeziehung muss mit dem externen Argument erfolgen, da das interne Argument bereits Abgleich unterlaufen ist. Im Itonama ist der Fall umgekehrt: Hier erfolgt Abgleich präferiert mit dem Spezifikator, weswegen Merkmalsabgleich zuerst mit dem externen Argument eingegangen wird. Anschließend ist nur noch das interne Argument für Abgleich zugänglich, was dazu führt, dass v seine Merkmale mit dem internen Argument erst spät abgleichen kann.

Die sprachspezifisch präferierte Sondierungsrichtung entscheidet also, welche Abgleichsoperation zuerst stattfindet. Kongruenz mit dem Argument, das im präferierten Suchraum liegt, bleedet Kongruenz mit dem Argument, das im nicht-präferierten Suchraum liegt.

6. Zusammenfassung

In diesem Papier habe ich verbale Kongruenz in Sprachen mit Inverssystem anhand von Daten aus den Sprachen Itonama und Baskisch analysiert. In Abschnitt 3 habe ich die Analyse von Béjar und Řezáč (2009) rekapituliert und gezeigt, dass ihre Analyse ausschließlich Sprachen erklären kann, die vorzugsweise Kongruenz mit dem internen Argument durchführen. Die Daten aus Itonama liefern aber ein Beispiel für eine Sprache, die bevorzugt mit dem externen Argument kongruiert. Um diese Diskrepanz erklären zu können, habe ich verschiedene Lösungsmöglichkeiten diskutiert.

Die erste Möglichkeit bestand darin, dass die Sonde auf T statt auf v sitzt. Bei dieser Möglichkeit kann die Annahme, dass Merkmalsabgleich unter c-Kommando durchgeführt wird, beibehalten werden. Diese Analyse kann die primäre Kongruenz erklären, scheitert aber an der Erklärung der in bestimmten Kontexten auftretenden Suffixe.

Auch die Annahme, dass zuerst Abgleich mit dem Spezifikator durchgeführt wird, kann die Daten aus Itonama scheinbar problemlos ableiten. Hier muss zwar Abgleich unter m-Kommando angenommen werden, dafür kann die Sonde auf v belassen werden. Allerdings scheitert diese Analyse an den Daten aus dem Baskischen.

Da in manchen Kontexten beide Argumente an unterschiedlichen Positionen kongruieren, wurde weiterhin der Vorschlag diskutiert, dass jede Sprache zugrundeliegend zwei Sonden hat, die auf v und T sitzen. Ich habe das Prinzip der Hauptsonde eingeführt und gezeigt, dass dieses Prinzip erklären kann, warum nur in bestimmten Kontexten mit beiden Argumenten kongruiert werden kann. Es gibt einen Parameter, der festlegt, ob die Hauptsonde auf v oder T sitzt und somit die Sprachen Baskisch und Itonama voneinander unterscheidet. Diese Idee verletzt jedoch das Prinzip des Strikten Zyklus und musste aus diesem Grund verworfen werden.

Schließlich konnte ich zeigen, dass sich sowohl die Daten aus dem Baskischen sowie aus Itonama erklären lassen, wenn man annimmt, dass Merkmalsabgleich unter m-Kommando stattfindet und das Hauptsondenprinzip mit dem Spezifikator/Komplement-Parameter kombiniert. In diesem Fall gibt es jeweils zwei Sonden, die beide auf v sitzen und von denen eine ihre Merkmale mit dem Spezifikator und die andere unter c-Kommando abgleicht. Das Hauptsondenprinzip bestimmt, welche der Abgleichsoperationen zuerst applizieren darf und kann somit erklären, warum sich Itonama und Baskisch im Kontext (1,2)>(1,2) unterschiedlich verhalten.

Literatur

Anagnostopoulou, Elena (2003): *The Syntax of Ditransitives. Evidence from Clitics*. De Gruyter Mouton, Berlin/New York.

Anagnostopoulou, Elena (2005): Strong and Weak Person Restrictions: a Feature-Checking analysis. *In:* L. Heggie und F. Ordoñez, eds, *Clitics and Affixation*. Benjamins, Amsterdam, pp. 199–235.

Anderson, Stephen R. (1992): *A-Morphous Morphology*. Cambridge University Press, Cambridge.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller und Philipp Weisser (2013): Ergatives Move too Early: On an Instance of Opacity in Syntax. *In:* F. Heck und A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 363-412.

Béjar, Susana (2003): Phi-Syntax: A Theory of Agreement. PhD thesis, University of Toronto.

Béjar, Susana und Milan Řezáč (2009): 'Cyclic Agree', *Linguistic Inquiry* 40(1), 35–73.

Brody, Michael (2002): On the Status of Representations and Derivations. *In:* S. D. Epstein und T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 19–41.

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

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels und J. Uriagereka, eds, *Step by Step. Papers in Minimalist Syntax in Honor of Howard Lasnik*. MIT Press, Cambridge, Mass., pp. 89–156.

Chomsky, Noam (2001): Derivation by Phase. *In:* M. Kenstowicz, ed., *Ken Hale: A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Crevels, Mily (2010): Ditransitives in Itonama. *In:* A. Malchukov, M. Haspelmath und B. Comrie, eds, *Studies in Ditransitive Constructions: A comparative Handbook*. De Gruyter Mouton, Berlin, pp. 678–709.

Crevels, Mily (2011): 'Who did What to Whom in Magdalena (Itonama)', *International Journal of American Linguistics* 77(4), 577–594.

Epstein, Samuel David und T. Daniel Seely (2002): Rule Applications as Cycles in a Level-Free Syntax. *In:* S. D. Epstein und T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65–89.

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

Harris, Alice (1981): *Georgian Syntax: A Study in Relational Grammar*. Cambridge University Press, Cambridge.

Pesetsky, David (1989): *Language-Particular Processes and the Earliness Principle*. MIT Press, Cambridge, Mass.

Řezáč, Milan (2004): Elements of Cyclic Syntax: Agree and Merge. PhD thesis, University of Toronto.

Richards, Marc (2008): Quirky Expletives. *In:* R. d'Alessandro, G. H. Hrafnbjargarson und S. Fischer, eds, *Agreement Restrictions*. De Gruyter Mouton, Berlin, pp. 181–213.

Silverstein, Michael (1976): Hierarchies of Features and Ergativity. *In:* R. M. W. Dixon, ed., *Grammatical categories in Australian Languages*. Australian Institute of Aboriginal Studies, Canberra, pp. 112–171.

Stump, Gregory T. (2001): *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge University Press, Cambridge.

Trommer, Jochen (2001): Distributed Optimality. PhD thesis, University of Potsdam.

Trommer, Jochen (2003*a*): Feature (Non-)Insertion in a Minimalist Approach to Spellout. In: *Proceedings of CLS 39*. pp. 469–480.

Trommer, Jochen (2003*b*): The Interaction of Morphology and Syntax in Affix Order. In: *Yearbook of Morphology 2002*. Dordrecht: Kluwer, pp. 283–324.

Ergatives Move Too Early:
On an Instance of Opacity in Syntax

Anke Assmann, Doreen Georgi, Fabian Heck,
Gereon Müller & Philipp Weisser[*]

Abstract

We examine the ban on Ā-movement of the external argument of a transitive verb that holds in many morphologically ergative languages. We argue that the prohibition against movement of the ergative subject should not be derived from restrictions on the movement of the ergative DP. Rather, we suggest that movement of the ergative argument is per se unproblematic, but if it applies, it applies too early, and thereby creates problems for its absolutive co-argument, which does not receive structural case. In morphologically accusative languages, no such movement asymmetry arises because arguments move too late to trigger the fatal consequences that moving ergatives cause. We present a co-argument-based analysis that implies a strictly derivational syntax in which the order of operations plays an important role in deriving properties of the grammar.

1. Introduction

In many morphologically ergative languages, ergative arguments cannot undergo Ā-movement (wh-movement, focussing, relativization). This phenomenon is an instance of the more general observation that languages exhibit extraction asymmetries, viz. that some kinds of linguistic expressions are less mobile than others. In the present paper, we suggest that movement asymmetries can arise because movement of an item α may create problems for another, sufficiently similar item β. We present a co-argument-based approach

---

[*] We are grateful for comments from and helpful discussion with the participants of the colloquium *Theory of Grammar* at the University of Leipzig as well as the audiences of *Generative Grammatik des Südens* in Freiburg/Breisgau, GLOW 35 in Potsdam, and the workshop on *Decomposition and natural classes in argument coding* of the Forschergruppe 742 in Leipzig. Particular thanks go to Michael Yoshitaka Erlewine, Peter Gallmann, and Erich Groat.

Rule Interaction in Grammar, 363–412
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

to displacement (α cannot move in the presence of β because α-movement creates problems for β-licensing) of the type that has sometimes been suggested for case assignment (α is assigned x-case in the presence of β; see Marantz 1991, Bittner and Hale 1996, Wunderlich 1997, Stiebels 2000, McFadden 2004).

As a case study on movement asymmetries, we focus on the ban on ergative movement in morphologically ergative languages. We argue that the prohibition against movement of the ergative subject should not be deduced from restrictions on the movement of the ergative. Rather we claim that movement of the ergative DP is per se unproblematic, but if it applies, it applies too early, and thereby creates problems for the absolutive co-argument of the ergative subject (cf. Polinsky et al.'s (2012) hypothesis that ergative displacement leads to a processing problem because removal of an ergative DP from a clause makes identification of the grammatical function of the absolutive DP difficult, but not vice versa). Here, we will argue that movement of the ergative prevents case assignment to the absolutive DP (cf. Aldridge 2004, Coon et al. 2011, where it is proposed that case movement of the absolutive creates an island for ergative movement). No extraction asymmetry arises in morphologically accusative languages because accusative or nominative arguments move too late to trigger the fatal consequences that early moving ergatives cause. In the formal account we develop, the different timing of movement in the two types of languages is a direct consequence of the background theory that derives morphological ergativity and accusativity in the first place. This theory as well as the relational, co-argument-based analysis of the ban on movement of ergatives proposed here implies a strictly derivational syntax in which the order of operations plays an important role in deriving properties of the grammar.

We will proceed as follows: In section 2 we introduce data from morphologically ergative languages that also show syntactic ergativity with respect to Ā-movement, that is, the ban on movement of ergative arguments of transitive predicates. Furthermore, we discuss problems of previous analyses of the phenomenon. Section 3 contains the assumptions and shows how morphological ergativity/accusativity is derived. Next, we illustrate in section 4 how the movement asymmetry in morphologically ergative languages arises in this system and why no such asymmetry results in morphologically accusative languages. In section 5 we address further predictions of the analysis. Finally, in section 6 we develop an approach to a repair strategy of the ban on ergative movement, the agent focus construction, within the system. Section 7 concludes.

2. Syntactic Ergativity in Ā-Movement

2.1. Data

In morphologically ergative languages (Comrie 1988, Dixon 1994), the internal argument of a transitive verb (DP_{int}) and the sole argument of an intransitive verb are encoded by the same morphological markers: They either bear the same case marker, called absolutive case, or they trigger the same agreement markers on the verb.[1] The external argument of a transitive verb (DP_{ext}) is encoded differently from the two other arguments: It bears ergative case or is cross-referenced by a different set of agreement markers on the verb. Many morphologically ergative languages also exhibit syntactic ergativity with respect to Ā-movement: DP_{ext} of a transitive verb cannot be questioned, relativized or focussed. DP_{int} of a transitive verb and the sole argument of an intransitive verb, however, can be freely extracted. Thus, the absolutive DPs cluster together, the ergative DP behaves differently. In this subsection, we present data from various morphologically ergative languages and different types of Ā-movement that illustrate this ban on ergative movement.[2]

---

[1] In what follows, we assume that the absolutive case involved is syntactic. We are aware of the possibility that morphological quirks can make a syntactic non-absolutive case look like absolutive on the surface (see Legate 2008).

[2] Not all morphologically ergative languages exhibit the ban on Ā-moving the ergative. Explanations for this variation are proposed in section 5.2. There is also variation as to which type of Ā-movement is subject to the constraint (see Stiebels 2006 on Mayan); we leave this issue unaddressed. Many Austronesian languages show constraints on Ā-movement similar to the one under discussion. However, for some of them it is unclear whether they are ergative (see, e.g., Chung 1998: 27-32, 99-111 on Chamorro). We therefore confine the discussion to languages whose status as being ergative is undisputed. The restrictions in Austronesian also differ from the one discussed here in that adjunct extraction in these languages is also very constrained if not impossible (cf. Keenan 1976 on Malagasy). Despite the limitation with respect to Austronesian and the variation among ergative languages in general, we take it that the ban on Ā-movement of the ergative argument in morphologically ergative languages instantiates a pattern and is not accidental. If it were, one would expect a similar ban to occur in accusative type languages, which is not the case as far as we know.

2.1.1. *Wh-Movement*

In Mayan languages, argument DPs do not bear overt case markers, but ergative and absolutive DPs trigger different kinds of agreement: DP$_{ext}$ triggers ergative agreement whereas DP$_{int}$ and the sole argument of an intransitive verb trigger absolutive agreement. Most Mayan languages are verb initial in affirmative sentences. If a DP is questioned, it is moved to the preverbal position. As the data with transitive verbs in (1) from Kaqchikel and in (3) from K'ichee' show, DP$_{int}$ can be questioned (see the b.-examples), but wh-movement of DP$_{ext}$ leads to ungrammaticality (see the c.-examples). The sentence without wh-movement is given in the a.-examples. The sole argument of an intransitive verb can also be questioned, as shown in (2) for Kaqchikel and in (4) for K'ichee'. It is thus possible to Ā-move DP$_{abs}$ but impossible to extract DP$_{erg}$.[3]

(1) *Wh-movement of DP$_{erg}$ vs. DP$_{abs}$ in Kaqchikel (Mayan)*:

a. N-Ø-u-löq' jun sik'iwuj ri a Karlos.
INCOMPL-3SG.ABS-3SG.ERG-buy INDEF book DET CL Carlos
'Carlos buys a book.'

b. Atux n-Ø-u-löq' a Karlos?
Q INCOMPL-3SG.ABS-3SG.ERG-buy CL Carlos
'What does Carlos buy?'

c. *Achike n-Ø-u-löq' jun sik'iwuj?
Q INCOMPL-3SG.ABS-3SG.ERG-buy INDEF book
'Who buys a book?'

---

[3] Unless references are provided, the Kaqchikel and K'ichee' examples in this paper are due to our informants Telma Can Pixabaj (K'ichee') and Rony Arnoldo Otzoy Chipix, Erika Edith Mux Son, and Herminia Son Bal (Kaqchikel). We used the following abbreviations in the glosses: 1/2/3 = $1^{st}/2^{nd}/3^{rd}$ person, ABS= absolutive, AF = agent focus, ANIM = animate, AP = antipassive, ART = article, CL = clitic, CLASS = class marker, COMPL = completive aspect, DAT = dative, DEIC = deictic element, DEP = dependent aspect, DET = definite determiner, DIR = directional, DUR = durative aspect, ENC = enclitic, ERG = ergative, EXCLAM = exclamative, FOC = focus, GEN = genitive, INCEP = inceptive aspect, INCOMPL = incompletive aspect, INDEF = indefinite, INSTR = instrumental, IPFV = imperfective aspect, textscitv = intransitive status suffix, LOC = locative, NEG = negative, NONFUT = non-future, PART = participle, PASS = passive, PFV = perfective aspect,PL = plural, POSS = possessive, POT = potential aspect, PREP = preposition, PROG = progressive aspect, PST = past, PUNC = punctual aspect, Q = question word, QUANT = quantifier, REL = relativization, RN = relational noun, RPST = recent past, SG = singular, SUF = suffix, TV = transitive status suffix.

(2) *Wh-movement of* DP_{abs} *in Kaqchikel*:

a. N-Ø-tze'en a Karlos.
INCOMPL-3SG.ABS-laugh CL Carlos
'Carlos laughs.'

b. Achike ri n-Ø-tze'en?
Q DET INCOMPL-3SG.ABS-laugh
'Who laughs?'

(3) *Wh-movement of* DP_{erg} *vs.* DP_{abs} *in K'ichee' (Mayan)*:

a. X-Ø-r-aj ri al Mari'y ri a Karlos.
COMPL-3SG.ABS-3SG.ERG-want DET CL Maria DET CL Carlos
'Carlos loved Maria.'

b. Jachin x-Ø-r-aj ri a Karlos?
who COMPL-3SG.ABS-3SG.ERG-want DET CL Carlos
'Who did Carlos love?'

c. *Jachin x-Ø-r-aj r-eech ri al
who COMPL-3SG.ABS-3SG.ERG-want 3SG.ERG-RN DET CL
Mari'y?
Maria
'Who loved Maria?'

(4) *Wh-movement of* DP_{abs} *in K'ichee'*:

a. X-Ø-kam ri a Karlos.
COMPL-3SG.ABS-die DET CL Carlos
'Carlos died.'

b. Jachin x-Ø-kam-ik?
who COMPL-3SG.ABS-die-ITV
'Who died?'

The same pattern is found in the unrelated language Kanamarí (Katukinan, Queixalos 2010). In Kanamarí, the grammatical function of a DP is signaled by case markers and word order: DP_{ext} bears an overt ergative marker and immediately precedes the verb, whereas DP_{int} and the single argument of an intransitive verb bear a phonologically zero absolutive marker and follow the verb in affirmative sentences. Ā-moved constituents are dislocated to the sentence-initial position. The single argument of an intransitive verb and DP_{int} of a transitive verb can be wh-moved (see (5-a) and (5-b)). The external argument of a transitive verb can, however, not be questioned (see (5-c)). The antipassive has to be used in order to extract DP_{erg} (see (5-d)).[4] It turns the agent DP into the sole absolutive marked DP of an intransitive verb which can then be extracted.

(5) *Wh-movement in Kanamarí* (Katukinan; Queixalos 2010):

a. hanian tu Nodia nah=hoho-nin?
who(m) Q Nodia ERG=call-DUR
'Whom is Nodia calling?'

b. hanian tu waokdyi-nin?
who(m) Q arrive.here-DUR
'Who is arriving here?'

c. *hanian tan na=dyuman tahi yu?
who here ERG-spread water Q
'Who spread water here?'

d. hanian tan wa-dyuman tahi yu?
who here AP-spread water Q
'Who spread water here?'

2.1.2. *Focus Movement*

If a DP is focussed in Mayan, it is also moved to the preverbal position. The data from K'ichee' in (6) and from Mam in (8), respectively, show the same ergative pattern as we saw with wh-movement: DP_{int} of a transitive verb can be extracted (see the b.-examples), but focussing of DP_{ext} leads to ungrammaticality (see the c.-examples). Focussing of the single argument of an intransitive verb is grammatical, see (7) and (9).

(6) *Focus movement of* DP_{erg} *vs.* DP_{abs} *in K'ichee'*:

a. K-Ø-u-loq' jun wuuj ri a Karlos.
INCOMPL-3SG.ABS-3SG.ERG-buy INDEF book DET CL Carlos
'Carlos buys a book.'

---

[4] Many Mayan languages exhibit, alongside the antipassive, yet another construction that enables extraction of the ergative argument, which is called the *agent focus* in the Mayanist literature. This construction is discussed in section 6; see Coon et al. (2011) (and references therein) on the difference between agent focus and antipassive.

b. Are ri jun wuuj k-Ø-u-loq' ri a
FOC DET INDEF book INCOMPL-3SG.ABS-3SG.ERG-buy DET CL
Karlos.
Carlos
'It is a book which Carlos buys.'

c. *Are ri a Karlos k-Ø-u-loq' ri jun
FOC DET CL Carlos INCOMPL-3SG.ABS-3SG.ERG-buy DET INDEF
wuuj.
book
'It is Carlos who buys a book.'

(7) *Focus movement of DP_{abs} in K'ichee'*:

a. Ka-Ø-tze'n-ik ri a Karlos.
INCOMPL-3SG.ABS-laugh-ITV DET CL Carlos
'Carlos laughs.'

b. Are ri a Karlos ka-Ø-tze'n-ik.
FOC DET CL Carlos INCOMPL-3SG.ABS-laugh-ITV
'It is Carlos who laughs.'

(8) *Focus movement of DP_{erg} vs. DP_{abs} in Mam (England 1989)*:

a. Ma chi kub' t-tzyu-ʔn xiinaq qa-cheej.
ASP 3PL.ABS DIR 3SG.ERG-grab-DIR man PL-horse
'The man grabbed the horses.'

b. Qa-cheej xhi kub' t-tzyu-ʔn xiinaq.
PL-horse DEP.3PL.ABS DIR 3SG.ERG-grab-DIR man
'It was the horses that the man grabbed.'

c. *Xiinaq chi kub' t-tzyu-ʔn qa-cheej.
man 3PL.ABS DIR 3SG.ERG-grab-DIR PL-horse
'It was the man who grabbed the horses.'

(9) *Focus movement of DP_{abs} in Mam (England 1989)*:

a. Ma tz-uul xiinaq.
ASP 3SG.ABS-arrive.here man
'The man arrived here.'

b. Xiinaq s-uul.
man DEP.ASP.3SG.ABS-arrive.here
'It was the man who arrived here.'

The same pattern is found in Kanamarí: DP_{abs} can be focussed (see (10-a) and (10-b)), but DP_{erg} cannot be focussed; antipassive is needed to extract the transitive agent (see (10-c) and (10-d)).

(10) *Focus movement in Kanamarí* (Queixalos 2010):
a. Maranmaran na=tyo kana tona tyo
Maranmaran GEN=daughter FOC go.away EXCLAM
'It's Maranmaran's daughter that went away.'
b. a-obatyawa kana Aro na=nuhuk kariwa
3SG-wife FOC Aro ERG=give white.man.LOC
'It's his own wife that Aro gave to the white man.'
c. *itiyan kawahiri kana na=duni tyon
this cat FOC ERG=catch rat
'It's this cat that caught the rat.'
d. itiyan kawahiri kana wa-duni tyon
this cat FOC AP-catch rat
'It's this cat that caught the rat.'

2.1.3. *Relativization*

In Jakaltek (Mayan), relativization exhibits a syntactically ergative pattern: It is possible to relativize DP_{int} of a transitive verb (see (11-a)) and the sole argument of an intransitive verb (see (11-b)), but it is impossible to relativize DP_{ext} of a transitive verb (see (11-c)).

(11) *Relativization of DP_{erg} vs. DP_{abs} in Jakaltek* (Campana 1992, Craig 1977):
a. … ch'en ome [xinliko …]
the.CLASS earrings buy.3ABS.1ERG
'… the earrings that I bought …'
b. X-Ø-w-il naj [xto ewi].
ASP-3ABS-1ERG-see CLASS go.3ABS yesterday
'I saw (the man) who went yesterday'
c. *… metx tx'i [xintx'a ni'an unin …]
the.CLASS dog bite.3ABS.3ERG little child
'… the dog that bit the child …'

Again, this pattern is also found in a number of typologically unrelated languages such as Dyirbal (Pama-Nyungan, Dixon 1994), Kanamarí (Katukinan, Queixalos 2010), Tongan (Austronesian, see Otsuka 2000, 2006).

(12) *Relativization of DP_{erg} vs. DP_{abs} in Dyirbal* (Pama-Nyungan; Dixon 1994: 169-170):

a. ŋuma-Ø [$_{CP}$ banaga-ŋu] yabu-ŋgu bura-n
father-ABS return-REL.ABS mother-ERG see-NONFUT
'Mother saw father who was returning.' *rel. of sole argument*

b. *yabu-Ø [$_{CP}$ bural-ŋu ŋuma-Ø] banaga-n^{y}u
mother-ABS see-REL-ABS father-ABS return-NONFUT
'Mother, who saw father, was returning.' *rel. of* DP_{ext}

c. yabu-Ø [$_{CP}$ bural-ŋa-ŋu ŋuma-gu] banaga-n^{y}u
mother-ABS see-AP-REL-ABS father-DAT return-NONFUT
'Mother, who saw father, was returning.' *antipassive*

(13) *Relativization in Kanamarí* (Queixalos 2010):

a. yo-hik nyan Nodia na=dahudyi-nin tukuna
1SG-know DEIC Nodia ERG=bring-DEP Indian
'I know the Indian that Nodia brought.' *rel. of* DP_{int}

b. yo-hik nyan waokdyi-nin anyan piya
1SG-know DEIC arrive.here-DEP this man
'I know the man who arrived here.' *rel. of sole argument*

c. *yo-hik nyan piya na=dahudyi-nin Hanani
1SG-know DEIC man ERG=bring-DEP Hanani
'I know the man who brought Hanani.' *rel. of* DP_{ext}

d. yo-hik nyan piya wa-dahudyi-nin Hanani
1SG-know DEIC man AP-bring-DEP Hanani
'I know the man who brought Hanani.' *antipassive*

(14) *Relativization in Tongan* (Austronesian; Otsuka 2006):

a. e fefine [na'e fili 'e Sione]
DET woman PST choose ERG Sione
'the woman (who) Sione chose' *rel. of* DP_{int}

b. *e fefine [na'e fili 'a Sione]
DET woman PST choose ABS Sione
'the woman (who) chose Sione' *rel. of* DP_{ext}

Assuming that relativization in all languages listed here involves Ā-movement (possibly of an abstract operator), this is an instance of the general pattern seen with wh-movement and focussing.

2.2. Previous Analyses

Two kinds of analyses of the ban on ergative movement have been proposed in the literature (cf. Campana 1992, Aldridge 2004, Coon et al. 2011 and Stiebels 2006). In this subsection, we discuss them briefly and illustrate some of their drawbacks. The analyses under discussion are the following:

1. Nothing is wrong with ergative movement as such; it is just that the relevant languages have a special (agent focus, AF) marker which does what the ergative marker does *and* signals the presence of an Ā-dependency at the same time. Given an optimality-theoretic approach, the agent focus construction can block the ergative+movement construction as suboptimal because the former leads to a better constraint profile than the latter (Stiebels 2006).
2. Case-driven movement (sometimes covert) of DP_{abs} blocks movement of DP_{erg}, either due to minimality (Campana 1992), or because DP_{abs} blocks the only escape hatch within vP (Aldridge 2004, Coon et al. 2011).

The problem with analysis 1 is that it only works for Mayan languages with the agent focus construction (AF). As such, it has nothing to say about languages which lack agent focus and which nevertheless show the ban on movement of the ergative argument (see section 2.1).

Analyses of type 2 have theoretical or empirical problems. To begin with, a minor technical flaw of Campana's (1992) analysis is that it is based on a non-standard concept of intervention.

Empirically, Campana (1992), Aldridge (2004), and Coon et al. (2011) all must assume that there is covert movement of DP_{abs}, which is hardly motivated on independent grounds.

Next, both Aldridge (2004) and Coon et al. (2011) must stipulate a ban on multiple vP-specifiers: The absolutive moved to the edge of a v-head can only block extraction of the ergative if v does not project another specifier that can serve as an escape hatch. However, parallel extraction of both ergative and absolutive is possible in at least some of the languages that exhibit the ban on moving the ergative in isolation (see section 5.1.2). This strongly suggests that v must be able to project multiple specifiers after all.

Furthermore, the analyses of Aldridge (2004) and Coon et al. (2011) predict that a similar movement asymmetry between co-arguments should be found in nominative-accusative languages. In their system, DP_{nom} of a transitive verb must move to the only escape hatch of v in order to get case from T. It should thus block extraction of the accusative marked DP. It is doubtful, however, whether such an asymmetry exists in accusative languages. In response to this problem, Coon et al. (2011) suggest that subjects in nominative-accusative languages are base generated outside vP while they are merged vP-internally in ergative-absolutive languages.

Finally, the type 2 analyses essentially derive an *absolutive island constraint* rather than an *ergative movement constraint*. As a consequence, the prediction is that DP_{abs} creates an island, i.e., the (covertly) moved DP_{abs} does not only block movement of DP_{erg} but movement of all elements inside vP like PP-adjuncts, DPs with oblique case, or (referential) adjuncts (which are VP-internal; see Aoun (1986)). Data from Mam in (15) and from Jakaltek in (16) show that the agent of a passivized verb and adjuncts of time and place in an intransitive context can be Ā-moved:

(15) *Wh-movement of passive agent in Mam* (England 1983*a*,*b*):
Al uʔn xhi kub' tzy-eet qa-cheej?
Q RN DEP-3PL.ABS DIR grab-PASS PL-horse
'By whom were the horses grabbed?'

(16) *Wh-movement of referential adjuncts in Jakaltek* (Craig 1977):
a. Bakin x-Ø-ul naj?
when ASP-3SG.ABS-arrive he
'When did he arrive?'
b. Bay chach yoyi?
where 2SG.ABS go
'Where are you going?'

In the analyses of Aldridge (2004) and Coon et al. (2011), this can be accounted for by assuming that intransitive vPs are never phases. Consequently, DP_{abs} does not have to move to the sole escape hatch of v to receive case and does not create an island. This accounts for the data in (15) and (16). However, a wrong prediction remains for transitive contexts, where v is always a phase. As the examples in (17-a-c) from Kaqchikel show, indirect objects, instrumental and locational constituents can be Ā-moved even in a transitive context.[5]

(17) *Wh-movement of oblique arguments in Kaqchikel:*

a. Achoq chi re n-Ø-u-ya' a Karlos
Q PREP DET INCOMPL-3SG.ABS-3SG.ERG-give CL Carlos
jun sik'wuj?
INDEF book
'To whom does Carlos give a book?'
(wh-movement of indirect object)

b. Achoq r-ik'in n-Ø-u-sël ri
Q 3SG.ERG-RN.INSTR INCOMPL-3SG.ABS-3SG.ERG-cut DET
ti'ij ri a Karlos?
food DET CL Carlos
'With what does Carlos cut the meat?'
(wh-movement of instrumental)

c. Akuchi n-Ø-u-ya' ri ti'ij
Q.3SG.ERG-RN.LOC INCOMPL-3SG.ABS-3SG.ERG-give DET food
ri a Karlos?
DET CL Carlos
'Where does Carlos put the meat?' *(wh-movement of locative)*

In view of this, the aim in what follows is to develop an account of the phenomenon (a) that derives the ban on ergative movement without predicting absolutive (and nominative) islands and (b) that relates this account to the nature of ergativity itself. The necessary background assumptions of the analysis are summarized in the following section.

---

[5] Henderson (2007) reports that extraction of certain adjuncts in Kaqchikel (including instrumental and locational adverbs) obligatorily requires the presence of the verbal marker *-wi*. (For reasons that are not clear to us, the examples from our Kaqchikel informants consistently lack *-wi*.) From this, one may conclude that extraction from vP is generally banned in Kaqchikel and that adjunct extraction from vP is exceptionally possible in the presence of *-wi*. However, Henderson (2007) also observes that there are adjuncts that do *not* require *-wi* when they undergo extraction (such as temporal adverbs and benefactives). We take this as evidence that vP in Kaqchikel is not an island. See Erlewine (2012) for further intriguing observations about extraction in Kaqchikel, which we have nothing to say about here.

3. Assumptions

3.1. Clause Structure

We adopt the following standard minimalist clause structure:

(18) $[_{\mathrm{CP}}\ \mathrm{C}\ [_{\mathrm{TP}}\ \mathrm{T}\ [_{\mathrm{vP}}\ \mathrm{DP}_{ext}\ [_{\mathrm{v'}}\ \mathrm{v}\ [_{\mathrm{VP}}\ \mathrm{V}\ \mathrm{DP}_{int}\ ]]]]]$

The internal argument is the sister of V, whereas the external argument is introduced as the specifier of v (Chomsky 1995, Kratzer 1996). There are two functional heads above v, viz., T and C. However, the projection of C will not occur in the following trees since it does not play an important role in the analysis of the ban on ergative movement.

3.2. Operations

All syntactic operations are feature-driven. The two basic operations are Merge for structure building (external and internal Merge) and Agree for argument encoding by case assignment/agreement. These are triggered by the following features (Heck and Müller 2007 and references cited there):

(19) *Two types of features that drive operations*:

a. Structure-building features (edge/subcategorization features) [•F•] trigger Merge.

b. Probe features [∗F∗] trigger Agree.

We take it that Agree and Merge both take place under *m-command* (i.e., Agree may affect a head and its specifier). Next, the Agree Condition and the Merge Condition in (20) and (21) demand that probe and structure building features are checked (application of these constraints at each derivational step derives the effects of the Earliness Principle; Pesetsky 1989).

(20) Agree Condition (AC):
Probes ([∗F∗]) participate in Agree.

(21) Merge Condition (MC):
Structure-building features ([•F•]) participate in Merge.

3.3. Locality of Movement

A crucial assumption of the analysis in section 4 is that Ā-movement to SpecC must make an intermediate stop in SpecT. This can be ensured in various ways: either by assuming that TP is a phase (Richards 2011), by stipulation (Chomsky 2005, Boeckx and Grohmann 2007), or by assuming that every phrase is a phase (for successive-cyclic movement through all intermediate phrase edges see, e.g., Sportiche 1989: 36, 45-47 Takahashi 1994, Boeckx 2003: 16-25, Müller 2004, Chomsky 2005: 18). We follow the last proposal and assume that movement takes place successive-cyclically, from one XP edge domain to the next one higher up. Given the Phase Impenetrability Condition (PIC; Chomsky 2001; (22)), this follows if every XP is a phase.

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

(23) *Edge*:
The edge of a head X comprises all specifiers of X (and adjuncts to XP).

In a model of syntax where all operations are feature-driven, it must be ensured that intermediate steps of movement, like movement to the edge domain of a phase as required under the PIC, are possible in the first place. A standard assumption is that an *edge feature* [•X•] (Chomsky 2007, 2008) that triggers intermediate movement can be inserted on any intervening phase head.

Departing from the standard, we assume that there is no minimality condition on Agree or Merge. Rather, we take it that minimality effects are derivable from other principles of grammar, such as the PIC (Chomsky 2001: 47, footnote 52; Müller 2004, Müller 2011). This means that if there is more than one DP in an accessible domain that can be attracted or agreed with, then in principle any of them can be targeted by the operation-inducing head.

3.4. Assignment of Structural Case

Every argument must receive abstract structural case in the syntax, otherwise the derivation crashes (Rouveret and Vergnaud 1980). Structural case is assigned by the functional heads v and T to argument DPs under Agree. This means that T and v, respectively, have valued case probe features [$*$c:$\alpha*$] that assign their value α to DPs with an unvalued case feature [c:□]. We follow a proposal by Murasugi (1992) (see also Jelinek 1993, Ura 2000: 206, Müller 2009), according to which in morphologically ergative as well as accusative languages T assigns the unmarked structural case (i.e., nominative = absolutive) and v assigns the marked structural case (i.e., ergative = accusative).[6] In intransitive contexts only the T head is active, so the single argument receives the unmarked case.[7] More specifically, assume that there is a single structural case feature *case*, abbreviated as 'c'. This feature can have the two values ext(ernal) and int(ernal), determined with respect to the vP, the predicate domain.[8] The unmarked case (nominative/absolutive) is represented as the external case [c:ext] and the marked case (ergative/absolutive) as the internal case [c:int]. Since T assigns unmarked external case and v assigns the marked internal case, these heads bear the following probe features:[9]

---

[6] This assumes that the ergative is a structural case. See Nash (1996), Alexiadou (2001), Woolford (2001, 2006), Legate (2008) for the opposite view. Woolford (2001, 2006) and Legate (2008) also assume that ergative is assigned by v; the only relevant difference is that they postulate that ergative assignment must go hand in hand with θ-assignment. Sometimes, it has been argued that ergativity has different sources (Aldridge 2004, Paul and Travis 2006, Legate 2008; see also footnote 1). The working hypothesis here is that morphological ergativity, at least in the languages that show the ban on Ā-moving the ergative argument, has a uniform base (see section 3.5).

[7] There are at least two other recent proposals on how to derive the difference between ergative and accusative alignment patterns that we will not further pursue here: (a) T assigns nominative=ergative, v assigns accusative=absolutive (Levin and Massam 1985, Chomsky 1995: ch.3, Bobaljik 1993, Laka 1993, Řezač 2003, Bobaljik and Branigan 2006); (b) T assigns ergative, v assigns accusative, nominative=absolutive is default (Bittner and Hale 1996).

[8] The concept external/internal case is independent of the concept external/internal argument: Both the external and the internal argument may, in principle, bear either external or internal case (depending on the alignment pattern).

[9] Throughout this paper, we assume that Agree results in valuation: DPs enter the derivation without a case value and get this value under Agree with a probe which provides a value. Note that this is the reverse of what is standardly assumed for Agree in phi-features where the goal provides the values for the probe. This is due to the nature of the feature case; case is not an inherent feature of DPs, in contrast to phi-features; rather, case is assigned to DPs (cf. Adger 2003, Pesetsky and Torrego 2007). We take case to be uninterpretable on both probe and goal.

(24) *Case probe features on T and v*:

a. T bears a probe [$*$c:ext$*$] that instantiates a matching [c:ext] goal on DP.

b. v bears a probe [$*$c:int$*$] that instantiates a matching [c:int] goal on DP.

We assume that argument encoding by case or agreement is the result of the same syntactic operation: Both case marking and verbal agreement are instances of an Agree relation that involves the feature *case*. The only difference is the locus of the morphological realization of this relation:[10]

(25) *Argument encoding by case or agreement*:

a. Argument encoding proceeds by case-marking if [c:α] is realized on DP.

b. Argument encoding proceeds by agreement if [$*$c:$\alpha*$] is realized on T/v.

3.5. Patterns of Argument Encoding

In ergative languages, DP$_{int}$ of a transitive verb and the sole argument of an intransitive verb (DP$_{int}$ or DP$_{ext}$) are treated alike, but differently from DP$_{ext}$ of a transitive verb. In accusative languages, DP$_{ext}$ of a transitive verb and the sole argument of an intransitive verb (DP$_{int}$ or DP$_{ext}$) cluster together: They bear nominative case or trigger the same kind of agreement. DP$_{int}$ of a transitive verb behaves differently; it receives accusative case or is cross-referenced by a different set of agreement markers. The question is how the difference between ergative and accusative encoding patterns can be derived if v assigns the marked case and T assigns the unmarked case in both types of languages. We adopt the analysis of argument encoding patterns proposed by Müller (2009) (see also Heck and Müller 2007), which relies on the *timing* of elementary operations. It turns out that the assumptions needed to derive the two basic

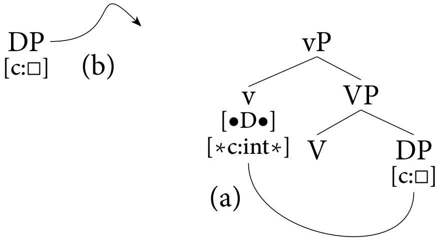
---

[10] In some languages, there is a one-to-one relation between case marking and agreement. In other languages, case/agreement mismatches may arise: Sometimes there is agreement with only a single argument or the resulting agreement pattern need not be identical to the one established for case (in particular, the case pattern may be ergative and the agreement pattern accusative). A possible analysis of such phenomena relies on delinking Agree for case and phi-features: In addition to case probes, there is secondary, purely phi-based Agree.

encoding patterns are also suited to account for the movement asymmetries described in section 2.1.

According to Müller (2009), ergative vs. accusative patterns of argument encoding result from different resolutions of conflicting earliness requirements for Agree and Merge on the vP level. The conflict emerges because v has a dual role in the present system: It participates in a Merge operation with DP$_{ext}$ and it also participates in an Agree relation with some DP with respect to case. It thus bears two operation-inducing features: [•D•] and [∗c:int∗]. Consider a simple transitive context with the two arguments DP$_{int}$ and DP$_{ext}$. Suppose that the derivation has reached a stage Σ where v has been merged with a VP containing DP$_{int}$, with DP$_{ext}$ waiting to be merged with vP in the workspace of the derivation. At this point, a conflict arises: AC (= (20)) demands that the next operation is Agree (case assignment) between v and DP$_{int}$, which is the only potential goal at this point of the derivation, (see (a) in (26)); MC (= (21)) demands that the next step is Merge of DP$_{ext}$ in Specv (see (b) in (26)).

(26) *Stage Σ:*



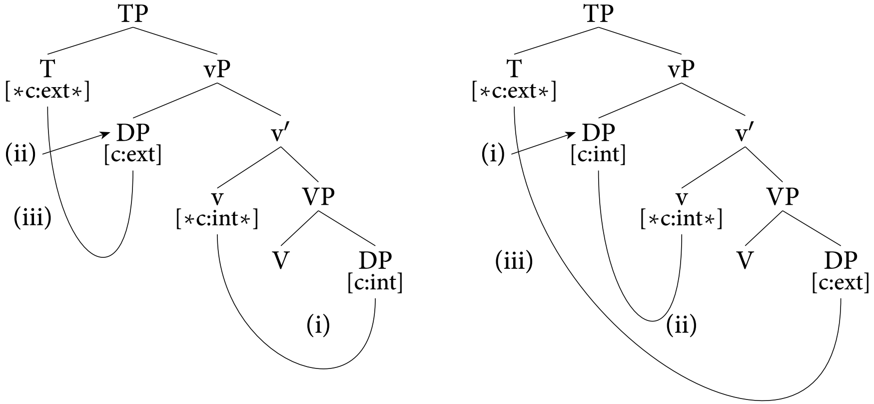
Assuming that only a single operation can apply at any stage of the derivation (pace Chomsky 2008), AC and MC need to be ordered.[11] This ordering has consequences for argument encoding. If Agree takes priority over Merge, then an accusative pattern arises; if Merge takes place before Agree, then an ergative pattern emerges. More precisely, the two patterns of argument encod-

---

[11] Three ways to resolve a conflict w.r.t. the order in which operations apply are thinkable: (a) The order is fixed (cf. "Merge over Move" in Chomsky 2000), (b) the order is free, or (c) operations apply simultaneously (Chomsky 2008). We adopt (a), assuming that Merge and Agree are ordered in a language-specific manner; as we will see, this has consequences for the argument encoding pattern *and* the extractability of core arguments in a language. Solution (b) is incompatible with the idea that operations apply as soon as their context of application is fulfilled (see Pesetsky 1989, Chomsky 1995: 233, Lasnik 1999: 198, among others). Otherwise, the "free" order between Merge and Agree, in fact, is a disjunction (Merge applies before Agree or the other way round). Simultaneous rule application as in (c) is at variance with a strictly derivational approach to syntax (see Brody 2002, Epstein and Seely 2002).

ing are derived as follows: If Agree applies before Merge, then v first assigns the internal case to DP$_{int}$ via Agree. DP$_{int}$ is the only available goal at this step of the derivation. [c:int] is then called accusative. In a second step, DP$_{ext}$ is merged. In a subsequent step, T is merged and DP$_{ext}$ receives the external case, called nominative (see the a.-derivation in (27)). An accusative pattern emerges. If, however, Merge takes priority over Agree, the structure-building feature [•D•] on v triggers Merge of DP$_{ext}$ first. After this, the case probe feature on v triggers Agree and assigns [c:int] to DP$_{ext}$; this case is usually called ergative. DP$_{int}$ later receives the unmarked [c:ext] from the case probe on T (see the b.-derivation in (27)); [c:ext] is called absolutive in this environment.

(27) a. *Agree before Merge: accusative* b. *Merge before Agree: ergative*



The derivation of the ergative pattern presupposes that a head prefers Agree with its specifier to Agree with an item included in the complement of that head: If DP$_{ext}$ is merged before v triggers Agree, it is DP$_{ext}$ in Specv that is assigned case by v, although DP$_{int}$ included in the complement of v is in the m-command domain of v, too, and has not yet been assigned a case value.[12] This preference for agreement with a specifier can be formulated as the *Specifier-Head-Bias* (cf. Chomsky 1986: 24-27, Chomsky 1995: 149, Kayne 1989, Koopman 1992: 557, and Koopman 2006; a similar idea, with the bias inverted, is presented in Béjar and Řezáč 2009.)

---

[12] The b-derivation of (27) presupposes either (a) that Agree can escape the PIC (under the assumption that every phrase is a phase), as suggested by Bošković (2007), among others, or (b) that the PIC is slightly less restrictive, as eventually proposed in Chomsky (2001).

(28) *Specifier-Head Bias (Spec-Head Bias)*:
Spec/head Agree is preferred to Agree under c-command.

Since Agree takes place under m-command, a situation may arise in which there are two goals in the m-command domain of a probe on a head α, viz., if there is a DP in the specifier of α and a DP in the c-command domain of α. The Spec-Head Bias states that in this situation Agree with the DP in the specifier of α is preferred over Agree with the DP in the c-command domain of α.[13] This critical situation emerges in languages with the order Merge before Agree on v after DP_{ext} is merged. The consequence of the Spec-Head Bias is that the internal case is assigned by v to DP_{ext} in Specv instead of to DP_{int} in the complement of v, resulting in an ergative alignment pattern. We take the Spec-Head Bias to replace standard minimality conditions like Relativized Minimality or the MLC (though with a somewhat different empirical coverage).

As mentioned before, only the T head is active in intransitive contexts both in languages with an ergative and with an accusative encoding pattern. As a consequence, the unmarked external case will be assigned to the single argument and an ergative or accusative encoding pattern emerges, depending on whether the single argument receives the same case as the internal or the external argument of a transitive verb.

With this final remark we finish the illustration of the analysis of argument encoding patterns developed in Müller (2009). In section 4 we will see that the same indeterminacy with respect to the order of elementary operations that emerges on the vP cycle also holds on the TP cycle because T triggers both Merge and Agree if one of the arguments of a transitive verb is to be extracted. Interestingly, if the indeterminacy on T is resolved in the same way as the indeterminacy on v (where it leads to morphological ergativity and accusativity, respectively), the ban on ergative movement in morphologically ergative languages and the absence of the corresponding effect in morphologically accusative languages follows automatically.

---

[13] This preference could also be derived by assuming that the probe agrees with the goal which is closer to α provided a notion of closeness that is based on a definition of path length from which it follows that the path from α to Specα is shorter than the path from α to an element in the complement domain of α (see e.g. Heck and Müller 2007). Here, we opt for the Spec-Head Bias which is compatible with equi-distance effects, which in turn pose a problem for path-based definitions of minimality.

3.6. Maraudage

A final assumption that is necessary to account for the extraction asymmetries described in section 2.1 concerns the behaviour of structural case features. Suppose that an argument can check more than one structural case feature (see Merchant 2006). This means that after a DP has received a structural case value, it is still an active goal for another structural case probe:

(29) *Activity of structural case features*:
Structural case features act as active goals.

Independent motivation for this assumption might come from the existence of *case stacking* (see Andrews 1996, Nordlinger 1998, Richards 2013, Assmann et al. 2013, see also Merchant 2006 and references therein). We take checking of [c:int] on a DP α with a conflicting value on a probe such as [$*$c:ext$*$] to be harmless as such; α will simply maintain its original case value. However, [$*$c:ext$*$] is then discharged and not available for further operations anymore.

In a transitive context with two structural case probes, the fact that a DP can check more than one structural case feature can lead to a situation where a DP α that already got a case value from probe P_1 also checks the case feature of probe P_2. As a consequence, the co-argument of that DP cannot receive case, which leads to the crash of the derivation. Put differently, DP α uses up a case feature that it does not need (because it already has a case value), but that would be absolutely necessary for its co-argument. We call this taking away of features that should normally be reserved for some other item "maraudage" (see Georgi et al. 2009, Georgi 2012, Müller 2011 on maraudage).[14]

In the present system, maraudage occurs in the following situation: Suppose there is a head γ which triggers Merge of a DP_1 and Agree for case. Under the order Merge before Agree, the structure-building feature of γ is discharged first and a DP is merged in Specγ. Due to the Spec-Head Bias, γ next checks its case probe with DP_1 in its specifier, although there may be another potential goal DP_2 in the complement domain of γ. Now, if DP_1 has already gotten a case value earlier in the derivation, it marauds the case feature of γ, with fatal consequences for its co-argument DP_2, which does not receive a case

---

[14] Similar concepts are suggested in Chomsky (2001: 15), Abels (2012: 105-108), Anagnostopoulou (2003: 272-274), Adger and Harbour (2007: 26), Béjar and Řezáč (2009), Heck and Richards (2010: 10); see also Trommer (2011) and Zimmermann (2013) for morphophonology.

value. Hence, DPs trigger maraudage in Spec-Head-configurations under the ranking Merge before Agree. The situation is abstractly depicted in (30):

(30) a. $[_{\text{XP}}\ \text{X}_{[*\text{c:ext}*]}\ [_{\text{ZP}} \dots \alpha_{[\text{c:int}]} \dots \beta_{[\text{c:}\square]} \dots]]$
b. $[_{\text{XP}}\ \alpha_{[\text{c:int}]}\ [_{\text{X}'}\ \text{X}_{[*\text{c:ext}*]}\ [_{\text{ZP}} \dots \text{t}_{\alpha} \dots \beta_{[\text{c:}\square]} \dots]]]$

In (30-a), an ambiguity arises: [∗c:ext∗] may be checked by either α or β because (a) there is no minimality condition on Agree, (b) both goal DPs can check structural case and (c) both DPs are in the c-command domain of the head X. If β checks the case feature, the derivation converges because both elements have structural case. If, however, α checks case with X, the derivation crashes because β is left with an unvalued structural case feature. Importantly, there is one converging derivation based on this configuration. In contrast, in (30-b) there is no ambiguity because α is in SpecX whereas β is in the c-command domain of X: Due to the Spec-Head Bias, X must assign Case to α. But since α already has structural case, it marauds the case feature that β needs, and hence this derivation crashes. Note that maraudage of case features is expected given (29); preventing it would require further stipulation.

The configuration in (30-b) will inevitably arise on the TP cycle in morphologically ergative languages if DP_{erg} is Ā-moved, given that Merge is preferred over Agree in the clausal domain in this language type. This will be shown to underlie the ban on ergative movement.

4. Analysis

The difference between morphologically ergative and accusative languages is explained by the order of the elementary operations Merge and Agree. Recall that this ordering of operations is necessary because there is an indeterminacy at the stage of the derivation where v is merged: It has a probe feature triggering Agree as well as a structure-building feature triggering Merge, but it can induce only a single operation at once. Crucially, the same indeterminacy may arise with T, given the assumptions laid down in section 3: If a DP is to be Ā-moved to SpecC, it must make an intermediate stop in SpecT, due to the PIC. This movement step is triggered by a category-neutral edge feature [•X•] instantiated on T. However, T also triggers Agree because it bears [∗c:ext∗]. Thus, if an element is to be Ā-moved to SpecC, then T bears two operation-inducing features, one that triggers Merge and another one that triggers Agree.

Hence, an ordering of the two operations is not only necessary for v but also for T. We make the natural assumption that the order of Merge and Agree that holds on the vP cycle is also maintained on the TP cycle; more generally: the same conflict resolution strategy is manifested throughout the extended projection (see Lahne 2008 for an application of this idea to a different empirical domain, viz., word order). This means that Agree is given preference over Merge in the case of conflict on the TP cycle in accusative languages, and Merge preempts Agree on the TP cycle in ergative languages. Together with the concept of maraudage and the Spec-Head Bias, this derives the ban on ergative movement in morphologically ergative language and the absence of extraction asymmetries in accusative languages.

4.1. Displacement in Languages with Ergative Encoding Patterns

4.1.1. Illegitimate Movement of the Ergative DP

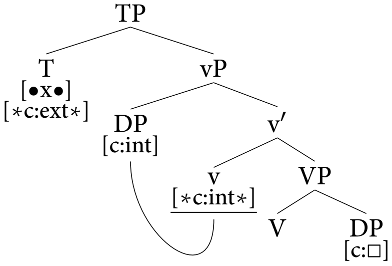
Suppose that the external argument of a transitive verb in a morphologically ergative language is to be extracted. In this type of language Merge takes priority over Agree. Thus, once v is introduced into the structure, it triggers Merge of the external argument. Afterwards, it assigns [∗c:int∗] to the external argument in its specifier (due to the Spec-Head Bias), see (31-a). Given the PIC, DP_{erg} must move from Specv to SpecT if it is to undergo subsequent Ā-movement to SpecC. Given that the "ergative" conflict resolution strategy Merge before Agree is also maintained on the TP cycle, internal Merge of DP_{erg} to the edge of T will have to *precede* Agree of T with a DP, see (31-b). Given the Specifier-Head-Bias, DP_{erg} in SpecT will maraud T's case probe (although it has already received case from v). The internal argument DP remains without a checked case feature, see (31-c). Assuming that all DPs must have their case features checked eventually (and that there is no such thing as a default case in standard transitive contexts), the derivation will crash. This derives the ban on ergative movement. In a nutshell, ergative movement is impossible because it deprives the remaining argument of absolutive case; movement of DP_{erg} per se is unproblematic.[15]

---

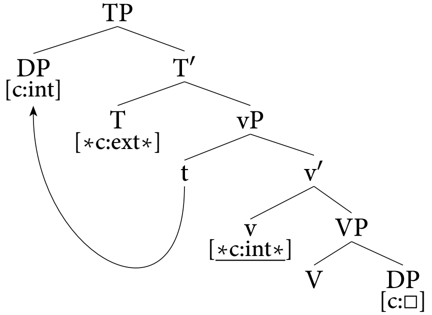
[15] In the following tree structures underlining signals a discharged probe; discharged edge features are not represented anymore; traces are only inserted as mnemonic devices.

(31) *Illegitimate movement of DP$_{erg}$:*

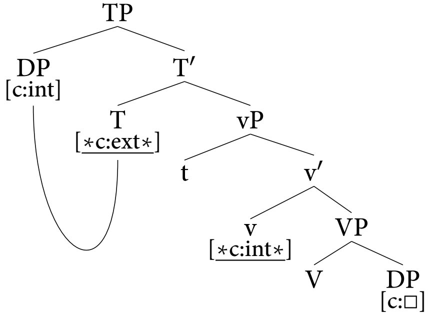
a. Structure after T is merged



b. Merge before Agree triggers movement of DP$_{erg}$ first



c. Specifier-Head Bias triggers maraudage of T



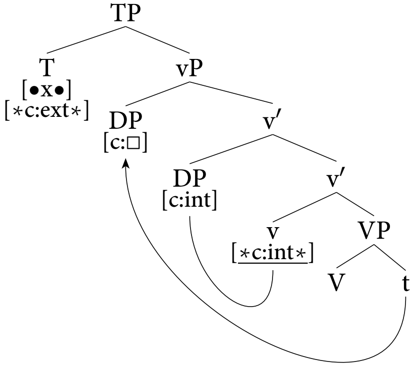
4.1.2. *Legitimate Movement of the Absolutive DP*

No such problem arises for movement of DP$_{abs}$. On the vP cycle in (32-a), the order Merge before Agree ensures that external Merge of DP$_{ext}$ and subsequent internal Merge of DP$_{int}$ (triggered by [•D•] and [•X•] on v) precede Agree. Movement of DP$_{int}$ to Specv is necessary because vP is a phase and

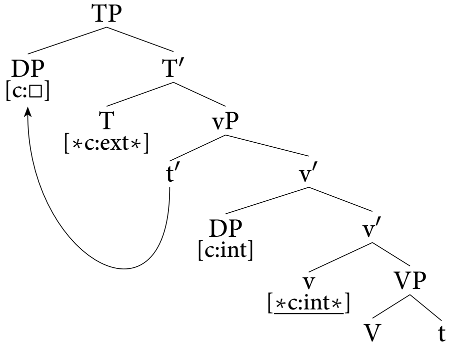
DP$_{int}$ would otherwise be trapped in the domain of the phase head. Next, v assigns [c:int] to DP$_{ext}$ in its specifier. Afterwards, T is introduced. Given that Merge applies before Agree, T first discharges its edge feature and attracts DP$_{int}$, which does not yet have a case value, see (32-b). Then T triggers Agree and due to the Spec-Head Bias it assigns [c:ext] (absolutive) to the DP in its specifier (32-c). Finally, DP$_{abs}$ moves to its final landing side SpecC. The derivation converges because both arguments receive structural case. It is thus possible to Ā-move DP$_{abs}$; DP$_{erg}$ has already been assigned case when DP$_{abs}$ moves to SpecT. Hence, maraudage does not take place.

(32) *Legitimate movement of DP$_{abs}$:*

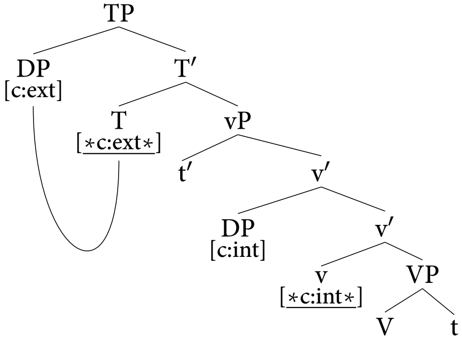
a. Structure after T is merged



b. Merge before Agree triggers movement of DP$_{abs}$ first



c. Finally, Agree with T ensures external case of DP_{abs}; no maraudage



Note that on the vP cycle, when both DP_{ext} and DP_{int} occupy a Specv position, optionality arises. Since there is no MLC-like constraint, the Spec-Head Bias does not discriminate between the two arguments and the derivation can proceed in two ways: If Agree takes place between v and DP_{ext}, a well-formed output results, see above; if, however, v Agrees with DP_{int} and assigns internal case to it, the derivation crashes because DP_{int}, which now bears [c:int], also marauds the external case assigned by T once it occupies SpecT. DP_{ext} is then left without case.

To summarize, an ergative DP_{ext} cannot be Ā-moved because intermediate movement to SpecT leads to maraudage: It applies *before* T can assign external case to DP_{int}, which needs the case value. The ergative DP thus moves too early. DP_{abs}, however, can be extracted because DP_{ext} is already assigned case within vP. In the following subsection we show that no extraction asymmetries arise in morphologically accusative languages; both DP_{int} and DP_{ext} can be Ā-moved.

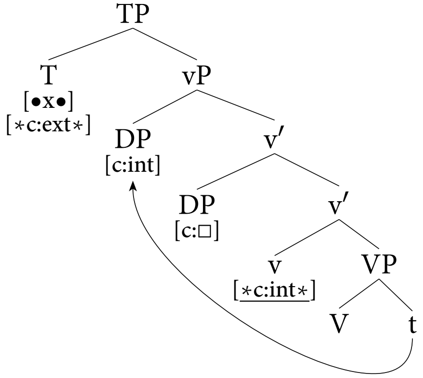
4.2. Displacement in Languages with Accusative Encoding Patterns

4.2.1. Legitimate Movement of the Accusative DP

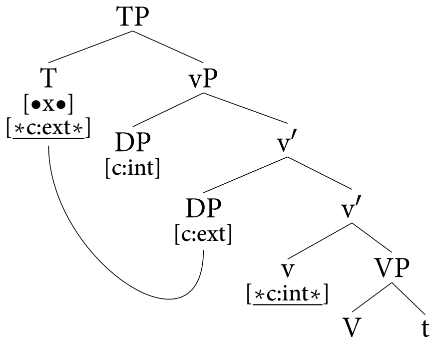
Suppose that the accusative marked DP is to be Ā-moved. The conflict resolution strategy Agree before Merge gives rise to an accusative pattern: v assigns the internal case to DP_{int} *before* DP_{ext} is merged. Next, DP_{int} moves to the edge of v to escape the vP-phase, see (33-a). Agree before Merge is also active on the TP cycle. Here it ensures that Agree with DP_{ext} in Specv can be carried out *before* DP_{int} undergoes successive-cyclic movement to SpecT (and then to a higher position), see (33-bc). This derivation converges because both arguments receive structural case. Note that at the point where T triggers Agree, there are two possible goals: If T assigns case to DP_{ext}, a well-formed output results. Since there is another DP in the c-command domain of T and there is no MLC-like constraint, T could also assign the case value to DP_{int}. However, this derivation crashes because DP_{ext} never gets case.

(33) *Legitimate movement of* $\textit{DP}_{acc}$*:*

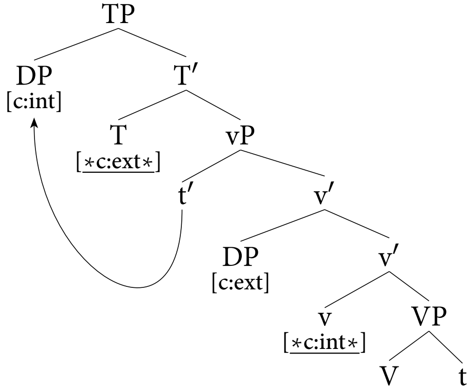
a. Structure after T is merged



b. Agree before Merge ensures external case of DP_{nom} first; no maraudage



c. Finally, movement of DP$_{acc}$ takes place to SpecT

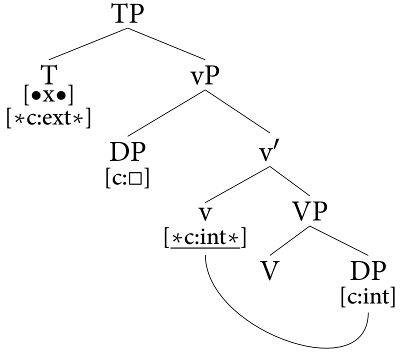


4.2.2. Legitimate Movement of the Nominative DP

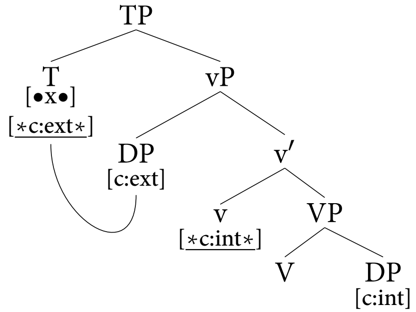
Similarly to movement of DP$_{abs}$, there is no problem for movement of DP$_{nom}$ because DP$_{acc}$ has already been assigned case when DP$_{nom}$ moves to SpecT and hence DP$_{nom}$ cannot cause maraudage. The initial step, the assignment of [c:int] to DP$_{int}$ (= accusative), is shown in (34-a). Then T assigns case to DP$_{ext}$ (= nominative) before DP$_{ext}$ moves to SpecT, see (34-b) and (34-c). Since both arguments receive structural case, the derivation converges. Note that T could in principle also assign case to DP$_{int}$ because both DPs are in the c-command domain of T and there is no MLC-like constraint. Again, this derivation crashes because DP$_{ext}$ does not receive structural case.

(34) *Legitimate movement of DP$_{nom}$:*

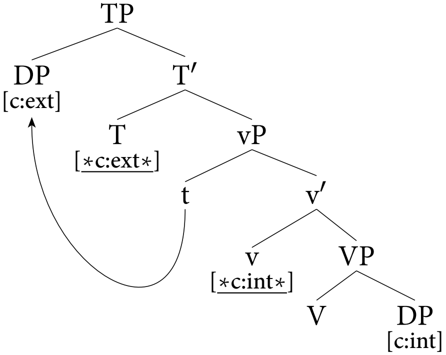
a. Structure after T is merged



b. Agree before Merge triggers valuation of DP_{nom} next



c. Finally, movement of DP_{nom} takes place to SpecT



4.3. Opacity

The above analysis instantiates an interesting case of opacity (Chomsky 1951, 1975, Kiparsky 1973, Arregi and Nevins 2012). The term opacity characterizes rule interactions that are not transparent: In cases of counter-feeding, a certain rule has not applied although its context is given; in cases of counter-bleeding, a rule has applied although its context is not given (see Kiparsky 1976). In our case, we are dealing with an instance of counter-bleeding of the following abstract pattern: There is a rule R_1 (Agree between T and a DP in what follows) that changes a structure AB into AC, and there is a rule R_2 (edge feature-driven Merge to SpecT in what follows) that changes A into D. Now, if an output DC is derived from an input AB, then both rules R_1 and R_2 must have applied. However, it is not clear from DC why rule R_1 could apply at all because its application context has been destroyed by application of R_2, i.e., rule R_2 should bleed R_1, but R_1 applies nevertheless (hence counter-bleeding). The only way to change AB into DC is to apply rule R_1 first and change AB into AC and then apply rule R_2 to change AC into DC. In the remainder of this section, we will discuss bleeding in ergative languages and an instance of counter-bleeding in accusative languages that arise through the interaction of Merge and Agree.

Consider first the derivation in which DP_{erg} is to be extracted, see (31). Merge of DP_{erg} (rule R_2) to SpecT *bleeds* Agree between T and DP_{abs} (rule R_1): Given that (internal) Merge of DP_{erg} precedes Agree due to the ergative order Merge before Agree (i.e., R_2 precedes R_1), and given the Spec-Head Bias, T must Agree with DP_{erg}. DP_{erg} therefore marauds the case feature that DP_{abs} would need. Agree between T and DP_{abs} is thus fatally prevented. Compare this with the derivation in (33) in which DP_{acc} is to be extracted. Movement of DP_{acc} to SpecT (rule R_2) creates a configuration that, on the surface, is identical in all relevant respects to the configuration that leads to maraudage if DP_{erg} is moved (compare (31-c)): There is a DP in SpecT that already has a case feature (assigned within vP), and T has a case probe. Thus, we might expect that DP_{acc} marauds the case feature of T just as DP_{erg} does in the same context. Hence, movement of DP_{acc} should bleed R_1, i.e., Agree between T and DP_{nom}. This should lead to the crash of the derivation. However, this is not the case; as we have seen, it correctly follows from the present approach that it is possible to extract DP_{acc}. The reason is that internal Merge of DP_{acc} in SpecT *counter-bleeds* Agree between T and DP_{nom}. Counter-bleeding results because the order of Merge and Agree, which stand in a bleeding relation if Merge applies before Agree, is inverted such that Agree preempts Merge (i.e., R_1 precedes R_2). The result is that movement of DP_{acc} to SpecT, which could potentially cause bleeding, comes too late; T has assigned case before DP_{acc} moves.

Note that the derivational order that creates counter-bleeding cannot be reconstructed by just looking at the output representation on the TP cycle: DP_{acc} in SpecT *does* occupy the preferred position for case valuation with T, compared with DP_{nom} in Specv, and there is no representational way to recover that DP_{acc} got there only after DP_{nom} was assigned case. Thus, unlike most other cases of syntactic opacity, which can be accounted for by positing devices like traces (like, e.g., *wanna*-contraction in Bresnan 1978 or reconstruction in Barss 1986), the opacity discussed here is of a type that cannot be accounted for in representational terms, at least not straightforwardly so. As such, it presents strong evidence for the derivational nature of syntax.[16]

Indeed, closer inspection reveals that *both* rule interactions discussed here are strictly speaking opaque because their effects cannot be read off the final output representations. The bleeding case additionally gives rise to a counter-feeding configuration: Movement of DP_{erg} to SpecC (its final landing site) could feed Agree between T and DP_{int}, but it does not. From looking at the final configuration, it is unclear why case assignment from T to DP_{int} is not available, given that DP_{ext} is not in SpecT anymore. The counter-feeding effect can be accounted for if traces are present, unlike the counter-bleeding effect with accusative movement.[17]

5. Predictions and Outlook

5.1. Predictions

The analysis presented in section 4 makes two falsifiable predictions: (a) The sole argument of an intransitive verb that bears ergative case/triggers ergative agreement should be extractable, and (b) the derivation converges if both arguments of a transitive verb are Ā-moved. In this subsection we illustrate that these predictions are borne out empirically.

5.1.1. *Extractability of the Sole Ergative Marked Argument of an Intransitive Verb*

The present analysis of the ban on ergative movement is a co-argument-based approach: Ā-movement of DP_{erg} is unproblematic per se, but it creates prob-

---

[16] Another case of this rare type of opacity is presented in Lechner (2010).

[17] As a matter of fact, opacity not only arises on the TP level, as discussed in the main text, but also on the vP level, in the derivation of the accusative pattern, given the system of case assignment in Müller (2009), Heck and Müller (2007) and the Spec-Head Bias: As soon as the external argument is merged in the specifier of v, it should be assigned the internal case of v due to the Spec-Head Bias and hence bleed assignment of the internal case to the internal argument (which would ultimately result in an ergative alignment pattern). However, DP_{int} does receive the internal case. Merge of DP_{ext} thus counter-bleeds internal case assignment. In the present analysis, this is again due to the order of the elementary operations Merge and Agree. In morphologically accusative languages, Agree applies before Merge, such that assignment of the internal case takes place before DP_{ext} is merged. At the point when the Spec-Head Bias could have an effect, Agree, i.e., case assignment by v, has already applied.

lems for the co-argument of DP_{erg}, which cannot get case. Crucially, the extraction asymmetry is not an effect of being ergative marked alone under this perspective. The account thus predicts that in a language with the ban on ergative movement in transitive clauses, the single argument of an intransitive verb which is ergative marked should be able to undergo Ā-movement. This is the case because there is no co-argument in the structure for which movement of the single ergative marked DP could have fatal consequences. Data from Mayan languages provide evidence that the prediction is correct. Some Mayan languages exhibit an aspect-based split with intransitive verbs. Usually, the single argument of an intransitive verb triggers absolutive agreement like the internal argument of a transitive verb does (leading to an ergative alignment pattern). In the imperfective/progressive aspect, however, the single argument is cross-referenced by the same affixes (the ergative affix set) as the external argument of a transitive verb (the accusative alignment pattern). This means that one and the same verb can bear the ergative and the absolutive affix set, depending on aspect; see the examples from Yukatek in (35). Aspect has no influence on the alignment pattern of transitive verbs: Here, DP_{ext} always triggers ergative agreement and DP_{int} absolutive agreement (36).[18]

---

[18] In the system of ergative vs. accusative alignment patterns presented in section 3.5, the single argument receives the unmarked case from T because v is not active in intransitive contexts. The question arises how aspect-based splits can be integrated into this analysis. One possibility is to assume that v can be reactivated in the imperfective/progressive aspect. Suppose that aspect is located on T: T with imperfective/progressive aspect selects only an active vP, whereas T with perfective aspect selects an inactive vP. Since v is merged before T, the single argument introduced within vP would get the marked case (the ergative in Mayan). It will also be assigned the unmarked case by T later on, but this will have no effect on the morphological realization as ergative because we assumed that a DP which checks multiple cases maintains the value of the first case feature it checks. An alternative analysis is proposed by Larsen and Norman (1979), Bricker (1981), Coon (2010*b*). They suggest that the imperfective/progressive aspect marker embeds a nominalized verbal projection. Thus, 'I am sleeping' is essentially a possessive structure meaning 'my sleeping is going on'. In Mayan languages, the possessum bears an affix that cross-references the possessor. The set of affixes used with possession is the same set that is used to cross-reference DP_{erg} on a transitive verb. It thus follows that "ergative" markers occur in imperfective/progressive aspect – they are possessive affixes (see also Smith-Stark 1976, Furbee-Losee 1976, Ayres 1981). This analysis is also compatible with the theory presented in section 4. Further accounts of aspect-based splits that are compatible with the present analysis can be found in Müller (2009) and Coon and Preminger (2012).

(35) *Yukatek, aspect split with intransitives (Bohnemeyer 2004: 18)*:

a. K-u=kim-il.
IPFV-3SG.ERG=die-INCOMPL
'He dies.'

b. H=kim-Ø-ih.
PFV=die-COMPL-3SG.ABS
'He died.'

(36) *Yukatek, no aspect split with transitives (Bohnemeyer 2004: 18)*:

a. K-u=hats'-ik-en.
IPFV-3SG.ERG=hit-INCOMPL-1SG.ABS
'He hits me.'

b. T-u=hats'-ah-en.
PFV-3SG.ERG=hit-COMPL-1SG.ABS
'He hit me.'

There are at least four Mayan languages that have both the ban on ergative movement and an aspect-based split with intransitives: Yukatek, Pocoman, Ixil and Chuj. We tested the prediction with examples from the two latter languages.[19] Ixil has four aspects: potential, inceptive, punctual and durative. In the latter, the single argument of an intransitive verb triggers ergative agreement like DP$_{ext}$ of a transitive verb. In the other aspects, it triggers absolutive agreement like DP$_{int}$ of a transitive verb, (cf. Lengyel 1978). The ban on ergative movement can be exemplified with constituent negation. If a DP is negated in Ixil, it is preceded by the negative element *yeʔl*, and the constituent [neg+DP] must be Ā-moved into the preverbal position (an instance of overt quantifier raising). This position is also targeted by wh-words and focussed constituents. (37-a) and (37-b) show that the absolutive marked DP$_{int}$ of a transitive verb can be negated, whereas the ergative marked DP$_{ext}$ cannot be negated. The single absolutive marked argument (in punctual aspect) can also be negated and extracted (see (37-c)), giving rise to an ergative pattern of Ā-movement. Crucially, the single ergative marked argument of an intransitive verb (in durative aspect) patterns with the absolutive marked DPs in that it can be negated, see (37-d).

---

[19] In Yukatek and Pocoman the use of agent focus is optional; DP$_{ext}$ can also be freely extracted. Therefore, these languages do not tell us much with respect to the prediction at hand. If the single ergative marked argument of a transitive verb is extracted without the AF, it is not clear whether AF is impossible or just optionally did not apply.

(37) *Negation in Ixil (Ayres 1981: 130)*:

a. Yeʔl in kat-et-il-in.
NEG 1SG PUNC-2PL.ERG-see-1SG.ABS
'It's not me who you saw.' *negated object*

b. *Yeʔl in in-w-il-ex.
NEG 1SG DUR-1SG.ERG-see-2PL.ABS
'It's not me who sees you.' *negated transitive subject*

c. Yeʔl in kat-ok-in.
NEG 1SG PUNC-enter-1SG.ABS
'It's not me who entered.' *negated intransitive subject*

d. Yeʔl in in-w-ok-eʔ.
NEG 1SG DUR-1SG.ERG-enter-SUF
'It's not me who is entering.' *negated intransitive subject*

The same pattern is found in Chuj. (38) shows that Chuj exhibits the ban on ergative movement with transitive verbs under focus. The focussed constituent is Ā-moved to the preverbal position. It is possible to focus DP_{abs} (38-c), but focussing of DP_{erg} requires the agent focus construction (38-b). In intransitive clauses, DP_{abs} can also be focussed, see (39).

(38) *Focus in Chuj, transitive verb (Davis 2010: ch.22, 37)*:

a. ʔix-Ø-y-ʔil waj Mekel ʔix Katal.
PST-3SG.ABS-3SG.ERG-see CL Michael CL Kathleen
'Kathleen saw Michael.'

b. Ha ʔix Katal ʔix-Ø-ʔil-**an** waj Mekel.
FOC CL Kathleen PST-3SG.ABS-see-**AF** CL Michael
'It is Kathleen who saw Michael.' *focussed transitive subject*

c. Ha waj Mekel ʔix-Ø-y-ʔil ʔix Ketel.
FOC CL Michael PST-3SG.ABS-3SG.ERG-see CL Kathleen
'It is Michael who Kathleen saw.' *focussed object*

(39) *Focus in Chuj, intransitive verb (Buenrostro 2009: 126)*:

a. Ix-Ø-way winh unin.
PST-3SG.ABS-sleep CLASS child
'The child slept.'

b. A jun unin ix-Ø-way-i.
FOC one child PST-3SG.ABS-sleep-ITV
'It was the child who slept.'

In the progressive aspect, the single argument of an intransitive verb triggers the same agreement as DP_{ext} of a transitive verb (ergative agreement); in other aspects it triggers absolutive agreement. Crucially, the ergative marked sole argument of an intransitive verb can be focussed like absolutive marked DPs; it is not necessary (and even impossible) to use the agent focus construction:

(40) *Chuj, focussing of an ergative marked single argument (Buenrostro 2009: 126)*:

a. Wan s-way winh unin.
PROG 3SG.ERG-sleep CLASS child
'The child is sleeping.'

b. A jun unin lanh s-way-i.
FOC one child PROG 3SG.ERG-sleep-ITV
'It is the child who is sleeping.'

Thus, Ixil and Chuj provide evidence that the first prediction of the present co-argument-based analysis is borne out: The single ergative marked argument of an intransitive verb can be Ā-moved.[20] It patterns with absolutive marked DPs in this respect. This shows that the extraction asymmetry in transitive clauses is not triggered by morphological ergative marking alone; rather, it is the presence of a co-argument that causes the ban on ergative movement (which should thus better be called a ban on transitive ergative movement).[21]

---

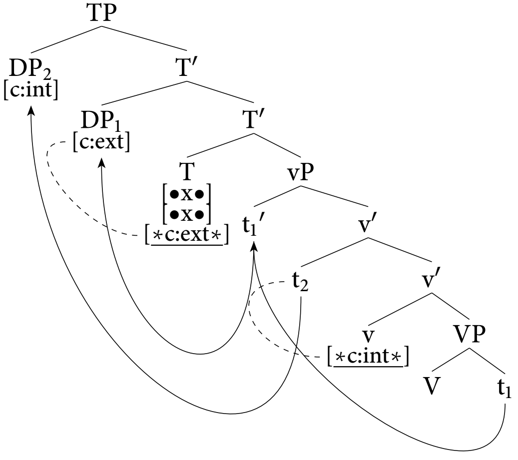
[20] The analyses in Aldridge (2004) and Coon et al. (2011), which are also co-argument based, make the same prediction. Under Stiebels's (2006) account, Ā-movement of the single ergative marked argument is predicted to require agent focus, just as the extraction of transitive agents does. The reason is that AF gives rise to a better constraint profile: It realizes the ergative and signals an Ā-dependency at the same time.

[21] The present account does not exclude the existence of languages in which the sole ergative marked argument of an intransitive verb cannot be extracted: If the split is semantic-based (i.e., each verb falls in exactly one semantic class), verbs of the class that assign ergative to their sole argument must be hidden transitives with a phonologically null DP_{int} (cf. Bobaljik 1993, Laka 1993, Nash 1996, Bittner and Hale 1996). Thus, there is a co-argument which does not get case when DP_{ext} is extracted. However, this analysis is not plausible for aspect-based splits because one and the same verb would have to be transitive and intransitive, depending on aspect.

5.1.2. *Extraction of Both Arguments of a Transitive Verb*

The second prediction of the present account is that in languages with the ban on ergative movement, DP$_{erg}$ *can* be Ā-moved after all if DP$_{abs}$ is extracted as well. The reason is that there is a derivation with Ā-movement of both DP$_{erg}$ and DP$_{abs}$ in which both arguments receive structural case. The initial step is as in (32-a): Given that Merge applies before Agree, v introduces DP$_{ext}$ (external Merge); next, DP$_{int}$ moves to the edge of the phase head v in order to be able to be moved to SpecC (internal Merge). Afterwards, v assigns [c:int] (the ergative) to DP$_{ext}$. As soon as T is merged, it attracts both DPs to its specifier, the edge of the TP phase. Since there is no MLC-like constraint, the order of movements is free. The DP that moves first lands in the inner specifier, and the DP that moves later ends up in the outer specifier of T. Finally, [c:ext] is valued by T via Agree. Since both DPs are in a specifier of T at that point, the Spec-Head Bias does not determine which DP must be the goal of case assignment. Thus, both DPs can be the goal (recall that there is no MLC). The derivation converges if DP$_{int}$ receives [c:ext] from T. In this case, there is no maraudage; see (41) (dotted lines indicate case assignment, continuous lines indicate movement).[22] Finally, both DPs are moved to SpecC.

(41) *Legitimate movement of DP$_{erg}$ and DP$_{abs}$*



---

[22] T could also assign the external case to DP$_{ext}$. In this case, however, DP$_{ext}$ would maraud the case feature that DP$_{int}$ needs and the derivation would crash. But since there is one converging derivation, grammaticality is ensured.

Data from K'ichee' and Kaqchikel confirm this prediction. In section 2.1, we have seen that K'ichee' exhibits the ban on ergative movement with wh-movement and focussing. In (42), both DP$_{erg}$ and DP$_{abs}$ are focussed, and AF is not necessary. Kaqchikel exhibits the ban on ergative movement if DP$_{erg}$ is questioned; cf. (1) and (2). In (43) and (44), DP$_{erg}$ is questioned and DP$_{int}$ is focussed/questioned as well; again, the AF construction is not employed.[23,24]

(42) *Focussing of DP$_{erg}$ and DP$_{abs}$ in K'ichee' (Can Pixabaj and England 2011: 26)*:

are k'u ri al Ixchel, are ri kinaq'
FOC PART DET CL Ixchel FOC DET beans
x-Ø-u-tzak-o.
COMPL-3SG.ABS-3SG.ERG-cook-TV
'... but as for Ixchel, it is beans that she cooked.'

---

[23] As noted in section 2.2, the analyses of Aldridge (2004) and Coon et al. (2011) differ from the present approach in that they predict Ā-movement of more than one DP to be impossible. In the OT account by Stiebels (2006), AF is wrongly predicted to occur with Ā-moved ergative DPs, regardless of whether another DP is extracted or not.

[24] Our account makes a wrong prediction with respect to double Ā-movement of DP$_{erg}$ and DP$_{abs}$ (we thank Erich Groat for pointing this out to us). The unwanted derivation runs as follows: After DP$_{ext}$ and DP$_{int}$ have been (externally and internally) merged to Specv, v assigns [c:int] to DP$_{int}$. T is merged and both DPs move to SpecT. The derivation converges if T assigns [c:ext] to DP$_{ext}$. As a result, an accusative pattern emerges. However, we are not aware of a morphologically ergative language showing morphological accusativity under double extraction. The unwanted derivation may be blocked as follows. First, suppose, following Chomsky (2000), that external Merge precedes internal Merge ("Merge before Move"). In addition, suppose that minimality holds after all, but only between multiple specifiers of a head α. In such a case, α can only enter into Agree with the DP in its innermost specifier. This is sufficient to block the unwanted derivation. In a morphologically ergative language Merge precedes Agree. If external Merge precedes internal Merge, then DP$_{int}$ occupies an outer Specv and DP$_{ext}$ an inner Specv. Consequently, minimality enforces that Agree targets DP$_{ext}$ and thus the accusative pattern is not derived. (For another potential approach that does without minimality, see Georgi 2013.) A similar issue may arise with scrambling: In Mayan, the order of post-verbal arguments is SO or OS with no difference in argument encoding. In principle, an accusative pattern might arise with OS word order: DP$_{int}$, which is scrambled above DP$_{ext}$ to Specv, might get [c:int] from v. Assuming that scrambling is movement, this undesirable result is again excluded if external Merge applies before internal Merge. (If VOS in Mayan comes about via fronting of vP, as Coon 2010*c* proposes for Chol, then there is no problem to begin with.)

(43) *Wh-movement of DP$_{erg}$ and focussing of DP$_{abs}$ in Kaqchikel*:
Achike ja ri jun sik'iwuj n-Ø-u-löq'?
Q.ANIM FOC DET INDEF book INCOMPL-3SG.ABS-3SG.ERG-buy
'Who buys a BOOK?'

(44) *Wh-movement of DP$_{erg}$ and DP$_{abs}$ in Kaqchikel*:
Atux achike n-Ø-u-löq'?
Q Q.ANIM INCOMPL-3SG.ABS-3SG.ERG-buy
'Who buys what?'

5.2. Open Questions

Not all morphologically ergative languages exhibit a ban on ergative movement. In some, DP$_{erg}$ can be freely extracted, e.g., in Chol (Mayan, Coon et al. 2011), Avar (Nakh-Dagestanian, Polinsky et al. 2012), Basque (isolate, Ortiz de Urbina 1989). The question arises as to how language variation with respect to extraction asymmetries can be integrated into the present analysis. The central parts of the analysis of the ban on ergative movement are the assumptions (a) that the order of Merge and Agree on T and v is identical, (b) that DPs that are to be moved to SpecC must make a stop-over in SpecT, and (c) that a DP can check more than one case feature. The extraction asymmetry in ergative languages may not arise if one of these assumptions is changed.

First, the order of Merge and Agree on T might, in principle, differ from the order on v. Merge before Agree on the vP cycle produces morphological ergativity. The same order on T results in the ban on ergative movement. The reverse order on T (Agree before Merge) has the consequence that movement of DP$_{erg}$ comes too late to effect maraudage because T assigned case to DP$_{int}$ earlier. However, this wrongly predicts the possibility of a ban on accusative movement in morphologically accusative languages: If the order in the T domain deviates from the order on v, then Merge before Agree on T may hold in some morphologically accusative languages (which have Agree before Merge on v). If DP$_{acc}$ is to be extracted, it would be merged in SpecT *before* T assigns case and would maraud the external case feature that DP$_{ext}$ needs.

Second, the status of T as a phase head may vary between languages. In some languages, T may not be a phase head and hence not bear edge features. This means that DP$_{erg}$ that is to be Ā-moved to SpecC does not have to go through SpecT. As a consequence, this DP need not maraud the case feature that T provides for DP_{abs} in a Spec-Head-configuration; recall that this was the fatal step in the derivation with illicit movement of DP_{erg}.

The third option to account for the absence of the ban on ergative movement is to assume that a DP cannot check more than one case. This may be so because (a) the number of cases a DP is able to check varies between languages, or (b) because the ergative is not a structural but rather an inherent case in some morphologically ergative languages, (see footnote 6). If (a) holds, DP_{erg}, which has already been assigned internal case by v and which moves to SpecT before T initiates Agree, cannot maraud the case feature of T. Assume that (b) holds: Since only structural case features keep a DP active for further case checking (see (29)), an inherently case marked DP_{ext} that is to be extracted is inactive and hence cannot maraud [c:ext] on T. As a consequence, Agree between T and DP_{int} is not bled, both arguments of a transitive verb receive case. This variant has been worked out in Heck and Müller (2013).

Tada (1993) observes that in languages of the Mayan family that exhibit the ban on ergative movement the absolutive marker appears to the left of the verb stem (high) while in those Mayan languages that lack the ban the absolutive marker appears to the right of the verb (low). Coon et al. (2011) call the first subgroup "HIGH-ABS" languages and the latter "LOW-ABS" languages. They propose that in HIGH-ABS languages absolutive is assigned by T while in LOW-ABS languages it is assigned by v. Due to the PIC, DP_{int} must move to Specv to receive absolutive case in HIGH-ABS languages, but not in LOW-ABS languages. As a consequence, the escape hatch Specv is blocked in HIGH-ABS languages only, which derives the ban on ergative movement and its variation within Mayan. Coon et al.'s (2011) explanation can, in principle, be transposed more or less directly into the present theory. To this end, suppose that the unmarked absolutive in Mayan is either valued by T (HIGH-ABS languages, as in Coon et al. 2011) or by V (LOW-ABS languages). For HIGH-ABS languages everything remains as it was. In LOW-ABS languages, absolutive on DP_{int}, having been valued by V, cannot be marauded by DP_{ext} simply because DP_{ext} is merged to high in the structure (Specv). Ā-extraction of the ergative argument is without consequences. As will become clear shortly, this analysis of LOW-ABS languages is, to a certain extent, similar to the analysis of the agent focus construction in Mayan in the next section.

6. Agent Focus in Mayan

A question that emerges in connection with the ban on ergative movement is how the external argument of a transitive verb can be questioned, relativized or focussed in languages that exhibit the ban on ergative Ā-movement. One possibility in Mayan languages in addition to the detransitivizing antipassive is the *agent focus* construction (AF). In this section we introduce the properties of this construction and we present an analysis of AF within the system developed in section 3.

6.1. Properties of Agent Focus in Mayan Languages

In a regular transitive clause without Ā-movement, both arguments receive structural case. The verb agrees in person and number with both DP$_{erg}$ and DP$_{int}$. The features of DP$_{ext}$ are cross-referenced on the verb by a set of affixes (the ergative affix set) that differs from the set which indicates the features of DP$_{abs}$ (the absolutive affix set). In addition, the verb carries the transitive status suffix (glossed as TV). An intransitive verb carries the intransitive status suffix (glossed as ITV) and the sole argument of the verb also triggers the absolutive agreement set on the verb, see the examples from Q'anjob'al in (45-a) and (45-b).

(45) *Agent focus in Q'anjob'al (Coon 2010a)*:

a. Max-ach y-il-a'.
ASP-2SG.ABS 3SG.ERG-see-TV
'She saw you.' *transitive verb, no extraction*

b. Max-ach way-i.
ASP-2SG.ABS sleep-ITV
'You slept.' *intransitive verb, no extraction*

c. *Maktxel max-ach s-laq'-a'?
who ASP-2SG.ABS 3SG.ERG-hug-TV
'Who hugged you?' *extraction of transitive agent without AF*

d. Maktxel max-ach laq'-on-i?
who ASP-2SG.ABS hug-AF-ITV
'Who hugged you?' *extraction of transitive agent with AF*

Ā-movement of DP$_{erg}$ is ungrammatical in Q'anjob'al (see (45-c)). The agent focus construction can be used instead to express the same content (see

(45-d)). In AF, both arguments receive structural case, just as in a regular transitive clause without extraction. None of the arguments is realized as an oblique; there is no demotion of arguments. Hence, AF is not a detransitivizing operation (in support of this view see the references in Aissen 1999). However, the verb can agree with only one of the two arguments of a transitive verb and cross-references this argument by the absolutive set of affixes.[25] In addition, the verb carries the intransitive status suffix. Furthermore, an additional suffix attaches to the verb, glossed as AF, see (45-d). To summarize, the AF construction is syntactically transitive, but morphologically intransitive: Apart from the AF-morpheme the verbal morphology looks like the one we find on intransitive verbs, but there are two core arguments.

Moreover, there are restrictions on the use of AF: It can only be used if a transitive agent is to be extracted (but see footnote 26); it cannot be used in a regular transitive clause without extraction (see (46-a)) or if a non-agent DP in a transitive clause is extracted (see (46-b)).

(46) *AF Restrictions in Tzotzil Aissen (1999, 455)*:
a. *I-kolta-on tzeb li Xun-e.
COMPL-help-AF girl the Juan-ENC
'Juan helped the girl.' *no extraction*
b. ??A li Xun-e, I-kolta-o li tzeb-e.
FOC the Juan-ENC, COMPL-help-AF the girl-ENC
'The girl helped JUAN.' *focussing of* DP_{int}

We thus need to account for the following properties of AF: (a) intransitive agreement morphology, (b) structural case assignment to both DPs, (c) obligatory extraction of DP_{ext}, and (d) impossibility of extracting DP_{int}.

6.2. Analysis of the Agent Focus Construction

Under the present analysis, the problem with Ā-movement of DP_{erg} is that its co-argument, the internal argument of a transitive verb, does not receive case. Following Ordóñez (1995) and drawing heavily from Coon et al. (2011),

---

[25] The choice of the agreement-triggering argument is regulated by language-specific rules: In some Mayan languages only the object triggers agreement, in others only the subject, and in a third group Silverstein hierarchies determine which argument agrees with the verb (see Stiebels 2006 for an overview). This choice does not have an impact on the analysis of AF that we will present in this section.

let us assume that in the AF construction DP$_{int}$ is assigned structural case by an added probe, represented as [*c:x*] (cf. Béjar and Řezáč 2009). This probe is morphologically realized by the AF-morpheme. Since the AF morpheme is always adjacent to the verbal root, we can conclude that the added probe is located very low in the structure, on V.[26] In addition, an intransitive v is merged that does not assign [c:int] (ergative case), but still introduces the external argument (this variant of v is independently needed to account for case assignment with unergative verbs: It introduces an external argument but does not assign ergative case to it). All other assumptions we made so far stay the same. In particular, the feature content of T does not change, it still assigns [c:ext] and triggers intermediate movement steps via edge features.

The assumption that an intransitive v is merged accounts for the intransitive morphology in the AF construction: Only a single argument is cross-referenced on the verb (via Agree with T), because v does not have a probe and hence cannot trigger Agree. The extractability of DP$_{ext}$ and the ban on extraction of DP$_{int}$ as well as the assignment of structural case to both DPs follow automatically from the assumptions in section 3.[27] We start with the operations in the vP; these are the same, regardless of whether DP$_{erg}$ or DP$_{abs}$ is to be extracted, see (47). First, the added probe on V enters into Agree with DP$_{int}$, which is the only available goal at that point of the derivation because V does not introduce a DP in its specifier. Afterwards, v is merged and it introduces DP$_{ext}$. Being an intransitive variant, v does *not* trigger Agree; hence, the order of operations does not play any role on the vP cycle. DP$_{ext}$ does not receive case from v, it therefore still needs a structural case value.

(47) *Operations applying in the vP*:
[$_{\mathrm{vP}}$ DP$_{ext}$ { [c:□]} [$_{\mathrm{v'}}$ v{[•D•]} [$_{\mathrm{VP}}$ V{ [*c:x*]} DP$_{int}$ { [c:x]}]]]
(ii) Merge
(i) Agree

Suppose that DP$_{ext}$ of a transitive verb is extracted, see (48). T has a case probe and an edge feature that triggers the intermediate movement step to SpecT. Given the order Merge before Agree in a morphologically ergative language, DP$_{ext}$ moves to SpecT. Due to the Spec-Head Bias, T assigns the external case

---

[26] Since the analysis of AF is borrowed from Coon et al. (2011), it accounts in the same way for their observation that embedded transitive clauses in Q'anjob'al exhibit the AF morpheme, too; see Coon et al. (2011) for details.

[27] Other accounts of the AF construction have been put forward by Larsen (1988), Tada (1993), Coon et al. (2011).

to DP$_{ext}$. But in contrast to the derivation without AF (cf. (31)), DP$_{ext}$ is in need of case from T because it did not receive a case value within vP. Since DP$_{int}$ gets case early in the derivation from V and does not depend on the case assigned by T (as it does in regular transitives), the derivation converges. Both DPs get structural case. DP$_{ext}$ can be moved to SpecC.

(48) [$_{\text{TP}}$ DP$_{ext}$ {[c:ext]} [$_{\text{T}'}$ T{[*c:ext*]} [$_{\text{vP}}$ t$_{DP_{ext}}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ V DP$_{int}$ {[c:x]}]]]]]

(iii) Merge

(iv) Agree

If DP$_{int}$ is Ā-moved, the derivation continues on the basis of (47) as follows: Given the order Merge before Agree, DP$_{int}$ is moved to SpecT before T assigns case. Due to the Specifier-Head Bias, DP$_{int}$ checks [c:ext] on T in addition to the case [c:x] it checked with the added probe on V. There is no case left which could be assigned to DP$_{ext}$. DP$_{int}$ marauds the case that DP$_{ext}$ needs; see (49). The derivation crashes. This is exactly the reverse pattern of what we saw in the derivation of the ban on ergative movement (cf. (31)): In AF, the Ā-moved DP$_{int}$ marauds the case that DP$_{ext}$ would need; in regular transitives, the Ā-moved DP$_{ext}$ marauds the case for DP$_{int}$.[28]

(49) [$_{\text{TP}}$ DP$_{int}$ {[c:ext]} [$_{\text{T}'}$ T{$\underline{\text{[*c:ext*]}}$} [$_{\text{vP}}$ DP$_{ext}$ {[c:□]G} [$_{\text{v}'}$ v [$_{\text{VP}}$ V t$_{DP_{int}}$]]]]]

(iii) Merge

(iv) Agree

To sum up, the analysis accounts for the fact that the external argument of a transitive verb can be Ā-moved under AF, whereas the internal argument cannot be extracted. The pattern is the reverse of what we find with extraction of DP$_{erg}$. However, one open question remains: Why can AF only be applied if an element is extracted? Under the present account, there is an AF derivation that converges if no DP is extracted: DP$_{int}$ gets case from the added probe on V and DP$_{ext}$ receives [c:ext] from T in its base position in Specv. [29]

---

[28] Coon et al. (2011) do not provide an explanation for this restriction on AF.

[29] One could pursue the idea that AF is a repair strategy that steps in only if the derivation without AF crashes. We will not pursue the issue any further here. As far as we can tell, no explanation is provided by Coon et al. (2011) either.

7. Conclusion

In this paper we have presented a relational, co-argument based account of the ban on ergative movement that holds in many morphologically ergative languages. We have argued that the extraction asymmetry cannot be brought about by restrictions on movement of the ergative DP if all constraints are principles of efficient computation or imposed by the interfaces and if traces do not exist as items that constraints can refer to, as is assumed in recent developments of the minimalist program. We have proposed that movement of the ergative is per se unproblematic, but if it applies, it creates problems for the absolutive co-argument of the ergative. The internal argument cannot get absolutive case because the ergative, by its very nature, moves early and marauds the case feature for the internal argument. No such movement asymmetry arises in morphologically accusative languages because movement of a DP applies late, after the co-argument already received its case feature. Hence, maraudage cannot take place. The different timing of operations in ergative vs. accusative languages is derived from the analysis of morphological ergativity and accusativity: The order Merge before Agree holds in ergative languages, whereas Agree before Merge holds in accusative languages on v and T. The analysis implies a strictly derivational syntax in which the order of operations plays an important role in deriving properties of the grammar.

Moreover, the varying order of Merge and Agree leads to opacity effects: In ergative languages, movement of DP_{erg} bleeds Agree between T and DP_{abs}, with fatal consequences; in accusative languages, movement of DP_{acc} counterbleeds Agree between T and DP_{nom}. Furthermore, the approach predicts that no ban on ergative movement arises (a) if DP_{abs} is extracted as well and (b) if the sole ergative marked argument of an intransitive verb is extracted. These predictions have been shown to be borne out empirically. Finally, we have suggested that the AF construction, a repair strategy used for extraction of DP_{erg} in Mayan languages, is another phenomenon in which the timing of operations plays an important role: Movement of DP_{int} bleeds Agree between T and DP_{ext}, the reverse of what we find with the extraction of DP_{erg} in a regular transitive clause. In sum, the present account provides an argument for the privileged status of specifiers in syntactic derivations (DPs in specifiers maraud features of a head); and it emphasizes the role of timing in grammar and thereby argues for a strictly derivational syntax.

References

Abels, Klaus (2012): *Phases: An Essay on Cyclicity in Syntax*. Mouton de Gruyter, Berlin.

Adger, David (2003): *Core Syntax*. Oxford University Press, Oxford, New York.

Adger, David and Daniel Harbour (2007): 'Syntax and Syncretisms of the Person Case Constraint', *Syntax* 10, 2–37.

Aissen, Judith (1999): 'Agent Focus and Inverse in Tzotzil', *Language* 75(3), 451–485.

Aldridge, Edith (2004): Ergativity and Word Order in Austronesian Languages. PhD thesis, Cornell University, Ithaca, New York.

Alexiadou, Artemis (2001): *Functional Structure in Nominals*. Benjamins, Amsterdam.

Anagnostopoulou, Elena (2003): *The Syntax of Ditransitives: Evidence from Clitics*. Mouton de Gruyter, Berlin.

Andrews, Avery (1996): 'Semantic Case-Stacking and Inside-Out Unification', *Australian Journal of Linguistics* 16, 1–54.

Aoun, Joseph (1986): *Generalized Binding*. Foris, Dordrecht.

Arregi, Karlos and Andrew Nevins (2012): *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Springer.

Assmann, Anke, Svetlana Edygarova, Doreen Georgi, Timo Klein and Philipp Weisser (2013): Possessor Case in Udmurt: Multiple Case Assignment Feeds Postsyntactic Fusion. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 23-63.

Ayres, Glenn (1981): 'On Ergativity and Aspect in Ixil', *Journal of Mayan Linguistics* 2(2), 128–145.

Barss, Andrew (1986): Chains and Anaphoric Dependence. Ph.d. thesis, MIT, Cambridge, Mass.

Béjar, Susana and Milan Řezáč (2009): 'Cyclic Agree', *Linguistic Inquiry* 40, 35–73.

Bittner, Maria and Ken Hale (1996): 'The Structural Determination of Case and Agreement', *Linguistic Inquiry* pp. 1–68.

Bobaljik, Jonathan (1993): Ergativity and Ergative Unergatives. *In:* C. Phillips, ed., *Papers on Case and Agreement II*. Vol. 19 of *MIT Working Papers in Linguistics*, MITWPL, MIT: Cambridge, Mass., pp. 45–88.

Bobaljik, Jonathan and Phil Branigan (2006): Eccentric Agreement and Multiple Case Checking. *In:* A. Johns, D. Massam and J. Ndayiragije, eds, *Ergativity*. Springer, pp. 47–77.

Boeckx, Cedric (2003): *Islands and Chains. Resumption as Stranding*. Benjamins, Amsterdam.

Boeckx, Cedric and Kleanthes K. Grohmann (2007): 'Putting Phases in Perspective', *Syntax* 10, 204–222.

Bohnemeyer, Jürgen (2004): 'Split Intransitivity, Linking, and Lexical Representation: The Case of Yukatek Maya', *Linguistics* 42(1), 67–107.

Bošković, Željko (2007): 'Agree, Phases, and Intervention Effects', *Linguistic Analysis* 33, 54–96.

Bresnan, Joan (1978): Contraction and the Transformational Cycle in English. Indiana University Linguistics Club.

Bricker, Victoria R. (1981): 'The Source of the Ergative Split in Yucatec Maya', *Journal of Mayan Linguistics* 2, 83–127.

Brody, Michael (2002): On the Status of Representations and Derivations. *In:* S. D. Epstein and T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 19–41.

Buenrostro, Cristina (2009): *Chuj de San Mateo Ixtatan*. El Colegio de México, Mexico.

Campana, Mark (1992): A Movement Theory of Ergativity. PhD thesis, McGill University, Montréal.

Can Pixabaj, Telma and Nora C. England (2011): Nominal Topic and Focus in K'ichee. *In:* R. Gutíérrez-Bravo, L. Mikkelsen and E. Potsdam, eds, *Representing Language. Essays in Honour of Judith Aissen*. Booksurge Publishing, pp. 15–30. Available at http://escholarship.org/uc/lrc_aissen.

Chomsky, Noam (1951): Morphophonemics of Modern Hebrew. Master's thesis, University of Pennsylvania.

Chomsky, Noam (1975): *The Logical Structure of Linguistic Theory*. Plenum Press, New York.

Chomsky, Noam (1986): *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and J. Uriagereka, eds, *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. *In:* M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (2005): 'Three Factors in Language Design', *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam (2007): Approaching UG from Below. *In:* U. Sauerland and H.-M. Gärtner, eds, *Interfaces + Recursion = Language?*. Mouton de Gruyter, Berlin, pp. 1–31.

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

Chung, Sandra (1998): *The Design of Agreement: Evidence from Chamorro*. Chicago University Press, Chicago.

Comrie, Bernard (1988): Linguistic Typology. *In:* F. J. Newmeyer, ed., *Linguistics: The Cambridge Survey*. Vol. I, Foundations, Cambridge University Press, Cambridge, pp. 447–461.

Coon, Jessica (2010*a*): A-Bar Extraction Asymmetries. Handout, talk given at the Workshop on the Fine Structure of Grammatical Relations, Leipzig University, December 2010.

Coon, Jessica (2010*b*): 'Rethinking Split Ergativity In Chol', *International Journal of American Linguistics* **76**(2), 207–253.

Coon, Jessica (2010*c*): 'VOS as Predicate Fronting in Chol', *Lingua* pp. 354–378.

Coon, Jessica and Omer Preminger (2012): Taking 'Ergativity' out of Split Ergativity: A Structural Account of Aspect and Person Splits. Ms., McGill and MIT.

Coon, Jessica, Omer Preminger and Pedro Mateo Pedro (2011): The Role of Case in A-Bar Extraction Asymmetries – Evidence from Mayan. Ms., McGill, Harvard, and MIT.

Craig, Colette (1977): *The Structure of Jacaltec*. University of Texas Press, Austin, Texas.

Davis, Philip W. (2010): Syntax and Semantics. Book draft, Rice University.

Dixon, R.M.W. (1994): *Ergativity*. Cambridge University Press, Cambridge.

England, Nora (1983*a*): *A Grammar of Mam, a Mayan Language*. University of Texas Press, Austin, Texas.

England, Nora (1983*b*): 'Ergativity in Mamean (Mayan) Languages', *International Journal of American Linguistics* **49**(1), 1–19.

England, Nora (1989): 'Comparing Mam (Mayan) Clause Structures; Subordinate vs. Main Clauses', *International Journal of American Linguistics* **55**(3), 283–308.

Epstein, Samuel David and T. Daniel Seely (2002): Rule Applications as Cycles in a Level-Free Syntax. *In:* S. D. Epstein and T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65–89.

Erlewine, Michael Yoshitaka (2012): Dissociating the Syntax and Morphological Realization of Kaqchikel Agent Focus. Ms., MIT.

Furbee-Losee, Louanna (1976): *The Correct Language: Tojolabal*. Garland, New York.

Georgi, Doreen (2012): A Local Derivation of Global Case Splits. *In:* A. Alexiadou, T. Kiss and G. Müller, eds, *Local Modelling of Non-Local Dependencies in Syntax*. Mouton de Gruyter, Berlin, pp. 305–336.

Georgi, Doreen (2013): Opaque Interaction of Internal Merge and Agree. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 413–461.

Georgi, Doreen, Fabian Heck and Gereon Müller (2009): Maraudage. Ms., Universität Leipzig.

Heck, Fabian and Gereon Müller (2007): Extremely Local Optimization. *In:* E. Brainbridge and B. Agbayani, eds, *Proceedings of the 26th WECOL*. California State University, Fresno, pp. 170–183.

Heck, Fabian and Gereon Müller (2013): On Accelerating and Decelerating Movement: A Case Study in Harmonic Serialism. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 511-558.

Heck, Fabian and Marc Richards (2010): 'A Probe-Goal Approach to Agreement and Non-Incorporation Restrictions in Southern Tiwa', *Natural Language and Linguistic Theory* 28, 681–721.

Henderson, Robert (2007): Observations on the Syntax of Adjunct Extraction in Kaqchikel. In: *Proceedings of the Conference on Indigenous Languages of Latin America III.*.

Jelinek, Eloise (1993): Ergative 'Splits' and Argument Type. *In:* J. Bobaljik and C. Phillips, eds, *Papers on Case and Agreement I*. Vol. 18 of *MIT Working Papers in Linguistics*, MITWPL, MIT: Cambridge, Mass., pp. 15–42.

Kayne, Richard (1989): Facets of Participle Agreement in Romance. *In:* P. Benincà, ed., *Dialect Variation and the Theory of Grammar*. Foris, Dordrecht, pp. 85–104.

Keenan, Edward L. (1976): Remarkable Subjects in Malagasy. *In:* C. Li, ed., *Subject and Topic*. Academic Press, New York, pp. 247–301.

Kiparsky, Paul (1973): 'Elsewhere' in Phonology. *In:* S. Anderson and P. Kiparsky, eds, *A Festschrift for Morris Halle*. Academic Press, New York, pp. 93–106.

Kiparsky, Paul (1976): Abstractness, Opacity and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton, The Hague.

Koopman, Hilda (1992): 'On the Absence of Case Chains in Bambara', *Natural Lanuage and Linguistic Theory* 10, 555–594.

Koopman, Hilda (2006): Agreement Configurations. *In:* C. Boeckx, ed., *Agreement Systems*. Benjamins, Amsterdam, pp. 159–201.

Kratzer, Angelika (1996): Severing the External Argument from its Verb. *In:* J. Rooryck and L. Zaring, eds, *Phrase Structure and the Lexicon*. Kluwer, Dordrecht, pp. 109–137.

Lahne, Antje (2008): Excluding SVO in Ergative Languages. *In:* F. Heck, G. Müller and J. Trommer, eds, *Varieties of Competition*. Vol. 87 of *Linguistische Arbeitsberichte*, Universität Leipzig, pp. 65–80.

Laka, Itziar (1993): Unergatives that Assign Ergative, Unaccusatives that Assign Accusative. *In:* J. Bobaljik and C. Phillips, eds, *Papers on Case and Agreement I*. Vol. 18 of *MIT Working Papers in Linguistics*, MITWPL, MIT: Cambridge, Mass., pp. 149–172.

Larsen, Thomas W. (1988): Manifestations of Ergativity in Quiché Grammar. PhD thesis, University of California, Berkeley.

Larsen, Tomas W. and William M. Norman (1979): Correlates of Ergativity in Mayan Grammar. *In:* F. Plank, ed., *Ergativity: Towards a Theory of Grammatical Relations*. Academic Press, London and New York, pp. 347–370.

Lasnik, Howard (1999): 'On Feature Strength: Three Minimalist Approaches to Overt Movement', *Linguistic Inquiry* 30, 197–217.

Lechner, Winfried (2010): Criteria for Diagnosing Movement. (And Some Remarks on the Duke of York). Ms., University of Athens.

Legate, Julie Anne (2008): 'Morphological and Abstract Case', *Linguistic Inquiry* 39(1), 55–101.

Lengyel, Thomas E. (1978): Ergativity, Aspect and Related Perplexities of Ixil-Maya. *In:* N. England, ed., *Papers in Mayan Linguistics*. University of Missouri, Columbia, pp. 78–91.

Levin, Juliette and Diane Massam (1985): Surface Ergativity: Case/Theta Relations Reexamined. *In:* S. Berman, ed., *Proceedings of NELS 15*. BLS, Amherst, Mass.

Marantz, Alec (1991): Case and Licensing. *In:* German Westphal, B. Ao and H.-R. Chae, eds, *Proceedings of the Eighth Eastern States Conference on Linguistics*. University of Maryland, pp. 234–253.

McFadden, Thomas (2004): The Position of Morphological Case in the Derivation: A Study on the Syntax-Morphology Interface. PhD thesis, University of Pennsylvania.

Merchant, Jason (2006): Polyvalent Case, Geometric Hierarchies, and Split Ergativity. *In:* J. Bunting, S. Desai, R. Peachey, C. Straughn and Z. Tomkova, eds, *Proceedings of the 42nd Annual Meeting of the Chicago Linguistics Society*. Chicago Linguistics Society, Chicago, pp. 47–67.

Müller, Gereon (2004): Phrase Impenetrability and Wh-Intervention. *In:* A. Stepanov, G. Fanselow and R. Vogel, eds, *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 289–325.

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. *In:* K. K. Grohmann, ed., *Explorations of Phase Theory. Features and Arguments*. Mouton de Gruyter, Berlin, pp. 269–308.

Müller, Gereon (2011): *Constraints on Displacement*. Benjamins, Amsterdam.

Murasugi, Kumiko (1992): Crossing and Nested Paths. PhD thesis, MIT, Cambridge, Mass.

Nash, Léa (1996): The Internal Ergative Subject Hypothesis. *In:* K. Kusumoto, ed., *Proceedings of NELS 26*. GLSA, Amherst, Mass, pp. 195–209.

Nordlinger, Rachel (1998): *Constructive Case: Evidence from Australian Languages*. CSLI, Stanford.

Ordóñez, Francisco (1995): 'The Antipassive in Jacaltec: A Last Resort Strategy', *CatWPL* 4, 329ïTh¡343.

Ortiz de Urbina, Jon (1989): *Parameters in the Grammar of Basque: A GB Approach to Basque Syntax*. Foris, Dordrecht.

Otsuka, Yuko (2000): Ergativity in Tongan. PhD thesis, Oxford University.

Otsuka, Yuko (2006): Syntactic Ergativity in Tongan: Resumptive Pronouns Revisited. *In:* A. Johns, D. Massam and J. Ndayiragije, eds, *Ergativity: Emerging issues*. Springer, Dordrecht, pp. 79–107.

Paul, Ileana and Lisa Travis (2006): *Ergativity in Austronesian Languages*. Springer, Dordrecht.

Pesetsky, David (1989): Language-Particular Processes and the Earliness Principle. Ms., MIT, Cambridge, Mass.

Pesetsky, David and Esther Torrego (2007): The Syntax of Valuation and the Interpretability of Features. *In:* S. Karimi, V. Samiian and W. K. Wilkins, eds, *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation*. Benjamins, Amsterdam, pp. 262–294.

Polinsky, Maria, Carlos Gómez Gallo, Peter Graff and Ekaterina Kravtchenko (2012): 'Subject Preference and Ergativity', *Lingua* **122**(3), 267–277.

Queixalos, Francesc (2010): Grammatical Relations in Katukina-Kanamari. *In:* S. Gildea and F. Quiexalos, eds, *Ergativity in Amazonia*. John Benjamins, pp. 235–283.

Řezač, Milan (2003): 'The Fine Structure of Cyclic Agree', *Syntax* **6**, 156–182.

Richards, Marc (2011): 'Deriving the Edge: What's in a Phase?', *Syntax* **14**, 74–95.

Richards, Norvin (2013): 'Lardil "Case Stacking" and the Timing of Case Assignment', *Syntax* **16**(1), 42–76.

Rouveret, Alain and Jean-Roger Vergnaud (1980): 'Specifying Reference to the Subject', *Linguistic Inquiry* **11**, 97–202.

Smith-Stark, Thom (1976): Ergativity, Grammatical Relations, Accessibility Hierarchies, Pocomam, and Cosmic Consciousness. Ms.

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

Stiebels, Barbara (2000): Linker Inventories, Linking Splits and Lexical Economy. *In:* B. Stiebels and D. Wunderlich, eds, *Lexicon in Focus*. Akademie-Verlag, Berlin, pp. 211–245.

Stiebels, Barbara (2006): 'Agent Focus in Mayan Languages', *Natural Language & Linguistic Theory* **24**, 501–570.

Tada, Hiroaki (1993): A/A-bar Partition in Derivation. PhD thesis, MIT, Cambridge, Mass.

Takahashi, Daiko (1994): Minimality of Movement. PhD thesis, University of Connecticut.

Trommer, Jochen (2011): Phonological Aspects of Western Nilotic Mutation Morphology. Habilitationsschrift, Universität Leipzig.

Ura, Hiroyuki (2000): *Checking Theory and Grammatical Functions in Universal Grammar*. Oxford University Press, New York.

Woolford, Ellen (2001): Case Patterns. *In:* G. Legendre, J. Grimshaw and S. Vikner, eds, *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 509–543.

Woolford, Ellen (2006): Case-Agreement Mismatches. *In:* C. Boeckx, ed., *Agreement Systems*. Benjamins, Amsterdam, pp. 317–339.

Wunderlich, Dieter (1997): 'Cause and the Structure of Verbs', *Linguistic Inquiry* 27, 27–68.

Zimmermann, Eva (2013): 'Vowel Deletion as Mora Usurpation: The Case of Yine', *Phonology* 30, 125–163.

Opaque Interaction of Internal Merge and Agree

Doreen Georgi[*]

Abstract

In this paper I present an argument for a strictly derivational model of syntax based on timing of operations. Empirical evidence comes from opaque interactions of elementary operations which show that internal Merge (IM) must be split into two types: IM to intermediate landing sites and IM to final landing sites. The split is motivated by the following observation: When both types of IM are triggered by the same head H, they apply at different points in the derivation. This becomes visible once they interact with Agree: In some languages, IM to final landing sites feeds/bleeds Agree initiated by H, whereas IM to intermediate landing sites has the opposite effect, i.e., it counter-feeds/counter-bleeds Agree. This effect can be derived by ordering operation-inducing features on H: One type of IM applies before and the other after Agree. The general implication is that Agree not only needs to be ordered with respect to Merge; a more fine-grained approach is needed that distinguishes between different types of (internal) Merge. Furthermore, reordering of the operation-inducing features on H predicts a certain range of cross-linguistic variation. Based on the attested variation, I argue for the need of extrinsic ordering.

1. Introduction

In this paper I present an argument for a strictly derivational model of syntax based on the timing of elementary syntactic operations. The evidence comes from opacity effects that show that internal Merge (IM) is not a uniform operation; rather, IM must be split into two types: (i) IM to intermediate landing

---

[*] For valuable comments on earlier versions of this paper I would like to thank the participants of the syntax colloquium at the University of Leipzig as well as the audiences at ConSOLE XXI (Potsdam, January 2013) and at the workshop Interaction of Syntactic Primitives (35th Annual Conference of the German Linguistic Society, March 2013). I profited enormously from the discussions on rule ordering with Anke Assmann, Fabian Heck, Gereon Müller, Martin Salzmann and Philipp Weisser. This research is supported by the DFG (Deutsche Forschungsgemeinschaft) within the project "Lokale Modellierung nicht-lokaler Abhängigkeiten in der Syntax" (Local modeling of non-local dependencies in syntax): http://www.uni-leipzig.de/~lomo.

Rule Interaction in Grammar, 413–461
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

sites, triggered by edge features, and (ii) IM to final landing sites in a movement chain, triggered by other features such as a wh-feature on C, the EPP on T, etc. This split is empirically motivated by the observation that when both types of IM are triggered by the same head H, they apply at different points in the derivation. This effect becomes visible once they interact with Agree: In some languages, IM to final landing sites feeds/bleeds Agree relations initiated by H, whereas IM to intermediate movement sites has exactly the opposite effect, i.e., it counter-feeds/counter-bleeds Agree. In the latter cases, the interaction of IM and Agree is opaque. The term opacity characterizes rule interactions that are non-transparent: When looking at the output of an opaque interaction, it is unclear (a) why a certain operation has not applied although its context is given (counter-feeding) or (b) why an operation has applied although its context is not given (counter-bleeding), cf. Kiparsky (1976). The cases at hand are opaque because internally merged XPs land in the same position SpecH (the Spec of the head H that triggered IM) whether IM is driven by edge features (intermediate IM) or by other features (final IM). The output structure is identical; nevertheless, the two types of IM have different consequences for Agree initiated by H, which is reflected in the presence or absence of a morphological marker. The effect can be modeled by ordering operation-inducing features on the triggering head H: edge feature-driven IM applies *after* Agree and non-edge feature-driven IM applies *before* Agree initiated by H. The consequence of this order is that the former type of IM applies *too late* to change possible Agree relations (the DP that is to be internally merged is still in its base position when Agree applies); the latter type of IM changes structural relations *before* Agree applies and can thus feed or bleed Agree relations (depending on the input), because Agree is structure-sensitive. This analysis of opacity effects crucially relies on timing of elementary operations and thereby provides an argument for a strictly derivational syntax (cf. also Řezač 2004, Heck and Müller 2007).

I will show that opacity effects with this abstract pattern can be found with a number of functional heads in the clause on the basis of the phenomena that follow: the anti-agreement effect in Berber, defective intervention in Romance and Icelandic, and possessor case/agreement in Hungarian. The analysis presupposes that intermediate movement steps are triggered by designated features (edge features) that are different from the triggers of final movement steps (contra Abels 2012).

Furthermore, I will argue on the basis of cross-linguistic variation that not all of the attested orderings of operation-inducing features on a head H are predicted by principles of the grammar (contra Pullum 1979); in particular, the Cyclic Principle (even in its strongest form as formulated in McCawley 1984, 1988) has nothing to say about orderings of operations that are triggered by the same head. Extrinsic ordering (or: parochial ordering in Pullum's 1979 terms) is thus necessary after all to account for the variation. Interestingly, not all logically possible orderings of the triggers of the two IM types and Agree seem to be attested for the phenomena examined in this paper. I suggest an account of this asymmetry which is based on specificity-driven ordering of operation-inducing features on a head. In addition, the variation found with defective intervention effects provides evidence that wh-movement uses SpecT as an intermediate landing site (for a recent defense of the opposite view see Abels 2003, 2012 and references cited there).

The paper is structured as follows: Section 2 gives a short overview of the types of rule interactions and the operations that interact in minimalist syntax. Section 3 introduces the theoretical assumptions and illustrates the abstract patterns of interactions that will be encountered in the data. Furthermore, it is shown how the opacity is resolved under the given assumptions. Section 4 presents concrete instantiations of the abstract patterns on the basis of several phenomena and presents detailed derivations. Section 5 discusses consequences of the analysis and examines whether its predictions with respect to cross-linguistic variation are borne out. Furthermore, the implications for the extrinsic/intrinsic dichotomy are discussed. Finally, section 6 concludes.

2. Rule Interaction in Grammar

This section introduces the four basic types of rule interactions: feeding, bleeding, counter-feeding and counter-bleeding. Furthermore, I present the syntactic primitives that can interact in Minimalism.

2.1. Types of Rule Interactions

In early transformational grammar, the grammar consists of two core components: the base component which generates the underlying structure (deep structure) of a linguistic expression, and a transformational component which relates the underlying structure to the surface structure of a linguistic expression (cf. Chomsky 1957, 1965 on syntax and Chomsky and Halle 1968 for application of this model to phonology). The transformational component consists of a number of rules that apply to the underlying structure and map it onto another structure. Often, more than one rule applies to derive the surface structure of a linguistic expression from its deep structure. It has been observed that in this case, the rules may interact in intricate ways: The application of a rule R_1 may facilitate or block the application of another rule R_2. The former case is an instance of feeding, the latter is an instance of bleeding. Kiparsky (1968, 1971, 1976) divided rule interactions into two types: transparent and opaque interactions. Transparent interactions comprise feeding and bleeding; opaque interactions comprise counter-feeding and counter-bleeding. These terms are defined as follows (X '>' Y means that X applies before Y):

(1) *Transparent interactions*:
- a. Feeding:
  - (i) A rule R_1 creates the context for the application of a rule R_2.
  - (ii) *Example* R_1: A → B > R_2: B → C
- b. Bleeding:
  - (i) A rule R_1 destroys the context for the application of a rule R_2.
  - (ii) *Example* R_1: A → B > R_2: A → C

(2) *Opaque interactions*:
- a. Counter-feeding:
  - (i) A rule R_1 creates the context for the application of a rule R_2 and should thus feed R_2.
  - (ii) However, empirical evidence shows that R_2 has not applied although R_1 has.
  - (iii) A rule has not applied although its context is given.
  - (iv) *Example* R_2: B → C > R_1: A → B
- b. Counter-bleeding:
  - (i) A rule R_1 destroys the context for the application of a rule R_2 and should thus bleed R_2.
  - (ii) However, empirical evidence shows that R_2 has applied although R_1 has as well.
  - (iii) A rule has applied although its context is not given.
  - (iv) *Example* R_2: A → B > R_1: A → C

The opaque interactions can be characterized as follows: It is expected that a rule R_1 facilitates (feeds) or blocks (bleeds) the application of another rule R_2, but this is not the case. On the surface, (i) rule R_2 has not applied although its context is created by the prior application of R_1 (counter-feeding), or (ii) R_2 has applied although its context should have been destroyed by the prior application of R_1 (counter-bleeding). Opacity can be described by reversing the order of operations: If R_1 applies before R_2 the former should feed or bleed the latter. But since this is not the case, R_2 must have applied before R_1; R_1 applied too late to facilitate or block the application of R_2.

Rule interactions have been a major topic in phonology and syntax since the earliest days of generative grammar. Opacity was first described by Chomsky (1951) for Hebrew phonology. For an overview of interactions of phonological rules see e.g. Chomsky and Halle (1968), Anderson (1969, 1974), Koutsoudas et al. (1974), Kenstowicz and Kisseberth (1977, 1979), Baković (2011); on rule interaction in morphology see Embick (2010), Arregi and Nevins (2012). As for syntax, the most comprehensive treatment of this topic is Pullum (1979); see also Ross (1967), Williams (1974), Kayne (1975), Perlmutter and Soames (1979), McCawley (1984, 1988), Řezač (2004), Lasnik (2001) and Brody (2002).

The topic of this paper are opaque interactions in syntax. A well-known example for counter-bleeding in syntax that has been recurring in the transformational literature is the interaction of Reflexivization and Imperative Subject Deletion (ISD) in English (cf. McCawley 1988: ch.6 and Pullum 1979: ch.1).

(3) *Informal definitions of transformations*:
  a. Reflexivization: If the direct object of a transitive verb is coreferent with the local subject, it must be realized as a reflexive pronoun.
  b. Imperative subject deletion: A deep structure 2nd person subject is deleted when it is the subject of an imperative.

In imperatives, the direct object is reflexivized if it is 2nd person and thus coreferent with the (phonologically empty) subject:

(4) *Reflexivization in imperatives in English*
  a. Defend yourself!
  b. *Defend you!

On the surface, it is unclear why Reflexivization could apply in (4-a) given that its context is not given because there is no coreferent subject. Since Reflexivization makes reference to the subject of a transitive verb, deletion of the subject by ISD could bleed Reflexivization, but it does not (counter-bleeding). This implies that Reflexivization applies before ISD, at a point in the derivation where the subject is still present; it is deleted only afterwards.

2.2. Rule Ordering in Minimalism: Conflicts in the Derivation

Minimalism is a derivational approach to syntax (cf. Chomsky 1995 et seq.). Syntactic structures unfold incrementally in a bottom-up fashion by successive applications of the elementary operations Merge and Agree (Chomsky 2000, 2001). These are defined as follows:

(5) *Merge and Agree*:

a. Merge (external and internal) is a structure-building operation. Merge is triggered by structure-building features [•F•].

b. Agree relates functional heads and arguments. It is (among other things) responsible for argument encoding: (i) ϕ-features are copied from DPs onto functional heads and (ii) case values are assigned by functional heads to DPs.
Agree applies under c-command and is triggered by probe features [*F*].

It is these basic operations that can interact in Minimalism. In particular, the structure-sensitive Agree operation can be fed or bled by Merge which builds structures and which thus may change structural relations.

That probe features and structure-building features must be ordered has been motivated independently: Some functional heads trigger more than one operation. Little v, for example, triggers ϕ-Agree and (external) Merge of the external argument. It thus has the following features: v { [•D•], [*ϕ*] }. The T head in English is another example: It triggers Agree with the subject DP and (internal) Merge to SpecT (the EPP property); it thus bears the following features: T { [•D•], [*ϕ*] }. Hence, at the point of the derivation where v or T is merged, a conflict arises (Müller 2009): The relevant head could either first discharge its structure-building feature and then the probe feature, or it could proceed the other way around. Assuming that only a single operation can apply at any given stage of the derivation, these operation-inducing features on v and T must be ordered.

It has recently been argued, although often rather implicitly, that the order of features on a head that triggers more than one operation is not arbitrary. Van Koppen (2005), Béjar and Řezáč (2009), Halpert (2012), and Assmann and Heck (2013) argue for a strict order of the two operations on certain heads.[1] Müller (2009), Heck and Müller (2007), Lahne (2008*a*), and Assmann, Georgi et al. (2013) on the other hand, argue that the order of Merge and Agree on v and T is in principle free and may vary between languages, i.e. it is determined in a language-specific fashion. The choice of the order in a given language is responsible for cross-linguistic variation, e.g. for the pattern of argument encoding (ergative vs. accusative), the extractability of core arguments and word order variation.

A simple example for a fixed order of Merge and Agree on T comes from English subject-verb-agreement where the operations are in a counter-bleeding relation: The T head is standardly assumed to trigger ϕ-Agree with the closest DP in its c-command domain, which is the subject DP base-generated in Specv, cf. (6). In addition, the T head has a [•D•]-feature (the EPP-feature) that triggers internal Merge of the subject to SpecT.

(6) *Basic structure of T′ in English*:
$[_{\mathrm{T}'}\ \mathrm{T}_{\{[*\phi:\square*],[\bullet \mathrm{D}\bullet]\}}\ [_{\mathrm{vP}}\ \mathrm{DP}_{ext\{[\phi:\textsc{2sg}]\}}\ [_{\mathrm{v}'}\ \mathrm{v}\ [_{\mathrm{VP}}\ \mathrm{V}\ \mathrm{DP}_{int}\ ]]]]$

In (6), the subject DP is in the c-command domain of T and T could thus Agree with it in ϕ-features; on the surface, however, the subject DP is no longer in the c-command domain of T because it has undergone EPP-driven movement to SpecT. Therefore, we expect that subject-verb-agreement is bled by EPP movement, but it is not, cf. (7) (the auxiliary *be*, which is located in T, is inserted to show that the subject has undergone EPP movement to SpecT).

---

[1] Van Koppen (2005) argues for internal Merge before Agree on the C head in languages with complementizer agreement to derive a bleeding effect; Halpert (2012) argues for the same order on an abstract head L in Zulu to account for the distribution of conjoint and disjoint morphology. Béjar and Řezáč (2009) derive hierarchy effects from the assumption that Agree between v and the internal argument of a transitive verb precedes external Merge of the external argument and may thus bleed Agree relations between v and the external argument; the proposal in Assmann and Heck (2013) also crucially relies on the order Agree before Merge on v.

(7) You are seeing John.

Hence, the operation Agree has applied although its context is not met in the output structure, cf. (8) with the subject moved to SpecT (checking the EPP on T) and the ϕ-probe on T valued by the subject.

(8) *Surface representation of TP in English*:
$[_{\text{TP}}\ \text{DP}_{ext\{[\phi:\text{2SG}]\}}\ [_{\text{T}'}\ \text{T}_{\{[\phi:\text{2SG}],\cancel{[\bullet D\bullet]}\}}\ [_{\text{vP}}\ \text{t}_{DP_{ext}}\ [_{\text{v}'}\ \text{v}\ [_{\text{VP}}\ \text{V}\ \text{DP}_{int}\ ]]]]]$

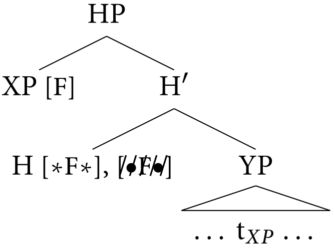
This effect is standardly derived by assuming that Agree takes place *before* the subject is EPP-moved, i.e., that the features on T are ordered as follows: T $\{ [*\phi*] > [\bullet\text{D}\bullet] \}$. Agree applies when the subject DP is still in the c-command domain of T; it is EPP-moved only afterwards.

In the data that will be examined in section 4, we will see that sometimes IM to the Spec of a head H has indeed a bleeding effect with respect to the Agree relation initiated by H, suggesting that IM takes place before Agree. On the other hand, in the same language that exhibits the bleeding effect with IM, IM sometimes also behaves like EPP-movement in English in that it counter-bleeds Agree initiated by H.

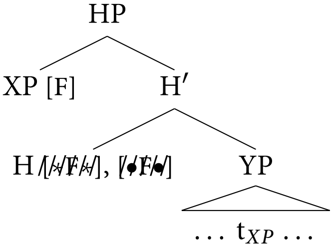
2.3. Opacity in the Present Data

The opaque interactions of internal Merge and Agree that will be presented in this paper are of the following type: A head H triggers Agree and internal Merge. It thus bears two operation-inducing features, a structure-building feature $[\bullet\text{F}\bullet]$ and a probe feature $[*\text{F}*]$ (a ϕ-probe or a case probe): H $\{ [\bullet\text{F}\bullet], [*\text{F}*] \}$. Sometimes the XP that moves to SpecH (and thereby checks the structure-building feature on H) feeds/bleeds Agree initiated by H, and sometimes XP movement has the opposite effect in the same position, i.e., it counter-bleeds/counter-feeds Agree. Thus, IM interacts transparently and opaquely with Agree in one and the same language. This situation is abstractly depicted for (counter-)bleeding in (9) and (10) which show the surface representation of HP:

(9) *Bleeding*:



(10) *Counter-Bleeding*:



On the surface, there is an XP with the features [F] in SpecH in both (9) and (10). XP movement is triggered by a structure-building feature on H that is then discharged (indicated by crossing-out the feature). Although XP occupies the same structural position in both examples, it has different consequences for the Agree relation that H triggers as well. In (9), Agree between the probe feature [*F*] and the matching feature [F] on XP is bled by XP movement: The probe on H is not discharged, indicating that Agree between the probe and the XP has not taken place. In (10), on the other hand, the probe [*F*] on H *is* discharged, resulting in valuation of the probe in case of ϕ-Agree and in case assignment to the XP for case-Agree, respectively; hence, H and XP must have entered into an Agree relation. Comparing the representation in (10) to the one in (9), one would expect bleeding in (10) as well because the XP occupies SpecH in both examples, but there is no bleeding, viz., this is an instance of counter-bleeding.

The split between the movement types that interact transparently and opaquely with Agree is not arbitrary. The split is conditioned by the type of IM: It depends on whether XP movement to SpecH is a final or an intermediate movement step in a movement chain. The background assumption is that movement of an XP to a position where it checks a structure-building feature [•F•] (its criterial position in the sense of Rizzi 2004, 2007) does not apply in one fell swoop. Rather, XP undergoes a sequence of short movement steps until it reaches its final landing site, i.e., it moves successive-cyclically. The positions in which the XP makes a stop-over but which are not the final landing sites are intermediate landing sites. Which positions constitute intermediate landing sites for which kinds of movement is a matter of an ongoing debate. In Chomsky's (2001) phase model, Specv and SpecC are identified as intermediate landing sites for movement out of vP and CP, respectively. I will come back to this issue in section 4. But before we continue, the notion of "final" landing site needs to be clarified. The term "final" is to be understood relative to a given movement chain and not in absolute terms: It denotes the position in a chain in which an XP checks a structure-building feature (in particular, a non-edge feature, see below) that triggered its movement in the first place – the XP moves in order to check that feature. This does not necessarily mean that this position is the *ultimate* landing site for the XP, i.e., that it cannot move on from that position. An XP in the final landing site of a chain α may undergo further movement in order to check another structure-building feature in the final landing site of another chain β.[2]

What is crucial for the analysis of the opacity effect is that IM to an intermediate landing site and IM to a final landing site can be distinguished on the basis of their triggers: The former is triggered by a special kind of structure-building feature (cf. Chomsky 2000, 2001), i.e., the edge-feature [•X•], a categorially underspecified structure-building feature; the latter is triggered by other kinds of structure-building features which encode a categorial or an interpretative property of the attracted XP, e.g. the feature [•WH•] on C that triggers IM of a wh-word, the EPP-feature [•D•] on T that attracts the closest DP, or the feature [•FOC•] that attracts a focus-marked XP, etc. We will see that edge feature-driven IM counter-feeds or counter-bleeds Agree, whereas non-edge feature-driven IM feeds or bleeds Agree. The pattern that arises from this is that intermediate movement steps of XP to SpecH triggered by edge features behave as if the XP is not moved at all with respect to Agree; final movement steps to SpecH triggered by non-edge features pattern the opposite way with respect to Agree.

These effects are derived by different orderings of the two types of IM triggers relative to the Agree trigger on H: The edge feature is discharged *after* the ϕ-probe on H, whereas the non-edge feature that triggers IM is discharged *before* Agree, cf. (11). Since Agree is structure-sensitive (to c-command, see section 3 below), *early* movements (i.e., the final movement steps) can influence possible Agree relations because they change structural relations. Inter-

---

[2] In this respect, *final* positions as defined in this paper (and as also used in Abels 2012) are different from the *criterial* positions introduced in Rizzi (2004, 2007). Rizzi assumes that once an XP has reached a criterial position, it is frozen in that position and cannot undergo any further movement steps (= *criterial freezing*, see Rizzi and Shlonsky 2007 on consequences of this assumptions for EPP-moved XPs). For this reason, I will not use the term *criterial* in what follows, although in some instances the final position of an XP may coincide with a criterial position in Rizzi's sense.

mediate movement steps apply *too late* to have this effect, they are still in their base position when the probe on H initiates a search, just as XPs that do not undergo movement at all.

(11) a. IM to the final landing site > Agree > IM to an intermediate landing site
  b. [•F•] > [*φ*] > [•X•]

3. Assumptions

I assume a strictly derivational model of syntax (cf. Chomsky 1995 et seq.) in which the structure unfolds in a bottom-up fashion by successive applications of the two basic operations Merge and Agree. The basic Minimalist clause structure is given in (12). The external argument is introduced in the specifier of the functional head v. Above vP, there are two more functional projections, TP and CP.

(12) *Clause structure*:
  $[_{\text{CP}}\ \text{C}\ [_{\text{TP}}\ \text{T}\ [_{\text{vP}}\ \text{DP}_{ext}\ [_{\text{v}'}\ \text{v}\ [_{\text{VP}}\ \text{V}\ \text{DP}_{int}\ ]]]]]$

All syntactic operations are feature-driven: Agree is triggered by probe features [*F*] and Merge is triggered by structure-building features [•F•] (I adopt the notation proposed in Sternefeld 2006, Heck and Müller 2007). Given this assumption, intermediate movement steps also need a trigger. As introduced in the discussion above, I take them to be triggered by edge features [•X•], which are categorially underspecified structure-building features. Furthermore, I assume that edge features are not freely available on functional heads; rather, they are inserted on a head H if H needs to attract an XP with a feature F although XP does not check the feature [•F•] on H (cf. *Phase Balance* in Heck and Müller 2000, 2003). For arguments that XPs must go through certain positions that are not their final landing sites see sections 4.1.3 and 4.2.3.

The operation Agree is defined in (13) (cf. Chomsky 2000, 2001). The crucial condition on Agree, which will be relevant in the derivation of opacity effects, is (13-a):

(13) Agree between a probe P and a goal G applies if
  a. P c-commands G,
  b. P has a feature [*F*] and G has a matching feature [F],

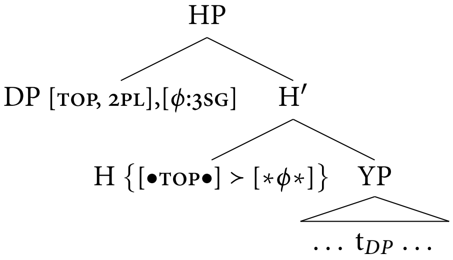
c. G is the closest matching goal for P.

d. Result: G values P (for ϕ-features) or P values G (for case features).

In what follows, case probes are represented as [∗c:value∗], i.e., they have a value which needs to be assigned to a DP without a value for case. If a case probe finds such a DP, it is discharged. ϕ-probes, on the other hand, do not have a value: [∗ϕ:□∗]. They seek for a value of a DP. If a ϕ-probe finds a DP with ϕ-features, the features of the DP are copied onto the probe and the probe is discharged. If Agree fails, i.e., if a probe does not find a goal, the derivation does not crash. Rather, default values are inserted either on the probe (in the case of ϕ-Agree) or on the goal (in the case of case Agree), cf. Béjar (2003), Preminger (2011). The default ϕ-features are [ϕ:3SG]; for the languages with a case split I will look at in this paper, the default case value is the nominative [C:NOM]. Furthermore, I adopt the Activity Condition (Chomsky 2001): A DP that has received a case value cannot be the goal for another Agree relation, neither for ϕ- nor for case Agree. Finally, I assume that traces or copies left by movement are not visible for Agree, i.e., they cannot serve as a goal for a probe. Alternatively, one might assume that traces/copies do not exist (because movement does not leave behind anything, see e.g. Epstein and Seely 2002, Unger 2010, Müller 2011, or because a multidominance approach is pursued, see e.g. Starke 2001, Abels 2004, 2012, Frampton 2004).

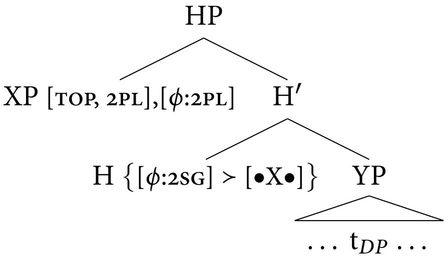
Given these assumptions, I will now show how IM and Agree can interact opaquely. An abstract counter-bleeding configuration is shown in (14) and (15). Suppose there is a head H which triggers IM and ϕ-Agree. For concreteness, it attracts an DP with a topic feature [TOP]. H also wants to Agree in ϕ-features with the topic DP, which bears the features 2nd person plural. If H first triggers IM as in (14), it moves the DP with the topic feature to its specifier ('[A] > [B]' on H indicates that the operation-inducing feature [A] is discharged before the feature [B]). Afterwards, the DP is no longer in the c-command domain of H. As a consequence, the ϕ-probe on H, which seeks for a goal in the next step, cannot target the DP with the topic feature anymore. It could have done so before the DP moved; hence, topic movement applies too early and thereby destroys the context for the application of Agree between H and $\text{DP}_{[\text{TOP}]}$. A default value (3sg) is thus inserted on the probe. This bleeding effect can be read off of the surface representation of HP: The DP with the topic feature is not in the c-command domain of H in (14) and hence it is clear why there is no ϕ-Agree. This effect will occur with final movement steps like e.g. topic movement or wh-movement to a head with a [•TOPIC•]- and a [•WH•]-feature, respectively.

(14) *Bleeding*:



However, in the same language that exhibits such bleeding effects, there are also configurations with the topic DP in SpecH but with full ϕ-agreement on H, cf. (15). Given (14), we would expect bleeding of ϕ-Agree here as well, but it does not occur. This instance of counter-bleeding can be derived if the probe and the structure-building feature on H are discharged in the reverse order, i.e., if the DP with the topic feature moves *after* Agree has applied (cf. the order [*ϕ*] > [•X•] on H in (15)). The consequences are the following: First, the probe seeks for a goal. At this point of the derivation, the topic DP is still in the c-command domain of H and therefore H is valued with the ϕ-features of that DP. Afterwards, H triggers movement of the topic DP.

(15) *Counter-Bleeding*:



On the surface, it is unclear why Agree between H and the topic DP has applied given that the DP is not in the right structural position to be a goal for H, i.e. it is not in H's c-command domain. The explanation is that at the point of probing, the DP has indeed been in the relevant domain, but this is disguised by the subsequent movement of the DP. This counter-bleeding pattern will arise with intermediate movement steps in a movement chain, viz., with movement steps that are triggered by edge features.

The conclusion is that in languages with both types of interactions in the HP, IM sometimes applies before and sometimes after Agree, depending on the nature of the landing site of the moved XP. The landing sites can be locally distinguished on H on the basis of their trigger: Intermediate movement steps are triggered by edge features on H; they are discharged *after* probe features; final movement steps are triggered by structure-buiding features on H that are not categorially underspecified; these are discharged *before* probe features. This analysis presupposes a strictly derivational model of grammar.

4. Case Studies

In this section I present three phenomena that instantiate the abstract pattern of an opaque interaction of Merge and Agree introduced in sections 1 and 3. The counter-bleeding pattern arises with the Anti-agreement effect for ϕ-Agree, and with the possessor case split in Hungarian for case Agree. Defective Intervention in an Icelandic dialect exhibits a counter-feeding interaction with respect to ϕ-Agree.

4.1. (Counter-)Bleeding: The Anti-Agreement Effect

4.1.1. Data and Rule Interactions

In a number of unrelated languages such as Berber, Breton, Welsh, Kinande, Kikuyu, Palaun and Turkish, ϕ-Agree between the verb and the subject reduces to default agreement (3sg) if the subject is Ā-moved (questioned, relativized, focussed) to the SpecC position of the minimal clause. This phenomenon is known as the Anti-agreement effect (AAE, cf. Ouhalla 1993, Richards 1997, Phillips 2001 for an overview). It is illustrated with wh-extraction data from Berber in (16). If the subject is not Ā-moved, the verb agrees with it in person, number and gender, cf. (16-a). If, however, the subject is questioned (Ā-moved to the minimal SpecC), as in (16-b), the verb shows default agreement (3sg masculine, glossed as 'participle').[3] Full agreement is ungrammatical in this context, cf. (16-c). Interestingly, in some of

---

[3] The verb form in (16-b) is glossed as 'participle', but Ouhalla (1993) argues that this form contains the default 3rd person masculine form of agreement. In most other languages with

the languages that exhibit the AAE when the subject is extracted to SpecC of the minimal clause, long-distance Ā-movement of the subject to SpecC of a higher clause does not result in default agreement in the embedded clause from which the subject is extracted; instead, there is full ϕ-agreement on the embedded verb (cf. (16-d)), as if the subject was not extracted at all.

(16) *Anti-agreement in Berber, wh-movement (Ouhalla 1993: 479f.)*:

a. zri-n imhdarn Mohand
saw-3PL students Mohand
'The students saw Mohand.' *no Ā-IM, full agr.*

b. man tamghart ay yzrin Mohand
which woman COMP see.PART Mohand
'Which woman saw Mohand?' *local Ā-IM, default agr.*

c. *man tamghart ay t-zra Mohand
which woman COMP 3SG.FEM-saw Mohand
'Which woman saw Mohand?' *local Ā-IM, full agr.*

d. man tamghart ay nna-n qa t-zra Mohand?
which woman COMP said-3PL that 3SG.FEM-saw Mohand
'Which woman did they say saw Mohand?' *long Ā-IM, full agr.*

In what follows, I will use the term "short" Ā-movement for Ā-movement to the SpecC position of the minimal clause and "long" Ā-movement for Ā-movement to the SpecC position of a higher clause.

In terms of rule interaction, the AAE can be described as follows: Short Ā-movement of the subject DP *bleeds* ϕ-Agree whereas long Ā-movement has the opposite effect. If a subject that is to undergo long-distance extraction uses SpecC of the minimal clause as an intermediate landing site, we have an instance of *counter-bleeding* when looking at the minimal CP: Under both short and long Ā-movement, there is a DP in the local SpecC position, but sometimes IM of the subject bleeds ϕ-Agree in that position, and sometimes it does not bleed ϕ-Agree in the same position. Put differently, ϕ-Agree did apply in the embedded clause under long extraction, although its context does not seem to be met, given that a DP in the minimal SpecC position can bleed agreement (under short extraction). Hence, on the surface, there is opacity. Importantly, opacity only arises if the subject DP that is to undergo long-

---

the AAE, the verb form under short Ā-extraction is completely identical to 3rd person singular agreement.

extraction makes a stop-over in SpecC of the embedded clause. For syntactic, morphological and semantic evidence for the successive-cyclic nature of long Ā-movement through the embedded SpecC see McCloskey (1979), Cole (1982), Clements et al. (1983), Torrego (1984), Barss (1986), Lebeaux (1990), Chung (1994), Fox (2000), Lahne (2008*b*) among many others. I adopt this assumption and present another argument in favor of successive-cyclic movement through intervening SpecC positions in AAE languages in section 4.1.3.

4.1.2. Analysis

First of all, I assume that subject-verb agreement in AAE languages is mediated by C and not by T (see also Ouali 2008 and Henderson (2009) for this assumption about some AAE languages).[4] The reason for this assumption is that short and long Ā-movement can be distinguished by the nature of their trigger on the minimal C head: With respect to short Ā-movement of DP, SpecC of the minimal clause is the final landing site for DP; this movement step is triggered by a wh-feature [•wh•] (or [•rel•], which triggers operator movement in a relative clause, etc.). For long Ā-movement, this position is only an intermediate landing site and movement to it is triggered by an edge feature [•X•]. The opacity effect is derived if final movement steps to SpecC apply before Agree and intermediate movement steps to the minimal SpecC apply *after* Agree:

(17) *Order of features on C in Berber type languages*:
C { [•wh•] ≻ [$*\phi*$] ≻ [•X•] }

The consequences of this ordering are as follows: If there is no subject extraction, as shown in (18), C only bears a ϕ-probe and no structure-building features that trigger IM. When the probe searches for the closest goal in its

---

[4] Since I assume that the ϕ-probe in AAE languages in on C, but the inflection shows up on the verb, there must be a connection between C and V: Either the verb picks up the inflection in C by V-to-C movement (cf. Sproat 1985, Jouitteau 2005 on Celtic) or the inflection is lowered to the T-V-complex in the morphological component. Lowering of the inflection is reminiscent of Affix-Hopping (Chomsky 1957). Furthermore, it seems plausible given the concept of Feature Inheritance, i.e., ϕ-feature transfer from C to T (cf. Chomsky 2004, Richards 2007); the only difference is that I assume that this transfer may apply late, in the morphological component (cf. morphological merger in Embick and Noyer 2001). See also Ouali (2008) on variation in the timing and mode of Feature Inheritance.

c-command domain, it finds the subject DP and is valued with the ϕ-features of this DP.

(18) *No Ā-extraction of the subject DP, ϕ-probe discharged*:

$[_{\text{CP}}\ \text{C}_{\{[*\phi*]\}}\ \ [_{\text{TP}}\ \text{DP}\ [_{\text{T}'}\ \dots\ ]]]$

└ ✓Agree ┘

If the subject is questioned in the minimal CP (short Ā-movement), C bears a ϕ-probe and the IM trigger [•WH•]. The structure-building feature is discharged first; it attracts the wh-subject to SpecC, cf. (19-a). Afterwards, the probe initiates Agree but since the closest goal, the subject DP, is no longer in C's c-command domain due to prior movement, Agree fails and the probe is valued by default, cf. (19-b). Hence, final movement of the subject DP to SpecC applies too early and thereby bleeds ϕ-Agree.

(19) *Short Ā-extraction of the subject DP*:

a. Step 1: Movement of the subject to SpecC, wh-feature discharged:

$[_{\text{CP}}\ \text{DP}_{wh}\ \text{C}_{[\bullet\textsc{wh}\bullet]\ \succ\ [*\phi*]\}}\ [_{\text{TP}}\ \text{t}_{DP}\ [_{\text{T}'}\ \dots\ ]]]$

└──── Move ────┘

b. Step 2: ϕ-Agree initiated by C, default valuation of the probe:

$[_{\text{CP}}\ \text{DP}\ \text{C}_{\{\cancel{[\bullet\textsc{wh}\bullet]}\ \succ\ [*\phi*]\}}\ [_{\text{TP}}\ \text{t}_{DP}\ [_{\text{T}'}\ \dots\ ]]]$

└ *Agree ┘

In contrast to that, long Ā-extraction that uses the minimal SpecC as an intermediate landing site is triggered by an edge feature. This movement applies late enough and therefore does not bleed Agree: First, the probe on the C head of the embedded sentence initiates Agree and finds the subject DP as its closest goal, cf. (20-a). Only afterwards does the subject move to SpecC, triggered by the edge feature on C, cf. (20-b). On the surface, the subject DP is no longer in the c-command domain of the embedded C, which gives rise to opacity; however, the subject was in the relevant domain at the point of ϕ-Agree.

(20) *Long Ā-extraction of the subject DP*:

a. Step 1: ϕ-Agree initiated by C, ϕ-probe discharged

$$[_{\text{CP}}\ \text{C}_{\{[*\phi*]\succ[\bullet\text{X}\bullet]\}}\ [_{\text{TP}}\ \underbrace{\text{DP}_{wh}}_{\checkmark\text{Agree}}\ [_{\text{T}'} \ldots\ ]]]$$

b. Step 2: Movement of the subject DP to the local SpecC, edge feature discharged

$$[_{\text{CP}}\ \text{DP}_{wh}\ [_{\text{C}'}\ \text{C}_{\{\cancel{[*\phi*]}\succ[\bullet\text{X}\bullet]\}}\ [_{\text{TP}}\ \text{t}_{DP}\ [_{\text{T}'} \ldots\ ]]]]$$

Move

Before closing this section, I will address some remaining issues that arise in the derivations presented above. The first question concerns short Ā-extraction in (19). Since short extraction applies early, the ϕ-probe on C cannot target the extracted subject DP anymore. But why can't it just target the closest target that is still in its c-command domain, i.e., the object DP if the verb is transitive? I propose that this is impossible due to the Activity Condition: The object is assigned case by v very early in the derivation. It thereby becomes inactive, i.e., invisible for a ϕ-probe. C thus cannot target the object.[5]

A further issue that arises is that of case assignment to the subject DP: If ϕ-Agree and case assignment go hand in hand, where does a subject that undergoes short Ā-movement to the minimal SpecC get its case value from? Since Agree between C and the subject DP is bled in this configuration C can also not assign a case to the subject. I propose that default valuation applies for case just as for ϕ-features. The case feature of the subject is valued with the default nominative case. See section 4.3 for another instance of default case valuation under short extraction.

---

[5] In section 4.2.3 I will argue that SpecT is an intermediate landing site for movement to SpecC. If this holds in general, as I will assume, a related problem arises: The AAE does not occur with long Ā-movement of the object. If, however, a wh-moved object goes through SpecT, there is a stage in the derivation at which the wh-object is in SpecT of the embedded clause and above the subject DP. Why can the ϕ-probe on C not Agree with the closer object DP in this configuration (since the object undergoes long extraction and checks the edge feature on the embedded C, it is still in C's c-command domain when the ϕ-probes searches for a goal)? As proposed above, this is impossible because the object is inactive for ϕ-Agree at that point of the derivation.

4.1.3. *More Evidence from Variation for SpecC as an Intermediate Landing Site*

The AAE data are only opaque if long Ā-movement makes a stop-over in the embedded SpecC position. However, there is an alternative analysis of the counter-bleeding effect in AAE languages that does not rely on this intermediate movement step: If short movement of a DP to SpecC causes bleeding but there is no bleeding under long Ā-extraction, one could propose that wh-movement to the matrix clause does not make a stop-over in the SpecC position of the embedded clause but moves directly to the matrix SpecC. Under this analysis, it follows why we find full subject-verb-agreement in the embedded clause: When the ϕ-probe on the embedded C searches for a goal, the wh-subject is still in its base position in the embedded clause and thus in the embedded C's c-command domain; it can only move when matrix C is merged with the IM trigger [•WH•]. But due to incremental bottom-up structure-building, matrix C is merged later than the embedded C head. This account does not require ordering of two different types of IM.

However, the assumption that long Ā-movement does not go through the embedded SpecC position causes problems for languages in which *both* long and short Ā-movement bleed ϕ-Agree. This is the case for example in some Italian dialects (Fiorentino and Trentino, cf. Brandi and Cordin 1998, Campos 1997) and in Ibibio (cf. Baker 2008). (21) illustrates that in Fiorentino full subject agreement under short extraction is ungrammatical; instead, default agreement must be used, compare (21-a) and (21-b). Default agreement is also necessary under long subject extraction, compare (21-c) and (21-d).

(21) *The AAE in Fiorentino (taken from Ouhalla 1993, Campos 1997)*:

a. Quante ragazze gli ha parlato con te?
how.many girls CL.3SG have.3SG spoken to you
'How many girls (it) has spoken to you?'

b. *Quante ragazze le hanno parlato con te?
how.many girls CL.3PL have.3PL spoken to you
'How many girls have spoken to you?'

c. Quante ragazze tu credi che gli ha telefonato?
how.many girls you think that CL.3SG have.3SG phoned
'How many girls do you think have phoned?'

d. *Quante ragazze tu credi che le hanno telefonato?
how.many girls you think that CL.3PL have.3PL phoned

In contrast to Berber, long Ā-movement of the subject also bleeds Agree in the embedded clause. But under the assumption that long extraction does not go through the embedded SpecC position, this is mysterious: The subject DP that is to undergo long extraction is still in the c-command domain of C when the latter probes and should thus value the probe. One solution to this problem could be to say that languages differ in the size of their locality domains: In AAE languages of the Berber type, long Ā-movement does not go through SpecC, which explains the absence of bleeding with long extraction; in AAE languages of the Trentino type, however, long Ā-movement does go through SpecC and DPs in the embedded SpecC cause bleeding (due to early movement prior to ϕ-Agree), regardless of whether SpecC is their final landing site or not.

In my view it is clearly undesirable to assume that languages differ in the size of their locality domains. In the languagues in which long Ā-movement does not apply successive-cyclically, there must be a non-local dependency between the matrix C and a DP in the embedded clause. This goes against the trend in contemporary minimalist syntax, where non-local dependencies are decomposed into a series of local dependencies (cf. Alexiadou et al. 2012). Under the analysis of the opacity effect presented in 4.1.2, there is another way to account for variation between AAE languages that I will adopt: Assume that languages do not have different locality domains for movement. In order to model bleeding with long extraction in the Trentino type languages, the embedded SpecC must be an intermediate landing site. So it must be in languages of the Berber type, with the consequence that the AAE data in these languages are indeed opaque. Cross-linguistic variation is accounted for by reordering of operation-inducing features on C: In the Trentino type languages, *both* edge-feature-driven IM and non-edge feature-driven IM apply *before* Agree. Hence, the wh-subject DP is moved out of the c-command domain of C before probing happens and there is thus always bleeding, regardless of the IM type.

(22) *Order of features on C in Trentino type languages*:
C { [•wh•], [•X•] ≻ [∗ϕ∗] }

In Berber type languages, on the other hand, the two IM triggers apply at different points in the derivation relative to Agree, giving rise to opacity; compare the order of features in (17) and (22).

Previous analyses of the AAE pattern include Ā-binding approaches (Brandi and Cordin 1998, Ouhalla 1993) and anti-locality approaches (Cheng 2006, Schneider-Zioga 2007). The basic idea of the former type of approach is that subject extraction leaves behind a *pro* in the subject position that must not be bound by its antecedent from an Ā-position (Principle B applied to Ā-binding). It is, however, bound under short subject extraction by the subject in SpecC. In order to avoid the violation of Principle B, *pro* must not be licensed in the first place. It is licensed by rich subject-verb-agreement; dropping the agreement, resulting in AAE, makes *pro* unavailable and thus avoids the principle B violation.

In the anti-locality approaches, it is argued that short extraction of the subject is too local in the sense of Grohmann (2003) (at least in some AAE languages). Hence, a repair mechanism must apply: The lower copy of the subject is spelled out, but not as a DP or a pronoun; rather, it is the anti-agreement morphology (default agreement) that is the spell-out of that copy.

Apart from a number of conceptual problems with these approaches that I will not address here, both of the approaches have problems to account for cross–linguistic variation. One approach needs to postulate that languages differ in the size of their locality domains, and the other must assume that they differ in the properties of empty elements such as traces. In the Ā-binding approach presented in Ouhalla (1993), long Ā-extraction passes through SpecC of the embedded CP. Since languages like Berber do not show the AAE with long extraction, Ouhalla assumes that the trace of the subject in the embedded SpecC position cannot act as an Ā-binder. To account for the occurrence of the AAE with long extraction in the Trentio-type languages, he proposes that the trace can be an Ā-binder in these languages. Hence, languages vary in the ability of traces to be Ā-binders. Apart from the question whether movement leaves behind traces, copies or maybe no element at all, it seems questionable that an empty element left by the same movement operation in the same position varies in this way.

In the anti-locality approaches, it is simply assumed that long extraction does not go through the embedded SpecC position but moves directly to the matrix SpecC (in languages of the Berber type). Since this movement is not too local, there is no repair, i.e., no anti-agreement morphology on the embedded verb. The Trentino pattern with anti-agreement showing up even under long extraction is not addressed in these approaches. One would probably have to assume that long movement does make a stop-over in the embedded

SpecC position. Hence, languages have to differ in their locality domains. As pointed out above, I take this to be undesirable. The reordering approach presented in this paper does not have to assume that languages differ in this respect, and it also does not need to postulate that elements left by movement (if there are such elements at all) have different binding properties. I take this to be a desirable outcome.

4.2. (Counter-)Feeding: Intervention Effects in Icelandic B

4.2.1. Data and Rule Interactions

In many languages, a dative XP blocks Agree relations between a probe on a head H that c-commands the dative XP and a lower DP that is c-commanded by H and the dative XP. This is remarkable because the dative XP itself can also not value the probe on H. This effect is known as defective intervention (Chomsky 2001). In Icelandic, dative experiencers (Exp) are defective interveners for Agree between T and a lower subject DP. As shown in (23-a), the raising verb can only bear default 3rd singular agreement because Agree with the subject in the embedded infinitive is blocked by Exp. Sigurðsson and Holmberg (2008) describe three Icelandic dialects that pattern differently with respect to the presence vs. absence of the intervention effect in the context of experiencer movement. Opaque interaction of Agree and IM is found in Icelandic B for number agreement as shown in (23-b) and (23-c) (see also Holmberg and Hróarsdóttir 2003 for a description of that dialect):[6]

(23) *Raising constructions in Icelandic B (Holmberg and Hróarsdóttir 2003)*:

a. það virðist/*virðast **einhverjum manni** [hestarnir
there seem.3SG/seem.3PL some man.DAT the-horses.NOM
vera seinir]
be slow

---

[6] As Sigurðsson and Holmberg (2008) show, person agreement patterns differently from number agreement. Agreement in 1st and 2nd person is blocked in all three dialects, regardless of the position of the dative experiencer. In what follows, I confine myself to number agreement. The facts on person agreement can be integrated into the present analysis if person and number are separate probes, as Sigurðsson and Holmberg (2008) have argued, and if the person probe, searching for local person arguments, probes *before* any movement operation has taken place; i.e., the person probe is the feature on T which is discharged before all other operation-inducing features.

‘It seems to some man that the horses are slow.’

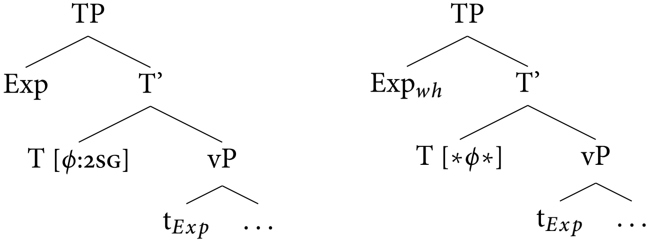
b. **Mér** virðast t$_{NP}$ [hestarnir vera seinir]
me.DAT seem.3PL the-horses.NOM be slow
‘It seems to me that the horses are slow.’

c. **Hvaða manni** veist þú að virðist/*viðast t$_{wh}$
which man.DAT know you that seem.3SG/seem.3PL
[hestarnir vera seinir]
the-horses be slow
‘To which man do you know that the horses seem to be slow?’

If the experiencer moves above the ϕ-probe on T in the matrix clause in order to check the EPP feature on T, this movement step feeds Agree between T and the subject DP in the embedded clause, cf. (23-b).[7] This is not surprising since, on the surface, the experiencer does not intervene anymore between probe and goal. What is striking, however, is that if the experiencer is moved to SpecC (questioned, relativized, topicalized), it does not feed Agree, although on the surface it also does not intervene anymore, cf. (23-c). This is a case of counter-feeding: Agree between T and the subject DP cannot apply although its context is given – the intervener is moved out of the way, just as in (23-b). Hence, a wh-moved experiencer behaves as if it is not moved at all with

Opacity arises at the stage of the derivation where TP is complete, cf. (24). Assume that wh-movement to SpecC makes a stop-over in SpecT (see section 4.2.3 for justification). At this stage in both (23-b) and (23-c), the experiencer is in SpecT. But in (23-b) it feeds Agree in this position whereas in (23-c) it has the opposite effect in the same position.

(24) a. Feeding (EPP-movement): b. Counter-feeding (wh-movement):



---

[7] Due to the V2-property of Icelandic, the finite verb moves to C and the dative subject in (23-b) undergoes further movement to SpecC in order to fill the sole preverbal position. What is important for the present discussion is that the subject has also checked the EPP in SpecT (see Holmberg and Hróarsdóttir 2003, Sigurðsson and Holmberg 2008 for evidence); subject movement to SpecT is thus a final movement step.

4.2.2. Analysis

The crucial difference between the feeding and the counter-feeding example again lies in the landing site of the experiencer: In (23-b), SpecT is the final landing site in the movement chain triggered by the EPP on T.[8] In (23-c), however, SpecT is only an intermediate landing site on the way to SpecC, where the feature [•WH•] is checked (for discussion of how the EPP is checked in this configuration see below). Intermediate movement steps are triggered by an edge feature on T and apply *after* Agree; final movement steps to SpecT are triggered by the EPP feature, represented as [•D•] in what follows, and apply *before* Agree. Just as in the AAE languages of the Berber type, the two movement types apply at different points in the derivation with respect to Agree.

(25) *Order of features on T (Icelandic B)*:
T { [•D•] ≻ [∗ϕ∗] ≻ [•X•] }

The consequences of this ordering are the following: If the experiencer is not moved at all and the EPP is checked by another element such as the expletive in (23-a), the experiencer is in the c-command domain of T and thus intervenes for ϕ-Agree between the probe on T and the lower subject DP, cf. (26).

(26) *No movement of the experiencer: default agreement*:
$[_{\text{TP}}\ \text{T}_{.\,[*\phi*]\}}\ [_{\text{vP}}\ \text{v}\ [_{\text{VP}}\ \boxed{\text{Exp}}\ [_{\text{V}'}\ \text{V}\ [_{\text{TP}}\ \text{DP}\ [_{\text{T}'}\ \ldots\ ]]]]]]$
*Agree

If the experiencer undergoes EPP-movement, this movement step applies before T initiates Agree, cf. (27-a). Afterwards, the ϕ-probe on T can target the subject DP since the experiencer is no longer in the c-command domain of T and thus does not intervene anymore, cf. (27-b).

(27) *EPP-movement of the experiencer, full agreement*:

a. Step 1: movement of the experiencer to SpecT, EPP discharged
Move
$[_{\text{TP}}\ \square\ [_{\text{T'}}\ \text{T}_{\{[\bullet\text{D}\bullet]\ \succ\ [*\phi*]\}}\ [_{\text{vP}}\ \boxed{\text{Exp}}\ [_{\text{v'}}\ \text{v}\ [_{\text{VP}}\ t_{Exp}\ [_{\text{V}'}\ \text{V}\ [_{\text{TP}}\ \text{DP}\ [_{\text{T}'}\ \ldots\ ]]]]]]]]$

---

[8] SpecT is not the ultimate landing site for the subject, which undergoes further movement to SpecC. Recall that "final" means final in a given chain. Movement to SpecT is the final step in the movement chain created to check the EPP in SpecT. See section 2.3 and footnote 7.

b. Step 2: Agree between T and the lower DP, ϕ-probe discharged:
[$_{\text{TP}}$ Exp [$_{\text{T'}}$ T$_{\{\cancel{[\bullet D\bullet]}\ \succ\ [*\phi*]\}}$ [$_{\text{vP}}$ t$_{Exp}$ [$_{\text{v'}}$ v [$_{\text{VP}}$ t$_{Exp}$ [$_{\text{V}'}$ V [$_{\text{TP}}$ DP [$_{\text{T}'}$ …]]]]]]]]
✓Agree

If the experiencer is questioned and must check the [•WH•]-feature on C, it uses SpecT only as an intermediate landing site. This movement step is triggered by an edge feature and applies after Agree. First, the ϕ-probe on T searches for a goal but Agree is blocked because the wh-experiencer still intervenes, cf. (28-a). It is moved out of the way only afterwards, cf. (28-b). Movement of the experiencer comes too late to feed Agree.

(28) *Wh-movement of the experiencer, default agreement*:

a. Step 1: Agree initiated by T fails, Exp$_{wh}$ still intervenes, default agreement
*Agree
[$_{\text{T'}}$ T$_{\{\ [*\phi*]\ \succ\ [\bullet X\bullet]\}}$ [$_{\text{vP}}$ $\boxed{\text{Exp}_{wh}}$ [$_{\text{v}'}$ v [$_{\text{VP}}$ t$_{Exp}$ [$_{\text{V}'}$ V [$_{\text{TP}}$ DP [$_{\text{T}'}$ …]]]]]]]

b. Step 2: Intermediate movement step to SpecT, edge feature discharged:
Move
[$_{\text{TP}}$ Exp$_{wh}$ [$_{\text{T'}}$ T$_{\{\ \cancel{[*\phi*]}\ \succ\ [\bullet X\bullet]\}}$ [$_{\text{vP}}$ t$_{Exp_{wh}}$ [$_{\text{v'}}$ v [$_{\text{VP}}$ t$_{Exp}$ [$_{\text{V}'}$ V [$_{\text{TP}}$ DP [$_{\text{T}'}$ …]]]]]]]]

With respect to the counter-feeding derivation in (28) a question arises: To obtain counter-feeding it is important that the wh-experiencer checks the edge feature on T and not the EPP feature. If it could check the EPP, movement to SpecT would apply early and would lead to feeding with wh-movement. However, since the EPP is always present on T (in contrast to the [•WH•]-feature on the embedded C head in the AAE derivation with long Ā-extraction), it is not clear why a wh-phrase cannot check it. I do not have an answer why the wh-phrase does not discharge the EPP, but it is clear for Icelandic that it really doesn't: As Holmberg and Hróarsdóttir (2003) show, the EPP can be checked by a different phrase in exactly the configuration where the experiencer undergoes wh-movement (or relativization, topicalization) to SpecC. This is an instance of stylistic fronting, an operation that moves the closest overt VP-internal category to SpecT to check the EPP (Holmberg 2000). Stylistic fronting is optional. An instance of this operation in the context of a wh-experiencer is given in (29) where *Òlafur* is fronted (Holmberg and Hróarsdóttir 2003: 1009):

(29) Hverjum hefur Òlafur virst vera gàfaður?
whoDAT has OlafNOM seemed be intelligent
'To whom has Olaf seemed to be intelligent?'

4.2.3. More Evidence from Variation for SpecT as an Intermediate Landing Site

On the surface, the interaction of Agree and IM in Icelandic B is opaque because, when looking at the output structure, neither the EPP-moved nor the wh-moved experiencer intervenes between the matrix T and the embedded subject; nevertheless, the wh-experiencer behaves as if it intervened. Derivationally, this could be derived by assuming that the wh-experiencer moves directly to SpecC without a stop-over in SpecT. Under this analysis, it is clear why the wh-experiencer causes an intervention effect: When T probes, the wh-experiencer is still in the c-command domain of T and thus blocks Agree. It is moved to SpecC only after C is merged (this is the solution proposed by Holmberg and Hróarsdóttir 2003). However, assuming that wh-movement does not stop in SpecT causes problems for cross-linguistic variation found with defective intervention: In another Icelandic dialect, Icelandic A, *both* wh-movement and EPP-driven movement of the experiencer feed Agree, i.e., (23-c) is grammatical with 3rd person plural agreement on the verb in this dialect (Sigurðsson and Holmberg 2008). But wh-movement of the experiencer should not cause feeding if wh-movement does not go through SpecT. When T probes for the lower DP, a wh-experiencer still intervenes; it is moved when C is merged with TP at a later stage of the derivation.

A similar pattern exists in Romance languages and Greek, although feeding cannot be read off of the agreement on the verb (McGinnis 1998, Anagnostopoulou 2003): In these languages, an experiencer blocks movement of (instead of ϕ-Agree with) a lower DP to SpecT. If the experiencer is cliticized (adjoined to T) or questioned (moved to SpecC), movement of the lower DP to SpecT becomes possible (feeding). Hence, both types of IM, i.e., movement to SpecC and to the T domain, feed Agree. Assuming that wh-movement does not go through SpecT is problematic for analyzing these languages as well: When T tries to attract the lower DP, a wh-experiencer still intervenes because it is only moved after C is merged, which only happens after the TP is complete. A way out would be to assume that in both Icelandic A, Greek and the Romance languages, direct wh-experiencer movement to SpecC must precede attraction of the subject DP, but the latter step is counter-cyclic if a wh-Exp does not go through SpecT. For Greek and the Romance languages, several solutions to this problem have been proposed in the literature: (i) phase-internal counter-cyclicity (Anagnostopoulou 2003), (ii) intervention is evaluated at the phase-level (McGinnis 2001), (iii) covert movement of the wh-experiencer to a low Ā-position in the c-command domain of T that is invisible to a probe on T because ϕ-Agree is an A-relation and therefore cannot target elements in Ā-positions (Legate 2002). Solution (i) gives up the Strict Cycle Condition within a phase (CP and vP) and thus standard locality assumptions. Solution (ii) suggests that intervention is computed when a phase is complete, i.e., at the CP level. Since a wh-experiencer is in SpecC at the end of derivation, it does not intervene on the surface. However, this means that the Minimal Link Condition is a strongly representational constraint (for the examples under discussion, it is evaluated at the end of the derivation) in a system that is assumed to be strictly derivational. Hence, it is argued that the data are not compatible with a strictly derivational model of grammar (this has also been pointed out in Řezač 2004). Finally, solution (iii) resorts to a covert movement step to a low Ā-position that is hardly motivated independently. In order to avoid all these problematic assumptions one could propose that languages differ in the size of their locality domains: In Icelandic B-type languages, wh-movement does not make a stop-over in SpecT, which accounts for the absence of feeding with wh-experiencers. In languages of the IcelandicA-/Romance/Greek type, however, wh-movement must proceed in smaller steps and does go through SpecT, explaining the feeding effect. Just as with the variation found with the AAE, the question is whether it is desirable to assume that languages differ in such a fundamental property like the locality domains for movement.

I will pursue another solution that (i) neither needs to weaken the Strict Cycle Condition nor (ii) needs to invoke strongly representational constraints (evaluated at the end of the derivation), nor (iii) postulates different locality domains in different languages. I assume that wh-movement always goes through SpecT, in all the languages discussed above (for the same assumption cf. Chomsky 2005, Gallego 2006, see also Müller 2004 and references cited there). This is needed to account for feeding in Romance-type languages and since languages do not vary in this property, it must hold for languages of the Icelandic B-type as well (and the data from this dialect are thus indeed opaque). As with the variation in AAE languages, I propose that cross-linguistic variation is accounted for by reordering of operation-inducing features on a head. In contrast to Icelandic B, the EPP and the edge feature are ordered before the probe feature in T in Icelandic A and Romance/Greek, cf. (30).

(30) Order of features on T in Icelandic A:
$T \{ [\bullet D\bullet], [\bullet X\bullet] \succ [*\phi*] \}$

As a consequence, the experiencer, whether it is a wh-experiencer or not, moves early to SpecT and does not intervene anymore at the point when T probes. In Icelandic B, the two IM triggers align differently with respect to the probe feature, which gives rise to opacity. In this way, the transparent (Romance, Icelandic A) and opaque (Icelandic B) interactions of Merge and Agree in the two types of languages can be accounted for in a uniform way. The only difference between them is the order of operation-inducing features on a head that triggers more than one operation – a purely lexical property. Languages do not differ in the constraints that determine the size of locality domains (like the PIC). The approach has the additional advantages that standard locality constraints are obeyed to and that the strongly representational version of the MLC as proposed in Anagnostopoulou (2003) can be avoided in a strictly derivational system.[9]

---

[9] Fabian Heck has pointed out to me that if the intervention of the dative experiencer in the present analysis is computed on the basis of an MLC-like constraint, the MLC would still have a representational residue because it evaluates the distance of two potential goal XPs relative to the T head, i.e., the TP must be scanned to see which goal is closer to T; however, this version of the MLC is arguably of a weaker type because it is evaluated before the derivation is complete, in contrast to the strongly representational version discussed in the main text. Note that the portion of the derivation that needs to be accessible to evaluate intervention can be further reduced if a probe simply targets the first available goal in its c-command domain and if the dative experiencer actually counts as a potential goal which can, however, not value the probe (as has been suggested in the literature on defective intervention, see Preminger 2011 for an overview of relevant accounts). As a consequence, the probe stops searching as soon as it finds the experiencer in its c-command domain and does not access material that is lower down in the structure.

To summarize, the Icelandic B data are not only opaque when looking at the surface position of the experiencer relative to the T head; they are also opaque when looking at the TP (before the derivation is complete): The experiencer is in SpecT, regardless of whether it checks the EPP or an edge feature in this position, but movement of the experiencer to SpecT has different consequences for Agree. Hence, Agree and Internal Merge interact opaquely.

4.3. (Counter-)Bleeding: Possessor Case and Agreement in Hungarian

4.3.1. Data and Rule Interactions

In this subsection I will present another instance of counter-bleeding. In contrast to the two previous instances of opaque interactions of Merge and Agree, the Agree relation will result in case valuation instead of ϕ-agreement.

The phenomenon that shows opacity in this respect is the case marking of the possessor in Hungarian. In Hungarian, the possessor precedes the possessum and the possessum agrees in ϕ-features with the possessor. The possessor exhibits a case split: It can bear either nominative or dative, cf. (31). As argued in Szabolcsi (1994), the two possessors occupy different structural positions: the dative possessor is in an operator position within the DP, which is taken to be SpecD. It is moved to that position from a lower position that is associated with nominative case. Evidence for this analysis comes from two facts: (i) the dative possessor precedes the D element *a(z)* whereas the nominative possessor follows this element (cf. (31-a) vs. (31-b)); (ii) only the dative possessor can be extracted out of the DP, cf. (32). As long as the possessor is not to undergo movement out of the DP, the choice between a nominative and a dative possessor is optional. Finally, note that both the dative and the nominative possessors precede a special class of determiners like 'each', cf. (31-c).

(31) *Possessors in Hungarian*:

a. (a) Mari kalap-ja
the Mari.NOM hat-POSS.3SG
'Mari's hat'

b. Mari-nak a kalap-ja
Mari.DAT the hat-POSS.3SG
'Mari's hat'

c. a te valamenyi tiik-od
the you.NOM each secret-POSS.2SG
'your every secret'

(32) *Extraction data*:

a. Mari-nak nem ismert-em [t′ t növér-é-t]
Mari.DAT not knew-1SG sister-POSS.3SG-ACC
'I never knew any sister of Mari'

b. *Mari nem ismert-em [t′ t növér-é-t]
Mari.NOM not knew-1SG sister-POSS.3SG-ACC
'I never knew any sister of Mari.'

With this empirical background, I make the following assumptions about the structure of the DP in Hungarian, illustrated in (33): The possessor (Poss) is merged as a sister of of N (the possessum) for reasons of theta-role assignment (cf. Delsing 1998, de Vries 2006, Georgi and Salzmann 2011). Determiner elements like 'each' are merged on top of NP, i.e., NP is the sister of Det. In order to derive the surface order Poss > Det, the possessor must move to the specifier of a functional projection above Det. I take this to be the projection of the functional head n. n has a trigger for this movement, represented as $[\bullet D_{Poss}\bullet]$ in what follows. In addition, n assigns dative case to the possessor and thus bears a case probe [⋆dat⋆]. The nominative is not assigned to a possessor; rather, it is the default value instantiated on a possessor if it does not Agree in case with n. Finally, D has a feature that triggers movement of the possessor to SpecD, but only if it bears dative case: $[\bullet D_{dat}\bullet]$ (recall that only dative possessors occupy SpecD, nominative possessors stay in the lower Specn position, preceding Det but following D elements).

(33) *Structure of the Hungarian DP*:

$$[_{\text{DP}}\ \square\ [_{\text{D'}}\ \text{D}\ [_{\text{nP}}\ \text{Poss}\ [_{\text{n'}}\ [_{\text{NP}}\ \text{Det}\ [_{\text{N'}}\ \langle\text{Poss}\rangle\ \text{N}]]\ \text{n}\ ]]]]$$

Move Move

The following interactions between Agree and Merge arise in the DP: Movement of Poss to Specn is obligatory since all possessors precede Det. Movement to this position may bleed dative case assignment by n, resulting in default nominative case on the possessor. Opacity arises when looking at the nP: Assuming that any Poss must go through Specn, both dative and nominative possessors occupy the same position, but with different consequences for case Agree: In one case dative assignment is bled by movement to Specn; in the other case, however, dative assignment does take place although Poss is in the same position Specn, an instance of counter-bleeding.

4.3.2. Analysis

Again, the bleeding and the counter-bleeding case can be distinguished by their trigger: If the possessor ends up with nominative (dative assignment is bled), it stays in Specn; Specn is thus its final landing site. If, however, the possessor moves on to SpecD (dative possessors), Specn is only an intermediate landing site for the possessor. Intermediate movement steps are triggered by an edge feature on n, whereas final movement steps are triggered by $[\bullet \mathrm{D}_{Poss} \bullet]$ on n. The opacity effect can be derived if the final movement step applies *before* Agree and the intermediate movement step applies *after* Agree. n-Poss-Agree in ϕ-features is not influenced by the case of the possessor, there is always full agreement in ϕ-features. Since ϕ-agreement is never bled by the movement of the possessor, this suggests that it applies before any movement operation has applied. The ϕ-probe on n thus always finds the possessor as a goal in its c-command domain.[10] The order of features on n is given in (34):

(34) *Order of features on n*:
$\mathrm{n} \; \{ \; [*\phi*] \succ [\bullet \mathrm{D}_{Poss} \bullet] \succ [*\mathrm{c{:}dat}*] \succ [\bullet \mathrm{X} \bullet] \; \}$

This order has the following consequences (for the sake of clarity, ϕ-Agree is not indicated, it applies first and always results in valuation of the ϕ-probe): If the possessor ends up with nominative in Specn, its movement is driven by the $[\bullet \mathrm{D}_{Poss} \bullet]$ on n. This operation applies early, cf. (35-a). Afterwards, n triggers dative assignment, but since the proper goal, the possessor, is no longer in n's c-command domain, it cannot be assigned dative case, cf. (35-b). The result is default nominative valuation.[11]Finally, D is merged with nP. Since

---

[10] Note that this is also evidence for the separation of case- and ϕ-Agree in Hungarian since the two apply at different points in the derivation. See also Marantz (1991), Bobaljik (2008), Baker and Vinokurova (2010), Keine (2010), Preminger (2011) among others for this conclusion.

[11] One might wonder why n cannot assign dative case to the possessum which is in its c-command domain. It does not bear a case value at that point of the derivation (and is thus not inactive) since its value is determined by the external head that selects the whole DP. Note that this is a very general question: Often, the possessor case is assumed to be assigned by a head that has both the possessor and the possessum in its c-command domain. It is unclear why it targets the possessor and not the possessum, a problem that a number of approaches face. See Georgi and Salzmann (2011) for extensive discussion and a number of references to papers that encounter the same question with respect to the possessor doubling construction in German. The following solutions come to mind: (i) Case assignment is tied to XPs and not to heads; (ii) dative assignment is tied to a specific property of the possessor DP, e.g. its thematic role; (iii) case stacking: n assigns dative to N, but in addition N is assigned the case from

the possessor does not bear dative, it cannot be attracted to SpecD by the feature [$\bullet$D$_{dat}$$\bullet$] on D. This feature is deleted by default.

(35) *Nominative possessor*:

a. Step 1: EPP-movement of Poss to Specn, EPP discharged:
[$_{\text{nP}}$ Poss [$_{\text{n'}}$ n $_{[\bullet D_{Poss}\bullet]\succ[*c:dat*]}$ [$_{\text{NP}}$ Det [$_{\text{N'}}$ t N]]]]]
Move

b. Step 2: Agree initiated by n fails, default case on Poss, case-probe discharged:
[$_{\text{nP}}$ Poss [$_{\text{n'}}$ n ~~[•D•]~~ $_{\succ[*c:dat*]}$ [$_{\text{NP}}$ Det [$_{\text{N'}}$ t N]]]]]
*Agree

Dative possessors move after n initiated case Agree because this movement is triggered by an edge feature.[12] First, n probes. finds the possessor in its c-command domain and assigns it dative case, cf. (36-a). Afterwards, the possessor undergoes movement to Specn, cf. (36-b). This movement comes too late to cause bleeding since Agree applied earlier. Finally, the dative possessor is moved to SpecD triggered by [$\bullet$D$_{dat}$$\bullet$]. This step is not indicated in (36).

(36) *Dative possessor*:

a. Step 1: Agree, n assigns dative to Poss, case-probe discharged:
[$_{\text{nP}}$ n $_{[*c:dat*]\succ[\bullet X\bullet]}$ [$_{\text{NP}}$ Det [$_{\text{N'}}$ Poss N]]]
✓Agree

b. Step 2: Edge-feature-driven Internal Merge of Poss to Specn:
[$_{\text{nP}}$ Poss n ~~[*c:dat*]~~ $_{\succ[\bullet X\bullet]}$ [$_{\text{NP}}$ Det [$_{\text{N'}}$ t N]]]
Move

---

the external head that selects the DP. Only the case assigned by the external head is realized in the postsyntactic morphological component, cf. Moravcsik (1995), Bejár and Massam (1999), Merchant (2006), Assmann, Edygarova et al. (2013).

[12] As with wh-movement of the experiencer in Icelandic where wh-Exp does not check the EPP on T, the question arises why a (future) dative possessor cannot check the feature [$\bullet$D$_{Poss}$$\bullet$] on n. If it did, it would move early and bleed dative case assignment. I do not have an answer to this question. In contrast to Icelandic, no other element seems to check that feature in this configuration.

The abstract pattern of interaction is identical to the one in AAE languages: Intermediate movement steps counter-bleed Agree because they apply after Agree; final movement steps bleed Agree because they apply before Agree. The only difference between the AAE and possessor case in Hungarian is the functional head that triggers IM and Agree, as well as the morphological reflex of the Agree relation. It is head-marking (agreement) in AAE languages and dependent-marking (case) in Hungarian.

5. Cross-Linguistic Variation

In this section I look at the cross-linguistic variation that the present account predicts for the AAE, defective intervention and the encoding of possessors. I will discuss what the attested variation tells us about (i) the extrinsic / intrinsic dichotomy and (ii) the role of independent principles of the grammar that have been argued to determine rule ordering.

5.1. Variation in AAE, Intervention and Possessor Encoding

In section 3 I assumed that the order of operation-inducing features on a head H that triggers more than one operation is in prinicple free, i.e., languages can vary in the order of features on H; it is determined language-specifically. Reordering of the features thus predicts a certain range of variation. If a head triggers Agree (in ϕ-features in (37)) as well as intermediate and final movement steps, represented by [•X•] and [•F•], respectively, we expect four permutations (ignoring the order between the two types of IM triggers when they both apply before or after Agree because in that case, their order cannot be determined by the presence or absence of the morphological reflex caused by the Agree relation):

(37) *Permutations of a probe feature and the two types of IM triggers*:
   a. [•F•], [•X•] ≻ [$*\phi*$] (transparent)
   b. [$*\phi*$] ≻ [•F•], [•X•] (transparent)
   c. [•F•] ≻ [$*\phi*$] ≻ [•X•] (opaque)
   d. [•X•] ≻ [$*\phi*$] ≻ [•F•] (opaque)

In (37-a) and (37-b) internal Merge and Agree interact transparently: Both types of IM feed or bleed Agree because they apply before or after Agree. The interaction of Agree and internal Merge in (37-c) and (37-d) is opaque because one type applies before and the other after Agree; as a consequence, one type feeds or bleeds Agree and the other counter-feeds or counter-bleeds Agree. In the table in (40) I cross-classified the empirical phenomena discussed in this paper and the four ordering patterns; the cells are filled with languages that instantiate the the respective patterns.

The data presented in section 4 are all instances of the pattern in (37-c). In sections 4.1.3 and 4.2.3, several examples of the pattern in (37-a) have been discussed: The data from Trentino type languages, where both long and short Ā-movement bleeds ϕ-Agree, and from Icelandic A or Romance-type languages, in which both EPP- and wh-movement of an experiencer cause feeding, served as an argument for (i) reordering of operation-inducing features in the first place, and (ii) for SpecC and SpecT as intermediate landing sites for long Ā-movement, respectively. I take it to be accidental that I did not find an instance of this pattern for possessor case/agreement.

Another instance of this pattern with possessor extraction can be found in other Uralic languages (cf. Nikolaeva 2002) for ϕ-Agree instead of case Agree: In Nentes (Samoyedic branch of the Uralic languages), for example, it is not the case value of the possessor that varies with final or intermediate movement of the possessor but ϕ-Agree between the possessor and the possessum that shows the same sensitivity to the type of possessor extraction: There is overt agreement with so-called internal possessors but not with external possessors:

(38) *Nenets (Nikolaeva 2002)*:

a. tyuku° Wata-h ti
this Wata-GEN reindeer
'this reindeer of Wata'

b. *tyuku° Wata-h te-da
this Wata-GEN reindeer-3SG
'this reindeer of Wata'

c. Wata-GEN tyuku° te-da
Wata-GEN this reindeer-3SG
'this reindeer of Wata'

External possessors pattern with dative possessors in Hungarian in that they precede the D element and can be extracted out of the DP; internal possessors pattern with nominative possessors in Hungarian in that they follow the

D element and cannot be extracted out of the DP. The case of the possessor is constant; it always bears genitive, regardless of the IM type it undergoes. Nenets is thus very similar to Hungarian. The only difference is that in Nenets it is head-marking (agreement on the possessor) that reflects the difference between the two IM types, whereas in Hungarian is is dependent-marking (case on the possessor). The Nenets pattern is derived if the ϕ-probe and the case probe on n in Hungarian are interchanged:

(39) *Order of features on n in Nenets*:
n { [∗c:gen∗] > [•D_{Poss}•] > [∗ϕ∗] > [•X•] }

As a consequence of this order, the possessor will always be assigned genitive because it is still in the base position when the case probe on n searches for a goal. Final movement to Specn applies before ϕ-Agree and thus bleeds it; hence, internal possessors do not show agreement on the possessum. External possessors that use Specn only as an intermediate landing site move after ϕ-Agree initiated by n and thus counter-bleed ϕ-Agree – this is also an instance of the pattern in (37-c).

Another instance can be found in Selkup (Samoyedic branch of the Uralic languages). Selkup exhibits a combination of the Hungarian and the Nenets system because head-marking and dependent-marking are both sensitive to the IM type of the possessor: Internal possessors neither have an overt case marker nor do they trigger agreement; external possessors bear an overt case marker (locative) and trigger agreement. This is derived under the order in (40) where both Agree operations apply in between final and intermediate movement:

(40) *Order of features on n in Selkup*:
n { [•D_{Poss}•] > [∗ϕ∗], [∗c:loc∗] > [•X•] }

Next, consider the pattern in (37-b). There are examples of this pattern for all three phenomena discussed in this paper. With respect to AAE, there are languages where subject-verb-agreement is not bled by any type of IM. In German or English, for example, there is always full subject-verb-agreement under short and long subject Ā-extraction. As for defective intervention, Sigurðsson and Holmberg (2008) describe a third dialect of Icelandic, Icelandic C, which instantiates this abstract pattern: Neither type of experiencer movement feeds Agree between T and the lower subject DP (with respect to number agreement; on person agreement in defective intervention environments see footnote 6). We have already seen this pattern in the discussion of possessor case in Hungarian: Recall that ϕ-Agree between the possessor and the possessum, which I take to be mediated by n, is never bled by any type of IM of the possessor. This suggests that both IM types apply after ϕ-Agree (cf. (34)) with the possessor in the c-command domain of n has taken place.

Interestingly, the pattern in (40-d) does not seem to be attested for any of the three phenomena. This will be further discussed in the next subsection.

Finally, note that the attested variation sheds new light on the timing of the insertion of edge features: Chomsky (2000, 2001) suggests that edge features are inserted once a head is inactive, i.e., once it has discharged all of its operation-inducing features. The edge feature, if present on a head H, is thus the last feature that is discharged on H. In contrast, Müller (2010*a*) proposes that edge features are inserted as long as the head H is still active; edge features are immediately discharged after their insertion, the remaining operation-inducing features on H are discharged afterwards. The former proposal rules out the patterns in (37-a) and (37-d) where a head H triggers operations *after* the edge feature has been discharged; the latter proposal is at odds with the pattern in (37-c) because the edge feature must have been inserted on the head H after all other operation-inducing features on H have been discharged, i.e., at a point where H was already inactive. Since these patterns are attested, we can conclude that either languages differ in the timing of edge feature insertion (some add edge features only to inactive heads and some only to active heads) or that we need a more flexible approach to edge feature insertion that allows for early as well as late insertion. I will not choose between these options and I will also not develop the second option here; I merely point out the relevance of the attested orderings in (37) to this.

Predicted and attested patterns:

	pattern (37-a)	pattern (37-b)	pattern (37-c)	pattern (37-d)
AAE	Trentino Fiorentino Ibibo (*both IM types bleed Agree*)	German English (*neither IM type bleeds Agree*)	Celtic Berber . . . (*IM types have different effects*)	?
Interv.	Icelandic A Romance (*both IM types feed Agree*)	Icelandic C (*neither IM type feeds Agree*)	Icelandic B (*IM types have different effects*)	?
Poss	? (*both IM types bleed Agree*)	Hungarian (ϕ-Agree) (*neither IM type bleeds Agree*)	Hungarian (case) Nenets (ϕ-Agree) (*IM types have different effects*)	?

On variation in

(i) the AAE see Ouhalla (1993), Richards (1997), Ouali (2008), Phillips (2001) and references cited there.

(ii) intervention effects see Boeckx (1999), Anagnostopoulou (2003), Sigurðsson and Holmberg (2008).

(iii) possessor case/agreement see Nikolaeva (2002) and references cited there.

5.2. An Unattested Pattern (?)

The pattern in (37-d) that is predicted by the present analysis does not seem to be attested for any of the phenomena I have looked at. This might be accidental but I think it is striking that the same gap exists for three different empirical phenomena. What is more, it has been stated explicitly in the literature on the AAE that the reverse pattern has never been found, i.e., a language in which long Ā-movement of the subject bleeds subject-verb-agreement but short Ā-extraction does not (Ouhalla 1993).[13] Similarly, there exist three different Icelandic dialects, but apparently there is no dialect that exhibits the fourth pattern. I have never encountered an example of this pattern in the considerable literature on intervention effects. It remains to be seen whether this gap also exists for other phenomena. But given that it seems to be unattested for AAE and intervention, I assume that it does indeed reflect a property of the grammar and I will make a proposal as to what it might be.

I hypothesize that the absence of the fourth pattern is due to Specificity (for application of this concept to syntax see Sanders 1974, Pullum 1979, Lahne 2008*b*, van Koppen 2005) which states that more specific rules apply before less specific rules. With respect to operation-inducing features this means that the more specific IM-triggering feature is discharged before the less specific one. I suggest that non-edge features that trigger final movement steps are more specific than edge-features that trigger intermediate movement steps. The reason is that non-edge features attract elements of a certain category or with a certain effect on interpretation (e.g. [•D•], [•WH•], [•TOP•]), whereas edge features are categorially underspecified structure-building features that do not ask for a certain category or interpretative feature on the XP they attract. Hence, an order in which an edge feature is discharged before a non-edge feature, as is the case in the pattern in (37-d), is excluded by Specificity.[14]

---

[13] Ouhalla (1993: 486): "[. . .] there is apparently no language which displays the AAE in relation to long extraction but not in relation to short extraction. When there is an asymmetry, it seems to be invariably the case that the AAE holds in relation to short extraction but not in relation to long extraction."

[14] If this hypothesis is on the right track, it also predicts that external Merge to SpecH, which is standardly taken to be triggered by categorially fully specified features, must precede intermediate movement steps to SpecH. Note that this would be incompatible with the Intermediate Step Corollary presented in Müller (2010*b*). The interaction of Agree with external vs. internal Merge is beyond the scope of the present paper. I leave it for future reseach to determine whether this prediction is borne out.

This principle does not say anything about the order of probe vs. structure-building features. So their relative order is free.

5.3. Extrinsic vs. Intrinsic Ordering

A central question in the debate on rule ordering has always been whether orderings are extrinsic or intrinsic. An order is extrinsic if it must be stipulated in a language-specific fashion, i.e., two languages can differ solely in the order of rules. An order is intrinsic if it is determined by independent principles of the grammar (Chomsky 1965: 223, fn.6).[15] From a conceptual point of view, intrinsic orderings are to be preferred, but the question is whether all orderings are predicted by meta-principles of the grammar. Indeed, Pullum (1979) claims that all orderings are determined by Universal Principles, i.e., extrinsic ordering is not necessary. The major principles that Pullum argues to determine rule ordering in syntax (but also in other components of the grammar) are the following:

(i) Obligatory Precedence Principle (or Immediate Characterization, cf. Ringen 1972, Perlmutter and Soames 1979)
(ii) Specificity Principle (or Elsewhere Condition, cf. Sanders 1974, Anderson 1969 among many others)
(iii) Cyclic Principle (or the Cycle, cf. Chomsky et al. 1956, Chomsky and Halle 1968)

The Obligatory Precedence Principle states that obligatory rules apply before optional rules. This principle is not easily applicable in Minimalism because in this framework, operation-inducing features *must* be discharged if the context for the operation is met (cf. Pesetsky 1989, Pesetsky and Torrego 2001, Preminger 2011). It is only the presence of the feature on a head that can be optional, but if it is present, the operation it triggers must obligatorily apply. I will thus disregard this principle in what follows.

---

[15] Pullum (1979: ch.1) criticizes the definition of the terms 'extrinsic' and 'intrinsic' as introduced in Chomsky (1965). He thus replaces them by the terms 'parochial' for extrinsic and 'universal' for intrinsic orderings with slightly different defining properties, in particular concerning the question whether languages can differ solely in their order of rules. In what follows, I will stick to the original terminology because it is more widely used and, as Pullum acknowledges, because these terms have by and large been used in the way that he defines parochial and universal orderings, and I will do so as well.

The Specificity Principle demands that more specific operations apply before less specific operations. The consequences of this principle depend on what 'specific' means. It can, for example, refer to the context of a (transformational) rule: A rule R_1 is more specific than a rule R_2 if the context of R_1 includes the context of R_2. I have argued in section 5.2 that this principle may account for the absence of the order in (37-d).

Finally, the Cyclic Principle demands that rules in a lower cyclic domain apply before all rules in a higher cyclic domain:

(41) *Cyclic Principle*:
Any operation to the cyclic domain D_x will precede any operation to the cyclic domain D_{x-1}.

Traditionally, S and NP (= CP and DP in current terminology) are taken to constitute cyclic domains. Note that the Cyclic Principle does not have anything to say about the order of operations that apply within the same cycle. Since there are a lot of instances of rules that apply within a single clause and hence, within the same domain S, McCawley (1984, 1988) proposes to increase the number of cyclic domains and thereby to maximize the applicability as well as the predictive power of the Cyclic Principle. McCawley assumes that every constituent, (i.e. every projection in an incremental model of syntax) is a cyclic domain.[16] As he points out, it is very unlikely under this assumptions that two rules or operations apply within the same cycle. And indeed, he illustrates that many (if not all) of the orderings that needed to be stipulated under the original version of the Cyclic Principle are predicted by his stronger version of the principle. For applications of each of the three principles see Pullum (1979: ch.1).

With these principles in mind, we can now look at the variation in the order of probe features and the two types of IM-triggers summarized in (40). I claim that none of the principles introduced above can account for the full range of variation that is attested for the AAE, defective intervention effects and the encoding of possessors by case and/or agreement. The reason is that the patterns in (37-a) and (37-b) are the exact opposite of one another. It is impossible that

---

[16] A similar though not identical proposal is made in Williams (1974). If the Williams cycle is applied to the inclusion hierarchy that Williams adopts, this yields the same effects as McCawley's proposal as long as the rules or operations apply within the same clause; for cross-clausal interactions, the Williams cycle and McCawley's version of the Cyclic Principle make different predictions.

any principle predicts one order of features as well as the opposite. Note that this point is independent of the existence of the pattern in (37-d); even if only 3 out of the four predicted patterns are attested, no principle predicts all three of them.

Let me go into some more details. As indicated above, the Obligatory Precedence Principle is not applicable at all in a minimalist framework. As for the Specificity Principle, it may be able to distinguish between the two types of IM triggers in terms of specificity and thereby account for the absence of a pattern, but I do not see how it could distinguish between probe features and structure-building features in this respect. A probe and a structure-building feature (which is not an edge feature) do not seem to be of a different degree of specificity in any sense of this term. And even if one would find a property relative to which the two features have a different degree of specificity, the Specificity Principle would predict an order of the two features, but at the same time it would exclude the reverse order. As pointed out above, this is problematic given that both the patterns in (37-a) and (37-b) are attested.

The Cyclic Principle, which is the principle that has been said to be predict most of the rule orderings in syntax, does also not predict all of the attested patterns in (40). Obviously, in its original form with only S and NP nodes being cyclic domains, it cannot determine the orderings of operation-inducing features because all of the operations triggered by a single head H apply in the same S (CP) and NP (DP) cycle, respectively. Crucially, this also holds under McCawley's stricter version with every projection being a cyclic domain. Note that the cases in which a single head triggers more than one operation are exactly of the type that McCawley said would be rare, if not inexistent. Under one interpretation of McCawley's version of the principle, they do apply in the same projection. However, there are indeed two possible interpretations of McCawley's Cyclic Principle: Agree and Merge apply (i) either in the same cycle (H') or (ii) Agree applies in a lower cycle than IM (H' vs. HP).[17] When looking at the point of the derivation where a head H that triggers Agree and IM is merged, the H' projection is created and any of the operations that is triggered by H applies within this projection, i.e., they apply in the same cycle, cf. interpretation (i). However, one could also argue that Agree and IM apply in different cycles, cf. interpretation (ii): Since Agree applies under c-command,

---

[17] I thank Gereon Müller for pointing out to me this ambiguity in the interpretation of McCawley's version of the Cyclic Principle.

it will only affect the H′ constituent, whereas IM, when looking at the output of Merge, led to the projection of the HP and could thus be assumed to apply within HP. But under neither of these two interpretations does the stricter version of the Cyclic Principle predict the full range of attested variation: If interpretation (i) is assumed, the Cyclic Principle does not make any predictions because the operations apply in the same cyclic domain; thus, the principle is too weak. If one adopts interpretation (ii), Agree is wrongly predicted to always apply before IM, but the reverse order is also attested, cf. (40); in this second case, the principle is too strong. Thus, the Cyclic Principle also does not predict the full range of orders of operation-inducing features, not even under the stronger version with every projection being a cyclic domain.

The conclusion is that none of three principles allows for all the attested orderings. Some principles do not make any predictions at all, whereas others prohibit orders that are attested. Hence, extrinsic language-specific (parochial) ordering of the operation-inducing features is needed after all.

6. Conclusion

In this paper I have argued for a more fine-grained approach to the order of elementary operations: Merge not only needs to be ordered with respect to Agree; rather, one must distinguish between (at least) two different types of (internal) Merge because one type applies before and the other after Agree. Evidence for this split of internal Merge into two types comes from the observation that in some languages, if a head H triggers IM and Agree, IM has different consequences for Agree initiated by H, although on the surface the internally merged XP always occupies the same structural position SpecH. The behaviour of an internally merged XP with respect to its effect on the Agree relation can be predicted by the nature of the landing site: If SpecH is a final landing site for XP in a movement chain, XP movement feeds or bleeds Agree; if, however, SpecH is an intermediate landing site for XP, XP movement has the opposite effect, i.e., it counter-feeds or counter-bleeds Agree by H. If intermediate and final landing sites have different kinds of triggers, the opacity can be derived by a different ordering of the two IM triggers relative to Agree: Intermediate movement steps are triggered by edge features and apply after Agree; final movement steps are triggered by non-edge features and apply before Agree. The pattern that arises from this ordering is that edge-feature-driven IM of XP has the same effect on the Agree relation by H as if XP is not moved at all.

The analysis has the following general implications: (i) It presupposes a strictly derivational model of syntax. (ii) The variation predicted through reordering of operation-inducing features is indeed attested, except for one gap. I have proposed that this gap may be due to Specificity. (iii) The attested orderings provide evidence for the need of extrinsic ordering: None of the principles that have been proposed to determine the order of rules (or: operations) can predict the full range of variation. Most of the principles are too strong, disallowing certain attested orderings; the strict version of the Cyclic Principle proposed by McCawley (1988) is arguably too weak, making no predictions at all. (iv) Variation in defective intervention effects provides evidence that movement to SpecC goes through SpecT.

References

Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Connecticut, Storrs, Connecticut.

Abels, Klaus (2004): Right Node Raising: Ellipsis or Across the Board Movement? *In:* K. Moulton and M. Wolf, eds, *Proceedings of NELS 34*. GLSA, Amherst, pp. 45–60.

Abels, Klaus (2012): *Phases: An Essay on Cyclicity in Syntax*. Mouton de Gruyter, Berlin.

Alexiadou, Artemis, Tibor Kiss and Gereon Müller (2012): *Local Modelling of Non-local Dependencies in Syntax*. Linguistische Arbeiten 547, de Gruyter, Berlin/New York.

Anagnostopoulou, Elena (2003): *The Syntax of Ditransitives: Evidence from Clitics*. Mouton de Gruyter, Berlin, New York.

Anderson, Stephen R. (1969): West Scandinavian Vowel Systems and the Ordering of Phonological Rules. PhD thesis, MIT, Cambridge, Massachusetts.

Anderson, Stephen R. (1974): *The Organization of Phonology*. Academic Press, New York.

Arregi, Karlos and Andrew Nevins (2012): *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Springer, Dordrecht.

Assmann, Anke, Svetlana Edygarova, Doreen Georgi, Timo Klein and Philipp Weisser (2013): Possessor Case in Udmurt: Multiple Case Assignment Feeds Post-syntactic Fusion. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 23-63.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller and Philipp Weisser (2013): Ergatives Move too Early: On an Instance of Opacity in Syntax. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 363-412.

Assmann, Anke and Fabian Heck (2013): Opaque Intervention. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 463-509.

Baker, Mark (2008): 'On the Nature of the Anti-Agreement Effect: Evidence from Wh-in-Situ in Ibibio', *Linguistic Inquiry* 39, 615–632.

Baker, Mark and Nadya Vinokurova (2010): 'Two Modalities of Case Assignment: Case in Sakha', *Natural Language and Linguistic Theory* 28, 593–642.

Baković, Eric (2011): Opacity and Ordering. *In:* J. Goldsmith, J. Riggle and A. Yu, eds, *The Handbook of Phonological Theory*. Wiley-Blackwell, Oxford, chapter 2.

Barss, Andrew (1986): Chains and Anaphoric Dependence. PhD thesis, MIT, Cambridge, Mass.

Bejár, Susana and Diane Massam (1999): 'Multiple Case Checking', *Syntax* 2(2), 65–79.

Bobaljik, Jonathan (2008): Where's Phi? Agreement as a Post-Syntactic Operation. *In:* D. Harbour, D. Adger and S. Béjar, eds, *Phi-Theory: Phi Features Across Interfaces and Modules*. Oxford University Press, Oxford, pp. 295–328.

Boeckx, Cedric (1999): 'Conflicting C-Command Requirements', *Studia Linguistica* 53(3), 227–250.

Brandi, Luciana and Patrizia Cordin (1998): Two Italian Dialects and the Null Subject Parameter. *In:* O. Jaeggli and K. Safir, eds, *The Null Subject Parameter*. Kluwer Academic Publishers, Dordrecht, pp. 111–142.

Brody, Michael (2002): On the Status of Representations and Derivations. *In:* S. Epstein and D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 19–41.

Béjar, Susana (2003): Phi-Syntax: A Theory of Agreement. PhD thesis, University of Toronto, Toronto.

Béjar, Susana and Milan Řezáč (2009): 'Cyclic Agree', *Linguistic Inquiry* 40(1), 35–73.

Campos, Héctor (1997): 'On Subject Extraction and the Antiagreement Effect in Romance', *Linguistic Inquiry* 28(1), 92–119.

Cheng, Lisa Lai-Shen (2006): 'Decomposing Bantu Relatives', *Proceedings of NELS* 36, 197–215.

Chomsky, Noam (1951): Morphophonemics of Modern Hebrew. Master's thesis, University of Pennsylvania.

Chomsky, Noam (1957): *Syntactic Structures*. Mouton, The Hague and Paris.

Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* M. Roger, M. David and 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, Massachusetts, pp. 1–52.

Chomsky, Noam (2004): Beyond Explanatory Adequacy. *In:* A. Belletti, ed., *Structures and Beyond. The Cartography of Syntactic Structures Vol. 3*. Oxford University Press, Oxford.

Chomsky, Noam (2005): 'Three Factors in Language Design', *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam and Morris Halle (1968): *The Sound Pattern of English*. Harper & Row, New York.

Chomsky, Noam, Morris Halle and Fred Lukoff (1956): On Accent and Juncture in English. *In:* M. Halle, H. Lunt, H. McLean and C. van Schooneveld, eds, *For Roman Jakobson. Essays on the occasion of his sixtieth birthday*. Mouton, The Hague, pp. 65–80.

Chung, Sandra (1994): 'Wh-Agreement and 'Referentiality' in Chamorro', *Linguistic Inquiry* 25, 1–44.

Clements, George, James McCloskey, Joan Maling and Annie Zaenen (1983): 'String-Vacuous Rule Application', *Linguistic Inquiry* 14, 1–17.

Cole, Peter (1982): 'Subjacency and Successive Cyclicity: Evidence from Ancash Quechua', *Journal of Linguistic Research* 2, 35–58.

Delsing, Lars-Olof (1998): Possession in Germanic. *In:* A. Alexiadou and C. Wilder, eds, *Possessors, Predicates and Movement in the Determiner Phrase*. Benjamins, Amsterdam, pp. 87–109.

Embick, David (2010): *Localism versus Globalism in Morphology and Phonology*. MIT Press, Cambridge, Mass.

Embick, David and Rolf Noyer (2001): 'Movement Operations after Syntax', *Linguistic Inquiry* 32, 555–595.

Epstein, Samuel David and T. Daniel Seely (2002): Rule Applications as Cycles in a Level-Free Syntax. *In:* S. Epstein and T. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65–89.

Fox, Danny (2000): *Economy and Semantic Interpretation*. MIT Press, Cambridge, Mass.

Frampton, John (2004): Copies, Traces, Occurrences, and All That. Ms., Northeastern University.

Gallego, Ángel J. (2006): Phase Sliding. Ms., UAB/UMD.

Georgi, Doreen and Martin Salzmann (2011): 'DP-Internal Double Agreement is not

Double Agree: Consequences of Agree-Based Case Assignment within DP', *Lingua* 121, 2069–2088.

Grohmann, Kleanthes K. (2003): 'Successive Cyclicity under (Anti-)Local Considerations', *Syntax* 6, 260–312.

Halpert, Claire (2012): Argument Licensing and Agreement in Zulu. PhD thesis, MIT, Cambridge, Mass.

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

Heck, Fabian and Gereon Müller (2003): 'Derivational Optimization of Wh-Movement', *Linguistic Analysis* 33, 97–148. (Volume appeared 2007).

Heck, Fabian and Gereon Müller (2007): Extremely Local Optimization. *In:* E. Brainbridge and B. Agbayani, eds, *Proceedings of the 26th WECOL*. California State University, Fresno, pp. 170–183.

Henderson, Brent (2009): 'Anti-Agreement and [person] in Bantu', *Selected Proceedings of the 38th Annual Conference on African Linguistics* 38, 173–181.

Holmberg, Anders (2000): 'Scandinavian Stylistic Fronting: How any Category can Become an Expletive', *Linguistic Inquiry* 31, 445–484.

Holmberg, Anders and Thorbjörg Hróarsdóttir (2003): 'Agreement and Movement in Icelandic Raising Constructions', *Lingua* 113(10), 997–1019.

Jouitteau, Mélanie (2005): La syntaxe comparée du Breton. PhD thesis, University of Nantes, Nantes.

Kayne, Richard S. (1975): *French Syntax: The Transformational Cycle*. MIT Press, Cambridge, Mass.

Keine, Stefan (2010): *Case and Agreement from Fringe to Core: A Minimalist Approach*. Linguistische Arbeiten 536, de Gruyter, Berlin.

Kenstowicz, Michael and Charles Kisseberth (1977): *Topics in Phonological Theory*. Academic Press, NY.

Kenstowicz, Michael and Charles Kisseberth (1979): *Generative Phonology*. Academic Press, New York.

Kiparsky, Paul (1968): Linguistic Universals and Linguistic Change. *In:* R. Bach, E. Harms, ed., *Universals in Linguistic Theory*. Holt, New York, pp. 170–210.

Kiparsky, Paul (1971): Historical Linguistics. *In:* W. Dingwall, ed., *A Survey of Linguistic Science*. University of Maryland Linguistics Program, College Park, pp. 576–642.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Koppen, Marjo van (2005): One Probe – Two Goals: Aspects of Agreement in Dutch Dialects. PhD thesis, Leiden University., Leiden. LOT-publications number 105.

Koutsoudas, Andreas, Gerald Sanders and Craig Noll (1974): 'The Application of Phonological Rules', *Language* 50, 1–28.

Lahne, Antje (2008*a*): Excluding SVO in Ergative Languages: A New View on Mahajan's Generalisation. *In:* F. Heck, G. Müller and J. Trommer, eds, *Varieties of Competition*. Universität Leipzig, Institut für Linguistik, pp. 65–80. Vol. 87 of *Linguistische Arbeitsberichte*.

Lahne, Antje (2008*b*): Where There is Fire There is Smoke: Local Modelling of Successive-Cyclic Movement. PhD thesis, University of Leipzig.

Lasnik, Howard (2001): Derivation and Representation in Modern Transformational Syntax. *In:* M. Baltin and C. Collins, eds, *The Handbook of Contemporary Syntactic Theory*. Blackwell, Malden, Mass., pp. 62–88.

Lebeaux, David (1990): Relative Clauses, Licensing, and the Nature of the Derivation. *In:* J. Carter, R.-M. Déchaine, B. Philip and T. Sherer, eds, *Proceedings of NELS 20*. G.L.S.A., University of Massachusetts, Amherst, pp. 318–332.

Legate, Julie Ann (2002): Phases in "Beyond Explanatory Adequacy". Ms., MIT, Cambridge.

Marantz, Alec (1991): Case and Licensing. *In:* G. Westphal, B. Ao and H.-R. Chae, eds, *Proceedings of the Eighth Eastern States Conference on Linguistics*. University of Maryland, pp. 234–253.

McCawley, James D. (1984): Exploitation of the Cyclic Principle as a Research Strategy in Syntax. *In:* W. de Geest and Y. Putseys, eds, *Sentential Complementation*. Foris, Dordrecht, pp. 165–183.

McCawley, James D. (1988): *The Syntactic Phenomena of English*. Vol. 1, University of Chicago Press, Chicago.

McCloskey, James (1979): *Transformational Syntax and Model Theoretic Semantics. A Case Study in Modern Irish*. Reidel, Dordrecht.

McGinnis, Martha (2001): Variation in the Phase Structure of Applicatives. *In:* P. Pica and J. Rooryck, eds, *Linguistic Variation Yearbook 1*. John Benjamins, Amsterdam, pp. 101–142.

McGinnis, Martha J. (1998): Locality in A-Movement. PhD thesis, MIT, Cambridge, Mass.

Merchant, Jason (2006): Polyvalent Case, Geometric Hierarchies, and Split Ergativity. *In:* J. Bunting, S. Desai, R. Peachey, C. Straughn and Z. Tomkova, eds, *Proceedings of the 42nd Annual Meeting of the Chicago Linguistics Society*. Chicago Linguistics Society, Chicago, pp. 47–67.

Moravcsik, Edith A. (1995): Summing up Suffixaufnahme. *In:* F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford University Press, New York, pp. 451–484.

Müller, Gereon (2004): Phrase Impenetrability and Wh-Intervention. *In:* A.

Stepanov, G. Fanselow and R. Vogel, eds, *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 289–325.

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. *In:* K. Grohmann, ed., *Explorations of Phase Theory. Features and Arguments*. Mouton de Gruyter, Berlin, pp. 269–308.

Müller, Gereon (2010*a*): 'On Deriving CED Effects from the PIC', *Linguistic Inquiry* 41, 35–82.

Müller, Gereon (2010*b*): Operator Islands, Maraudage, and the Intermediate Step Corollary. Ms., University of Leipzig. To appear in Balázs Surányi (ed.), *Minimalist Approaches to Syntactic Loality*.

Müller, Gereon (2011): *Constraints on Displacement: A Phase-Based Approach*. John Benjamins, Amsterdam.

Nikolaeva, Irina (2002): The Hungarian External Possessor in a European Perspective. *In:* C. Hasselblatt and R. Blokland, eds, *Finno-Ugrians and Indo-Europeans: Linguistic and Literary Contacts*. Shaker, Maastricht, pp. 272–285.

Ouali, Hamid (2008): On C-to-T-Feature Transfer: The Nature of Agreement and Anti-Agreement in Berber. *In:* R. D'Alessandro, S. Fischer and G. Hrafnbjargarson, eds, *Agreement Restrictions*. de Gruyter, pp. 159–180.

Ouhalla, Jamal (1993): 'Subject Extraction, Negation and the Anti-agreement Effect', *Natural Language and Linguistic Theory* 11, 477–518.

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

Pesetsky, David (1989): 'Language Particular Processes and the Earliness Principle', Ms.

Pesetsky, David and Esther Torrego (2001): T to C Movement: Causes and Consequences. *In:* M. Kenstowicz, ed., *Ken Hale: A Life in Language*. MIT Press, Cambridge, Massachusetts, pp. 355–426.

Phillips, Colin (2001): Disagreement between Adults and Children. *In:* A. Mendikoetxea and M. Uribe-Etxebarria, eds, *Theoretical Issues on the Morphology-Syntax Interface*. ASJU, San Sebastian, pp. 359–394.

Preminger, Omer (2011): Agreement as a Fallible Operation. PhD thesis, Massachusetts Institute of Technology, Cambridge, Mass.

Pullum, Geoffrey K. (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York & London. (PhD thesis from 1976, University of London).

Řezač, Milan (2004): Elements of Cyclic Agree: Agree and Merge. PhD thesis, University of Toronto.

Richards, Marc D. (2007): 'On Feature-Inheritance: an Argument from the Phase Impenetrability Condition', *Linguistic Inquiry* 38, 563–572.

Richards, Norvin (1997): What Moves Where When in Which Language? PhD thesis, MIT, Cambridge, Mass.

Ringen, Catherine (1972): 'On Arguments for Rule Ordering', *Foundations of Language* 8, 266–273.

Rizzi, Luigi (2004): On the Form of Chains: Criterial Positions and ECP Effects. Ms., University of Siena.

Rizzi, Luigi (2007): 'On Some Properties of Criterial Freezing', *CISCL Working Papers on Language and Cognition. STIL* 1.

Rizzi, Luigi and Ur Shlonsky (2007): Strategies for Subject Extraction. *In:* U. Sauerland and H.-M. Gärtner, eds, *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Mouton de Gruyter, Berlin, pp. 115–160.

Ross, John R. (1967): Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, Mass.

Sanders, Gerald (1974): 'Precedence Relations in Language', *Foundations of Language* 11, 361–400.

Schneider-Zioga, Patricia (2007): 'Anti-Agreement, Anti-Locality and Minimality. The Syntax of Dislocated Subjects', *Natural Language and Linguistic Theory* 25, 403–446.

Sigurðsson, Halldór Ámann and Anders Holmberg (2008): Icelandic Dative Intervention: Person and Number are Separate Probes. *In:* R. d'Alessandro, G. H. Hrafnbjargarson and S. Fischer, eds, *Agreement Restrictions*. Mouton de Gruyter, Berlin, pp. 251–280.

Sproat, Richard (1985): 'Welsh Syntax and VSO Structure', *Natural Language and Linguistic Theory* 3, 173–216.

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

Sternefeld, Wolfgang (2006): *Syntax*. Stauffenburg, Tübingen. Two volumes.

Szabolcsi, Anna (1994): The Noun Phrase. *In:* F. Kiefer, ed., *Syntax and Semantics, vol. 27: The Structure of Hungarian*. Academic Press, San Diego, pp. 179–274.

Torrego, Esther (1984): 'On Inversion in Spanish and Some of Its Effects', *Linguistic Inquiry* 15, 103–129.

Unger, Christina (2010): A Computational Approach to the Syntax of Displacement and the Semantics of Scope. PhD thesis, Universiteit Utrecht.

Vries, Marc de (2006): 'Possessive Relatives and (Heavy) Pied Piping', *The Journal of Comparative Germanic Linguistics* 9, 1–52.

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

Opaque Intervention

Anke Assmann & Fabian Heck[*]

Abstract

Arguments differ in their capacities to function as antecedents for certain associates (e.g., floating quantifiers and parasitic gaps in German). These differences cannot always be read off the argument's surface position but are sometimes opaque. We argue that such cases of opaque syntactic intervention can be derived in a strictly derivational fashion. Concretely, we assume that the capacities of an argument to function as an antecedent can be traced back to an earlier stage of the derivation where the antecedent-associate relation is established in a transparent manner. Later stages of the derivation may alter the configuration but not the associations established before. Within this derivational approach, there is no need for representational devices such as copies/traces or constraints on representations like the MLC.

1. Introduction and Overview

The purpose of this paper is to account for cases of opaque intervention[1] as they arise with binding of parasitic gaps (PGs) and association with floating quantifiers (FQs) in German. In what follows, we subsume FQs and PGs under the term "associate"; the argument that establishes a relation with the associate is called its "antecedent". Our central claim is that an argument's capacity to function as the antecedent of an associate cannot always be read off the argument's surface position but is often opaque. To begin with, one can observe that the relations under discussion (FQ-association, PG-binding) are

---

[*] For helpful comments and questions we would like to thank the audiences at Generative Grammatik im Süden (Universität Stuttgart, May 2010) and at the GLOW workshop 'Empty Categories in Syntax' (Universität Potsdam, March 2012) as well as the participants of the Colloquium on Syntax and Morphology at the Universität of Leipzig. Particular thanks go to Klaus Abels, Petr Biskup, Sandra Döring, Gisbert Fanselow, Eric Fuss, Doreen Georgi, Stefan Keine, Ivona Kučerová, Denisa Lenertová, Gereon Müller, Martin Salzmann, Florian Schäfer, Radek Šimík, Volker Struckmeier, Philipp Weisser, and Malte Zimmermann.

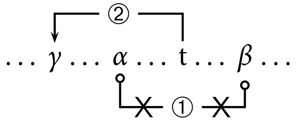
[1] For now, the notion of intervention can be understood in a pre-theoretical sense; see sections 4 and 5 for qualification.

Rule Interaction in Grammar, 463–509
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

subject to an intervention restriction: An antecedent α cannot associate with β if there is another potential antecedent γ that intervenes on the surface between α and β. Crucially, however, sometimes association between α and β is impossible although no γ intervenes. And sometimes, there is an intervening γ, and yet association between α and β is not disrupted. These cases, which are at variance with the surface intervention effect observed elsewhere, are, we claim, to be analyzed as instances of opacity. Before moving on, we would like to briefly recall the notion of opacity familiar from generative phonology (Chomsky 1951, Kiparsky 1976, Kenstowicz and Kisseberth 1979).

In derivational phonology, there are two types of opaque interaction between two rules R_1 and R_2: counter-feeding and counterbleeding. In the first type, application of R_1 provides the application context for R_2 but the surface form ψ indicates that R_2 has not applied; thus, application of R_1 does not feed application of R_2, which is usually explained by assuming that R_2 applies before R_1. As a consequence, the application context for R_2 created by R_1 comes about too late to feed R_2. Similarly, in the second type, application of R_1 destroys the application context for R_2, yet ψ suggests that R_2 has applied nevertheless; thus application of R_1 does not bleed application of R_2. Again, the explanation consists of assuming that R_2 applies before R_1, thus R_1 does not have a chance to destroy the application context for R_2. In the first case, ψ does not transparently indicate why R_2 has not applied; in the second case, ψ does not show in a transparent manner why R_2 has in fact applied. In this sense, ψ (or the rule interaction deriving it) is opaque.

Syntactic association that does not obey surface intervention can be understood in terms of opacity, too. Concretely, if association between α and β is impossible although there is no intervening γ on the surface, then, we claim, this is because there is an earlier derivational stage where γ does intervene between α and β (blocking association between them) but later moves away and thus no longer intervenes on the surface (counter-feeding). This is schematically illustrated in (1), with the order of steps indicated by the number labels.[2]

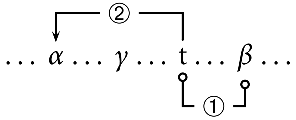
(1)


---

[2] Here and in what follows, traces are indicated for expository reasons only. Thus, we will not take a firm stand on whether movement leaves a copy behind, a trace, or nothing at all.

If association between α and β is possible although some γ intervenes on the surface, then there must be a previous stage of the derivation where α occupies a position such that γ does not intervene between α and β (allowing association between them). In a later step, α moves across γ, thereby giving the surface impression of intervention (counter-bleeding), see (2).

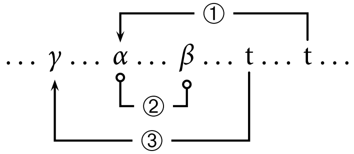
(2)



As noted before, the relation between an antecedent and its associate is not always opaque. There are also well-behaved cases of association without intervention (feeding) and cases where association is blocked by surface intervention (bleeding). The question then is what determines whether an antecedent relates (or does not relate) to an associate in an opaque or in a transparent way. It turns out that there are asymmetries between arguments with respect to whether they can act as antecedent for an associate in the presence of a particular co-argument (leading to feeding and counter-bleeding) or whether they cannot (resulting in bleeding and counter-feeding). Descriptively, of those arguments that (ultimately) precede the associate, it is always the one that has been merged *lowest* that becomes its antecedent.

The analysis that we propose follows the same general pattern for both types of association. Simplifying somewhat, we propose that the associate is merged to a fixed position to the left of its potential antecedents. Association requires that the antecedent must move to the left of the associate (to c-command it). The surface true associations (involving feeding or bleeding) can then be traced back (a) to the order in which the arguments (the potential antecedents α, γ etc.) are merged and (b) to the idea that their relative order is preserved if they move together across the associate. As a consequence, (a) and (b) interact to the effect that of the arguments that move across the associate β the one that has been merged lowest moves first (see α in (3)). It then immediately establishes association and thus becomes the antecedent, blocking association of any other argument that moves later (such as γ in (3)).

(3)



The non-surface true associations (involving counter-feeding and counter-bleeding) are accounted for by further movement of the antecedent, which changes the relative order of the potential antecedents, thereby rendering feeding and bleeding opaque. These are the movements indicated in (1) and (2).

A desideratum of a derivational account of opacity is that the order of rules is intrinsic, that is, it follows from independent properties of the system. It will be shown that the rule order involved in the present study is (almost) completely determined by independently motivated principles of the derivational architecture (such as strict cyclicity or the requirement that feature checking applies as early as possible).

We proceed as follows. In section 2, we illustrate the opacity effects that arise with FQ-association (2.1) and PG-binding (2.2). We make explicit the theoretical background (section 3) and then show how the opacity effects observed in section 2 can be made to follow for both empirical domains under investigation (sections 4.1–4.2). Section 5.1 discusses some complications that arise with *wh*-in-situ. In section 5.2, we explain why the findings presented in the present study are at variance with certain diagnostics that have been argued to distinguish verb classes (in German) with respect to the underlying word orders they project. Section 5.3 illustrates that the findings provide an argument against the idea that scrambling is to be derived by base generation. Finally, we show that the analysis suggests that the theory of tucking-in is not well-suited to account for order preservation effects in a strictly derivational framework (section 5.4). Section 6 provides a conclusion.

2. Observations

2.1. Floating Quantifiers

German possesses a FQ *alles* "all", which obligatorily associates with a *wh*-phrase (Pafel 1991, Reis 1992).[3] This is illustrated in (4-a-c), where association is represented by co-indexation. In (4-a), *alles* associates with a subject *wh*-phrase (marked by nominative case), in (4-b,c) the associates of *alles* are a direct and an indirect object (marked by accusative and dative), respectively.

---

[3] Giusti (1991) argues that *alles* combining with *wh*-phrases in German must be distinguished from other uses of *alles*.

(4) a. Wer$_2$ hat euch alles$_2$ geholfen?
who$_{nom}$ has you all helped
“Who all helped you?”
b. Wen$_2$ habt ihr alles$_2$ kennengelernt?
who$_{acc}$ have you all met
“Who all did you meet?”
c. Wem$_2$ habt ihr alles$_2$ geholfen?
who$_{dat}$ have you all helped
“Who all did you help?”

The grammatical function of the associate has no impact on the morphology of *alles*. Thus, *alles* is morphologically invariant. In this respect, it differs from its kin, the FQ *all(es)* “all” (and also from the FQ *beide* “both”) in German. Syntactically and semantically, *alles* and *all(es)* must also be distinguished (see in particular Reis 1992). In what follows, we focus on invariant *alles*.[4]

If a subject *wh*-phrase associates with floating *alles*, then no indefinite object, be it indirect or direct, may intervene between the *wh*-phrase and the FQ, see (5-a,b). Also, an indirect object *wh*-phrase cannot be separated from *alles* associated with it by an indefinite direct object (5-c).

(5) a. *Wer$_1$ hat einem Professor alles$_1$ gratuliert?
who$_{nom}$ has a professor$_{dat}$ all congratulated
“Who all congratulated a professor?”
b. *Wer$_1$ hat einen Professor alles$_1$ vergöttert?
who$_{nom}$ has a professor$_{acc}$ all idolized
“Who all idolized a professor?”
c. *Wem$_1$ hat sie einen Professor alles$_1$ vorgestellt?
who$_{dat}$ has she a professor$_{acc}$ all introduced
“Who all did she introduce a professor to?”

It thus seems as if an indefinite noun phrase that intervenes at the surface between a *wh*-phrase and a FQ associated with the latter disrupts the association between the two.

However, if the *wh*-phrase that associates with floating *alles* is a direct or an indirect object, then an indefinite subject *can* intervene between the *wh*-

---

[4] Discussions of variant *all(es)* (sometimes with reference to scrambling) can be found in Link (1974), Reis and Vater (1980), Giusti (1990), Haider (1993, 214-215), Merchant (1996). Floating *beide* is addressed in Reis and Vater (1980).

phrase and the FQ, see (6-a,b). Similarly, a direct object *wh*-phrase can be separated from a FQ associated with it by an indefinite indirect object (6-c).

(6) a. Wem$_1$ hat ein Professor alles$_1$ geholfen?
who$_{dat}$ has a professor$_{nom}$ all helped
"Who all did a professor help?"
b. Wen$_1$ hat ein Professor alles$_1$ beleidigt?
who$_{acc}$ has a professor$_{nom}$ all insulted
"Who all did a professor insult?"
c. Wen$_1$ hat sie einem Professor alles$_1$ vorgestellt?
who$_{acc}$ has she a professor$_{dat}$ all introduced
"Who all did she introduce to a professor?"

Against the background of the intervention facts in (5), this state of affairs can be interpreted as an instance of counter-bleeding: although the surface intervention of the indefinite is expected to bleed association of the *wh*-phrase with the FQ (as it does in (5)), no such intervention effect occurs in (6).[5]

The generalization emerging from the asymmetries between (5) and (6) can be stated as follows.

(7) *Intervention Asymmetry for FQs:*
A *wh*-phrase α and a FQ β can associate in the presence of an (indefinite) co-argument γ that precedes β if and only if γ is higher on the hierarchy $nom > dat > acc$ than α.

Finally, it can be observed that the intervention asymmetries vanish if the intervening argument is definite (and not indefinite): a definite argument can

---

[5] Beck (1997, 41-46) is the only discussion of intervention effects with indefinites and invariant *alles* we are aware of. There, it is proposed that indefinites interpreted as narrow scope existentials block LF-movement of *alles* (assumed to be required for semantic interpretation, cf. footnote 26) while those interpreted as generic or specific (possibly existentials with wide scope) do not. Against this background, the examples in (6) should necessarily involve a generic or specific reading of the indefinites, which they do not in our view; the examples in (5) should become grammatical if a generic or specific reading of the indefinites is forced (in which case we could analyze them as lacking [INDEF], cf. section 4.1). Interestingly, all of Beck's (1997) grammatical examples that exhibit surface intervention and that, according to Beck (1997), involve a generic or specific reading of the indefinite instantiate the type of configuration that is analyzed as counter-bleeding in the present approach. Finally, note that the more recent theory of intervention effects proposed in Beck (2006) no longer covers invariant floating *alles*, thus calling for an alternative account anyway.

freely intervene between a *wh*-phrase and a FQ associated with it, no matter what the grammatical functions of the *wh*-phrase and the intervener are. Obviously, definite noun phrases are not the right type of element to bleed association between a *wh*-phrase and a FQ. This is illustrated in (8), which provides minimal pairs with respect to (5).

(8) a. Wer$_1$ hat dem Professor alles$_1$ gratuliert?
who$_{nom}$ has the professor$_{dat}$ all congratulated
"Who all congratulated the professor?"
b. Wer$_1$ hat den Professor alles$_1$ vergöttert?
who$_{nom}$ has the professor$_{acc}$ all idolized
"Who all idolized the professor?"
c. Wem hat sie den Professor alles vorgestellt?
who$_{dat}$ has she the professor$_{acc}$ all introduced
"Who all did she introduce the professor to?"

In what follows, we turn to asymmetries that arise with PG-binding in German. It will turn out that these are very similar to the asymmetries discussed in the context of FQ-association.

2.2. Parasitic Gaps

Generally, PGs must be bound by an element in A-bar position (Taraldsen 1981, Engdahl 1983, Chomsky 1982).[6] In addition to that, German PGs exhibit an interesting asymmetry, similar to the asymmetries that show up with FQs, as discussed in section 2.1.

To begin with, if an indirect object *wh*-phrase binds a PG, then no direct object must intervene between the PG and the *wh*-phrase (9-a). Thus, a direct object that intervenes at the surface between a PG and its binder appears to disrupt the binding relation between the two. In contrast, if the PG is bound by a direct object *wh*-phrase, then an indirect object may intervene between the *wh*-phrase and the PG bound by it (9-b) without causing ungrammaticality.[7]

---

[6] On PGs bound by *wh*-moved categories in German, see Bayer (1984), Fanselow (1993), and Lutz (2001).

[7] Asymmetries with a subject binding a PG cannot be observed for independent reasons: due to the anti-c-command condition on PGs (Engdahl 1983, Chomsky 1982, or Safir 1987), subjects can hardly act as binders for PGs.

(9) a. *Wem$_2$ hat Fritz das Buch [anstatt PG$_2$ zu helfen] geklaut?
who$_{dat}$ has Fritz the book$_{acc}$ instead to help stolen
"Who did Fritz steal the book from instead of helping him?"
b. Was$_2$ hat Fritz der Maria [anstatt PG$_2$ wegzuwerfen] zu essen angeboten?
what$_{acc}$ has Fritz the Maria$_{dat}$ instead away to throw to eat offered
"What did Fritz offer Maria to eat instead of throwing it away?"

Speaking in terms of surface intervention and opacity, intervention of a direct object bleeds the binding relation between an indirect object *wh*-phrase and a P; intervention by an indirect object, however, does not bleed binding of a PG by a direct object *wh*-phrase: again, an instance of counter-bleeding.

As first argued by Felix (1985), scrambled elements in German can also bind PGs (see also Webelhuth 1992: 175-176, Mahajan 1990, Grewendorf and Sabel 1999).[8] In (10), for instance, the direct object has scrambled across the adjunct clause, thereby binding a PG within the adjunct.

(10) Hans hat Maria$_2$ [ohne PG$_2$ anzuschauen] t$_2$ geküsst.
Hans has Maria$_{acc}$ without at to look kissed
"Hans kissed Maria without looking at her."

As has been noted at various occasions (Mahajan 1990: 60, Fanselow 1993: 35, Müller 1995: 232, 261-264; also den Dikken and Mulder 1991 for Dutch), if a PG is bound by a scrambled indirect object, then the direct object must not intervene between the PG and its binder, see (11-a). What has gone unnoticed in the literature so far is the fact that binding of a PG by a scrambled indirect object is impossible even if the direct object has scrambled to a position where it does *not* intervene between the indirect object and the PG, see (11-b).

---

[8] Independently, the same claim was made for Dutch at about the same time, see Bennis and Hoekstra (1984). The question as to whether scrambled categories can bind PGs is still under debate. Huybregts and van Riemsdijk (1985) argue against binding of PGs by scrambling in Dutch; see also Fanselow (1993, 2001), Haider (1997), Kathol (2001), and Haider and Rosengren (2003) on German.

(11) a. *wenn jemand der Anette$_2$ das Buch [ohne PG$_2$ zu
if someone the Anette$_{dat}$ the book$_{acc}$ without to
vertrauen] ausleiht
trust lends
"if someone lends Anette the book without trusting her"
b. *wenn jemand das Buch der Anette$_2$ [ohne PG$_2$ zu
if someone the book$_{acc}$ the Anette$_{dat}$ without to
vertrauen] ausleiht
trust lends

In its surface position, the indirect object in (11-b) is expected to be able to bind the PG because the direct object does not intervene. But this is not the case. Thus, (11-b) instantiates counter-feeding.

Unsurprisingly, if a scrambled direct object binds a PG, then it is possible to scramble an indirect object to a position to the left of the direct object, where it does not intervene between the binder and the PG (12-a). (The same result obtains if the indirect object scrambles to a position to the left of the subject, example omitted.) More interestingly, the indirect object may be even scrambled to a position where it intervenes between the binder and the PG on the surface without disrupting the binding relation, see (12-b).[9] The same state of affairs holds if scrambling of a PG-binding object targets a position to the left of the subject (as already noted in Fanselow 1993), see (12-c) (example with a dative object omitted). Thus (12-b,c) can be understood as instantiating the pattern of counter-bleeding, again bearing strong similarity to what was observed with respect to FQs in section 2.1.

(12) a. dass Hans der Maria das Buch$_2$ [ohne PG$_2$ durchzulesen]
that Hans the Maria$_{dat}$ the book$_{acc}$ without through to read
zurückgibt
back gives
"that Hans returns the book to Maria without reading it through"
b. dass Hans das Buch$_2$ der Maria [ohne PG$_2$ durchzulesen]
that Hans the book$_{acc}$ the Maria$_{dat}$ without through to read
zurückgibt
back gives

---

[9] Judgments are not uniform: while Lee and Santorini (1994, 267) agree with our view, Müller and Sternefeld (1994, 375) and Müller (1995, 263) judge similar examples as ungrammatical.

c. dass das Buch$_2$ jemand [ohne PG$_2$ durchzulesen]
that the book$_{acc}$ someone$_{nom}$ without through to read
weggeworfen hat
away thrown has
"that someone threw the book away without reading it"

To summarize, the facts from PG-binding in German may be captured by the following generalization:

(13) *Intervention Asymmetry for PGs:*
A noun phrase α can bind a PG β in the presence of a co-argument γ that precedes β if and only if γ is higher on the hierarchy $nom > dat > acc$ than α.

Again, (13) is almost identical to the generalization in (7), suggesting a unified account.

There are two environments where scrambled categories have been argued to not act as interveners for the binding relation between a scrambled co-argument and a PG in German. First, as (9-b) already suggests and as noted by Fanselow (1993, 35), subjects never act as interveners for PG-binding. This is also the case when the binder has undergone scrambling instead of *wh*-movement (12-c,d). Second, as again noted by Fanselow (1993) (see also Kathol 2001: 329) if the intervening element binds a PG itself, then it does not interfere with the PG-binding relation of a co-argument to its left (14). Both observations will be accounted for by the analysis in section 4.2.[10]

(14) wenn jemand der Anette$_2$ das Buch$_3$ [anstatt PG$_2$ PG$_3$ zu
if someone the Anette$_{dat}$ the book$_{acc}$ instead to
schenken] nur leiht
give only borrows
"if one only borrows Anette the book instead of giving it to her as a present"

---

[10] It has been argued by den Dikken and Mulder (1991) for Dutch that the intervention asymmetry for PGs does not hold if two weak object pronouns precede an adjunct clause containing the PG. Müller (1995, 263), who discusses den Dikken and Mulder (1991) in another context, seems to presuppose that the same facts hold for German. Our judgment is that clauses with two weak object pronouns are on a par with examples involving full NPs, i.e., the intervention asymmetry is also true for cases with two weak object pronouns.

2.3. Interim Summary

To briefly summarize the findings presented so far, association of an argument α with the FQ *alles* in German is blocked if there is an (indefinite) co-argument γ that intervenes between α and the FQ such that γ is lower on the case hierarchy *nom* > *dat* > *acc* than α. Moreover, similar facts obtain with binding of PGs in German. Abstracting away from certain particularities, the observations from these two domains can be captured by a unified version of the generalizations from sections 2.1–2.2, which is given in (15).

(15) *Generalized Intervention Asymmetry:*
An antecedent α can establish a relation with an associate β in the presence of a co-argument γ that precedes β if and only if γ is higher on the hierarchy *nom* > *dat* > *acc* than α.

The fact that there is a generalization that captures both phenomena under investigation suggests a unified account. Such an account will be presented in section 4. Before we turn to the analysis, however, we must prepare the ground by laying out the theoretical assumptions.

3. Theoretical Background

3.1. EFs and the ISC

The analysis to be presented in section 4 is couched in the probe-goal framework of Chomsky (2000, 2001, 2007). In this framework, the two central operations that form syntactic dependencies are Agree and Move. We adopt the standard view that Agree between two features, the probe and the goal, takes place under c-command (but cf. footnote 27 for qualification). Move is assumed to be subject to the (strict version of the) Phase Impenetrability Condition (PIC, Chomsky 2000), see (16).[11]

(16) *Phase Impenetrability Condition:*
The domain of a head H of a phase HP is not accessible to operations outside of HP. Only H and its edge domain are accessible.

---

[11] Whether Agree is also subject to the PIC is still under debate. Chomsky (2000, 2001) assumes that it is, Bošković (2007) and Müller (2010, 2011) assume that it is not. Since the present proposal draws on the theory of edge feature insertion proposed by Müller (2010, 2011), we are committed to adopt the latter view.

Chomsky (2000, 2001) assumes that only CP and vP are phases. Various works have criticized the conceptual basis of this assumption (Bošković 2002, Boeckx and Grohmann 2007, Epstein 2007, Richards 2011). Here, we depart from it, instead pursuing the idea that every phrase qualifies as a locality domain in the sense of (16) (cf. Koster 1978, van Riemsdijk 1978, Müller 2004, 2010, 2011). In other words: Every phrase is a phase and is therefore subject to (16). On the conceptual side, this gets around the question as to why CP and vP should be special as opposed to other phrasal categories. As a consequence, movement leaving any phrase β must make an intermediate stop at the edge domain of β (see also Sportiche 1989, Takahashi 1994, Boeckx 2003); extraction out of β in one fell swoop is blocked by the PIC.[12]

The notion of edge domain that we presuppose in the present study is given in (17) (where α is a genuine complement of a head β if and only if α and β are sisters and β projects a specifier).

(17) *Edge Domain:*
α is in the edge domain of a head β if and only if α is not a genuine complement of β.

Standardly, the probe-goal theory incorporates the idea that all movement is feature-driven. It must then be ensured that there is a feature that drives movement to the phase edge, as is indirectly required by the PIC. To this end, Chomsky (2007, 2008) proposes that phase heads can have edge features (EFs) inserted on them (cf. already Chomsky 2001: 34). An EF is a property of a head H, indicating that H can merge with some category, potentially creating a specifier of H. We adopt this proposal.

Following Müller (2010, 2011), we assume that EF-insertion is constrained: an EF can be inserted on a head H only if H is still active, that is, if H bears at least one other feature that needs to be discharged by Merge or Agree.[13] This is called the Edge Feature Condition (EFC) in Müller (2010, 2011), where it

---

[12] Note that for the account presented in section 4 to effectively work, it would be sufficient to assume that vP (and CP) are phasal categories: The only intermediate landing site crucially invoked there for the purpose of deriving opacity effects is, in fact, the edge of vP. By sticking to the main text assumption, we merely keep as close to the theory of edge feature insertion proposed by Müller (2010, 2011), which the analysis in section 4 makes use of.

[13] Further constraints on the insertion of EFs are necessary. For instance, it must be ensured that no EF is inserted and attracts a *wh*-phrase that is supposed to remain in-situ. We will not discuss this issue here (but see, e.g., Müller 2011; for theories that avoid the problem but

is also shown that it leads to what is dubbed the Intermediate Step Corollary (ISC), see (18).

(18) *Intermediate Step Corollary:*
Intermediate movement steps to specifiers of X (triggered by EFs) must take place before the final specifier is merged within XP.

To see why (18) holds, consider the following scenario. Suppose the derivation has constructed a VP containing a *wh*-phrase that is ultimately to end up in the specifier of an interrogative C-head. Assume that the *wh*-phrase already occupies the edge domain of VP. In the next step, a v-head is merged with VP. The *wh*-phrase must now move to Specv in order to remain PIC-accessible for further operations. By assumption, such intermediate movement is triggered by an EF. An EF can only be inserted on v as long as v is active. Suppose that v introduces an external argument, and that it therefore bears yet another feature, namely a subcategorization feature that is to be discharged by merging the external argument. Consequently, v remains active as long as the external argument has not been merged. It follows that if an EF is to be inserted on v, this must happen before the external argument is introduced.[14]

Suppose next that the features on a head H that trigger Merge are organized on a stack, and that only the element on the top of the stack is accessible. Once the topmost element is discharged, it is removed, and the second topmost element becomes accessible. If EFs are always inserted on top of this stack, then they must be discharged before any other feature on the stack can be accessed (realizing the last-in-first-out property typical of stacks). Thus, the *wh*-phrase in the scenario above must discharge the inserted EF before the external argument can discharge v's subcategorization feature. Provided the Strict Cycle Condition (SCC, Chomsky 1973), a version of which is (19), this

---

depart from the present approach in that they do not rely on features to derive successive-cyclic movement see, e.g., Takahashi 1994, Bošković 2002, Heck and Müller 2003, Stroik 2009).

[14] Usually, v is assumed to bear ϕ/case features, too. These could equally serve to keep v active, even after the external argument is merged. Thus, insertion of an EF could, in principle, apply after introduction of the external argument. Ultimately, however, this option fails: checking of v's ϕ/case features after the external argument was merged violates strict cyclicity (19) because the creation of a specifier extends the current phrase marker (see Müller 2010, 2011). One may wonder how an EF can be inserted on a head H that does not bear a subcategorization feature that keeps H active. We will not go into this matter here, but see Müller (2011, 179-185).

means that the *wh*-phrase targets an inner specifier of vP while the external argument is merged to the outer Specv.

(19) *Strict Cycle Condition:*
If Σ is the root of the current phrase marker, then no operation can take place exclusively within Ω, where Ω is dominated by Σ.

(20) illustrates the derivation sketched above from the point on where v has been merged. The horizontal arrays to the right of the structures symbolize the status of v's stack at each point. The bottom of the stack is marked by #. [*u*D] stands for v's subcategorization feature.

(20) a.	[$_{\text{vP}}$ v [$_{\text{VP}}$ … *wh* …]]		*u*D	#	(EF-insertion) →
b.	[$_{\text{vP}}$ v [$_{\text{VP}}$ … *wh* …]]	EF	*u*D	#	(Move/EF-deletion) →
c.	[$_{\text{vP}}$ *wh* v [$_{\text{VP}}$ … t$_{wh}$ …]]		*u*D	#	(Merge/*u*D-deletion) →
d.	[$_{\text{vP}}$ subj *wh* v [$_{\text{VP}}$ … t$_{wh}$ …]]			#	…

As is clear from (20-d), the interaction of the ISC and the SCC leads to a configuration where a phrase undergoes movement to a position that ultimately ends up as an inner specifier below a phrase that later undergoes external Merge to become the outer specifier of the same head.[15]

3.2. Scrambling and EFs

If all movement is feature-driven, then the question arises as to what kind of feature drives scrambling. Fanselow (2001, 2003) argues at length that no common morpho-syntactic feature could possibly be a plausible candidate to serve as the trigger for scrambling. A straightforward alternative consists in assuming the existence of some abstract feature [SCR] (see McGinnis 1998, Grewendorf and Sabel 1999, Sauerland 1999, among others). This move has been criticized by Fanselow (2001, 2003) as an unmotivated stipulation (see also Haider 2010: 180), who thus concludes that scrambling does not involve movement to begin with. We would like to point out that if EFs are needed anyway to account for successive-cyclic movement, then one may as well identify them as the trigger of scrambling. This idea is not without precedent. Richards (2004, 4) (building on Chomsky 2001: 34) proposes that scrambling is EPP-driven

---

[15] The resulting structure is similar to what Richards (2001) calls "tucking-in", however, its derivation is not. We return to the issue of tucking-in in more detail in section 5.4.

movement to Specv.[16] We incorporate this proposal, suggesting that in languages such as German EFs on v and on T trigger scrambling.[17] Moreover, we assume that an EF can only attract categories that lie within its c-command domain.

Recently, Chomsky (2007, 11) suggested that the existence of multiple specifiers indicates that EFs do not delete when they have triggered Merge once but rather remain active (a possibility already alluded to in Chomsky 1995). We adopt this suggestion here, assuming that EFs that trigger scrambling behave just like EFs that trigger successive-cyclic movement in that they do not necessarily delete once they have triggered Merge. As a consequence, they can, in principle, attract an arbitrary number of categories that are within their search space. This is what happens in the case of multiple scrambling. If an EF can attract multiple categories, then it suffices to assume that a head can receive one EF per derivation. In what follows, this will be our working hypothesis.

It is often assumed that probing by a feature is subject to the Minimal Link Condition (MLC; Rizzi 1990, Fanselow 1991, Ferguson 1993, Chomsky 1995), a version of which is given in (21).

(21) *Minimal Link Condition:*
If in a representation $\alpha \ldots [\ldots \beta \ldots [\ldots \gamma \ldots] \ldots]$ both β and γ are of the right type to establish a relation R with α, then α can establish R only with β (but not with γ).

If scrambling is triggered by EFs, then one expects it to be subject to the MLC.[18] This, however, does not seem to be the case in German: In this lan-

---

[16] Somewhat similarly, Kitahara (2002), Miyagawa (2001), and Bailyn (2004) claim that scrambling to SpecT discharges an EPP feature. However, these works do not establish any link to successive cyclic movement.

[17] Although, provided the PIC, English must allow for EF-insertion on v (and perhaps on other heads, too), it does not exhibit scrambling. Thus, German differs from English in that in the former, a category attracted by an EF to Specv can remain there while in the latter it cannot. We do not know what this difference derives from but we would like to indicate that a related problem arises for any theory that assumes successive-cyclic movement: in many languages, *wh*-phrases cannot remain in positions that they pass through successive-cyclically. This cannot entirely be due to a requirement that *wh*-phrases must end up in SpecC for scope reasons because often they cannot even remain in a position P if the minimal phrase dominating P undergoes subsequent pied-piping to SpecC (see Heck 2009 and references therein).

[18] Successive-cyclic movement, the parade case of EF-driven movement, is often argued to be subject to the MLC, too.

guage (as opposed to, e.g., Dutch), α can scramble across β even if β, too, could in principle undergo scrambling (see, e.g., Fanselow 2001: 407, Haider 2010: 142).

One could, of course, assume that scrambled arguments bear a feature that non-scrambled arguments lack (rendering the latter irrelevant for the MLC with respect to the scrambling probe). This, however, would again beg the question as to what kind of feature is involved in scrambling. To avoid this problem, we would like to suggest that a probe P can in principle skip some potential goal G, thereby targeting a lower goal G′ (see section 3.3 for some qualification). In other words, we propose that the MLC, as it stands in (21), should be dispensed with.[19] In section 3.3, we present another argument that suggests that the MLC is problematic (see also the remarks on multiple PGs at the end of section 4.2). In section 4 we show that the intervention effects observed in section 2, which might suggest an analysis in terms of the MLC, actually follow from a strictly derivational approach without further ado (in particular without reference to the MLC).

3.3. Parallel Movement

It has been observed that movement of co-arguments is often order preserving, that is, it results in structures that give the impression as if movement had applied "in parallel" (see Müller 2001, Richards 2001, Sells 2002, and Williams 2003 for various cases of order preserving movement and different explanations thereof). Arguably, order preservation effects obtain if co-arguments are attracted by the same feature (see Richards 2001, McGinnis 1998, Anagnostopoulou 2003). To illustrate, consider multiple object shift in Danish (see Vikner 1989, 1995). (22-a,b) show that if two object pronouns in Danish undergo object shift, then their relative order is preserved.

(22) a. Peter viste hende$_2$ den$_3$ jo t$_2$ t$_3$.
Peter showed her it indeed
"Peter indeed showed it to her."
b. *Peter viste den$_3$ hende$_2$ jo t$_2$ t$_3$.
Peter showed it her indeed

---

[19] Müller (2004, 2011) makes a related point and offers an account of superiority that does without the MLC; also cf. Chomsky (2008, 151), Fanselow and Lenertová (2011, 184) for pertinent remarks.

In section 4, we will argue that EF-driven movement (subsuming scrambling and successive-cyclic movement) exhibits this property, too.[20] This plays a crucial role when it comes to explaining the asymmetries illustrated in section 2. But before we turn to the analysis let us present the mechanics that we assume to be responsible for such order preservation effects.

Ignoring the external argument for the moment, suppose that (22-a) comes about because v in Danish can be equipped with a feature that obligatorily attracts all pronouns from within VP into its specifier domain, forming multiple specifiers within vP. Since it is exclusively pronouns that are attracted (there is no object shift with full noun phrases in Danish), we assume that the feature driving object shift is an EF relativized to pronouns; in what follows, we write EF$_{pron}$ for short.[21] Now, if both the SCC (19) and the MLC (21) are respected, then these assumptions result in the partial derivation in (23).

②

(23) [$_{\text{vP}}$ den$_3$ hende$_2$ v . . . [$_{\text{VP}}$ . . . t$_2$ t$_3$ . . .]]

①

Ultimately, however, (23) leads to (22-b), which is ungrammatical. It thus looks as if one either has to give up the SCC or the MLC. In section 3.2, we already suggested that the MLC should be dispensed with.[22] Note, however, that by simply giving up the MLC one does not get order preservation effects for free: Without the MLC, EF$_{pron}$ can choose which pronoun it attracts first, resulting in one order or the other. In what follows, we make a proposal which is compatible with the idea that scrambling is triggered by EFs (as proposed in section 3.2), which obeys strict cyclicity (at the expense of the MLC), and which allows to account for order preservation effects.

Consider again multiple object shift in Danish. To ensure that weak pronouns must move, one may assume that they bear a feature [*u*PRON] that requires checking against EF$_{pron}$. Suppose that the EF$_{pron}$ scans down the tree in search for a goal G. Once it has found a weak pronoun, it may eliminate

---

[20] Similarly, Richards (2001, 60-73), argues on the basis of different facts that multiple short scrambling in Japanese must be order preserving if they target the same specifier domain.

[21] It is not implausible to assume that pronouns, which form a closed class, have a particular property which an EF can be relativized to. In contrast, assuming that scrambling is also due to a relativized EF would bring back the problem as to what the EF should be relativized to in this case, cf. section 3.2.

[22] Hunter and Malhotra (2009) propose yet another account of order preservation effects that retains strict cyclicity at the expense of giving up the MLC.

its [*u*PRON]. At this point, we assume, the pronoun is taken from the tree and placed on a stack.[23] EF$_{pron}$ then proceeds scanning. If it finds another goal G′, it may again eliminate its [*u*PRON] and place G′ on the stack, on top of G, etc. At one point, EF$_{pron}$ will have exhausted its search space. Now the derivation starts to remerge the pronouns it has collected on the stack. In the process of remerging, only the pronoun on top of the stack is accessible. Once the topmost pronoun has been remerged, the one below it is promoted and therefore becomes accessible for the next step of remerge. As a result, the order in which the elements on the stack are remerged is the inverse of the order in which they have been attracted. This re-establishes the original order of G, G′, etc. within the vP domain.[24] The relevant part of the derivation of (22-a) thus proceeds as illustrated in (24-a-e).

(24)	a.	$[_{vP}$ v $[_{VP}$ … h(ende)$_2$ d(en)$_3$ …]]			#	(put *h* on stack) →
	b.	$[_{vP}$ v $[_{VP}$ … t$_2$ d$_3$ …]]		h	#	(put *d* on stack) →
	c.	$[_{vP}$ v $[_{VP}$ … t$_2$ t$_3$ …]]	d	h	#	(remerge *d*) →
	d.	$[_{vP}$ d$_3$ v $[_{VP}$ … t$_2$ t$_3$ …]]		h	#	(remerge *h*) →
	e.	$[_{vP}$ h$_2$ d$_3$ v $[_{VP}$ … t$_2$ t$_3$ …]]			#	…

There is nothing that forces EF$_{pron}$ to attract a pronoun. It could, in principle, skip a higher pronoun and attract a lower one instead because there is no MLC (or it could attract nothing at all). However, if a pronoun fails to be attracted by EF$_{pron}$, its [*u*PRON] will not delete and the derivation crashes. There is, however, one assumption to be made yet: an EF cannot first attract a lower goal and then, in a later step, a higher goal. In other words, when an EF-probe encounters a possible goal G, the derivation must decide once and for all whether G is attracted or not. If G is skipped, no backtracking is possible. This naturally follows from the top-down manner the EF scans its search space.

---

[23] Note that this stack exclusively contains phrases selected for movement and is thus different from the stack proposed by Müller (2010, 2011) and introduced in section 3.1, which hosts the structure building features of a head (subcategorization features and EFs).

[24] Stroik (2009) and Unger (2010) independently (and based on different assumptions) derive order preservation effects by making use of (some kind of) a stack. Richards (2001) puts forward a different approach to order preservation effects which sacrifices strict cyclicity in order to maintain the MLC. In section 5.4, we discuss why we favor the stack approach.

Turning to scrambling and successive-cyclic movement, the same mechanism applies. The only difference is that the EF triggering these movements is not relativized to any particular category; rather it is a bare EF. Accordingly, categories do not undergo scrambling or successive-cyclic movement because they bear an inherent feature that requires checking. As a consequence, scrambling and successive-cyclic movement can, in principle, skip other potential goals without consequences (cf. section 3.2). Of course, if a bare EF attracts more than one category, then all attracted elements have to pass via the stack, as was the case with attraction by EF_{pron} above. In such a situation, one expects multiple scrambling (or successive-cyclic movement) to show order preservation effects. (As with weak pronouns, backtracking never is an option.) This is exactly what the analysis in section 4 will make use of.

4. Analysis

We are now ready to show how the observations from sections 2.1 and 2.2 can be derived. Since we already introduced some aspects of the mechanics that the analysis is based on, in particular the derivation of the ISC and the analysis of order preserving movement, we will, every now and then, gloss over some of the details, simply speaking of EF-movement and parallel movement instead.

4.1. Floating Quantifiers and Opacity

Let us start by clarifying our assumptions about the FQ *alles* in German. We assume, at least for now, that *alles* adjoins to VP (cf. Bobaljik 1995; but see section 5.1 for qualification).[25] Suppose that, semantically, *alles* associates with a *wh*-phrase and that semantic association corresponds to an Agree relation in the syntax, targeting the feature [INDEF].[26] Thus, the FQ *alles* bears a probe

---

[25] Adjunction applies after the outermost specifier has been merged. This may derive from the idea that adjunction, as opposed to specifier-formation, is not feature-driven (see, e.g., Adger 2003) and therefore not subject to the earliness requirement on checking (see below). Adjoined elements and specifiers are treated alike by the PIC.

[26] Beck (1997) argues that *alles* undergoes LF-raising in order to semantically combine with the meaning of the whole interrogative clause. In contrast, Zimmermann (2007) argues for an analysis along the lines sketched above.

[*u*INDEF] that must enter into Agree with [INDEF] on an indefinite. We presuppose that this can only happen if the indefinite c-commands the FQ.[27]
We now turn to the account of bleeding effects with floating *alles*. Let us start by illustrating the case of an indefinite object disrupting the association between a *wh*-subject and a FQ:

(25) *Wer$_1$ hat einem Professor alles$_1$ gratuliert?
who$_{nom}$ has a professor$_{dat}$ all congratulated
"Who all congratulated a professor?"

We enter the derivation of (25) at the point where v has been merged with the VP containing the object. The object in (25) precedes the FQ, the latter being adjoined to VP. The object therefore must, at some point, move out of VP, presumably to a specifier of vP. Since it is the only internal argument, the object occupies the complement position of V. No specifier is present in VP and thus the object is on the edge of VP (cf. the notion of edge domain in (17)). Therefore, it is PIC-accessible. Movement of the object is triggered by an EF on v. Due to the EFC, EFs can only be inserted on a head as long as the head is active. Thus, the EF that is supposed to attract the object must be inserted prior to merging the subject: Recall that what keeps v active is the subcategorization feature that is to be discharged by merging the external argument. An EF is thus inserted on top of the feature stack of v. Since only the topmost element of the stack is visible, the EF must be discharged first. This leads to movement of the object to Specv, see ① in (26).

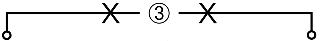
①
(26) [$_{\text{vP}}$ einem Professor$_2$ [$_{\text{VP}}$ FQ$_2$ [$_{\text{VP}}$ … t$_2$ …]]]
②

In line with the requirement that probe features are discharged as early as possible once they have been introduced into the derivation (see Pesetsky 1989, Chomsky 1995: 233, Lasnik 1999, among others), [*u*INDEF] on the FQ

---

[27] This departs from the standard assumption that the probe must c-command the goal (but cf. Adger 2003, Koopman 2006, Baker 2008 for diverging views). The same non-standard assumption will be made for PGs (see section 4.2) but not for Agree relations in general. That is, we still take it that, generally, a probe [$u\phi$] must c-command the goal [ϕ]. Note in passing that if one were to follow Fitzpatrick (2006) in assuming that adverbial FQs are merged with an empty *pro*, and provided that it were *pro* that bears [*u*INDEF], then, given the standard view, [*u*INDEF] could not enter into Agree anyway due to the lack of c-command; similarly for PGs, see section 4.2.

and [INDEF] on the indefinite object immediately enter into Agree (step ② in (26)).[28] Thus, when the subject is merged in the next step, it cannot establish Agree with the FQ, as indicated by the crossed-out association line in (27), because the probe on the FQ has already been eliminated by step ② in (26).



(27) [$_{vP}$ wer$_1$ einem Professor$_2$ [$_{VP}$ FQ$_2$ [$_{VP}$ … t$_2$ …]]]

Since Agree with a FQ, by assumption, is a precondition for semantic association with it, the derivation is doomed to crash at the semantic interface.[29]

Bleeding that involves two objects, an indirect *wh*-object and an indefinite direct object (see (28)) is derived along similar lines. The only assumption we have to add (which is often made for German anyway, cf. section 5.2) is that the indirect object is merged above the direct object within VP.

(28) *Wem$_1$ hat sie einen Professor alles$_1$ vorgestellt?
who$_{dat}$ has she a professor$_{acc}$ all introduced
"Who all did she introduce to a professor?"

In the derivation of (28), the direct object must again reach a position to the left of the FQ, minimally Specv. But this time, it must first move to SpecV in order to remain PIC-accessible. The reason is that after the indirect argument is merged to SpecV, the direct object, being the sister of the verb, counts as a genuine complement in the sense of (17). As V still bears the subcategorization feature for the indirect object, an EF can be inserted on it, but only before the indirect object is merged.[30] Once inserted, the EF has to be discharged before the subcategorization feature for the indirect object can be accessed.

---

[28] Arguably, Merge of the subject also discharges an uninterpretable feature on v, namely its subcategorization feature, an operation that must be procrastinated if Agree between the object and the FQ is supposed to apply first. Here, and in what follows, we assume that there is an inherent preference of Agree over Merge in the vP-domain (see Heck and Müller 2007 for a related proposal).

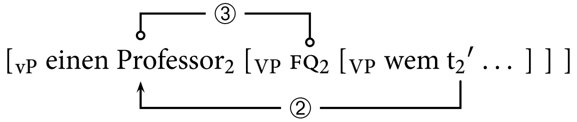
[29] Note that the indefinite object in (26) has checked [*u*INDEF] on the FQ. Thus, syntactically, (25) is fine. Yet, it is not well-formed. Our hunch is that, semantically, the non-*wh* indefinite is not of the right type to combine with the FQ.

[30] Technically, the direct object is *not* a genuine complement before the indirect object is merged. There is no need to invoke structural look-ahead here, though: By inspecting the subcategorization features on V, the derivation can determine early whether the direct object will end up as a genuine complement within VP or not.

Therefore the direct object moves first. Only then is the indirect object *wh*-phrase merged to an outer specifier of VP (29).


(29) [$_{\text{VP}}$ wem einen Professor$_2$ [$_{\text{V}'}$ … t$_2$ …]]

In the following steps, the FQ is adjoined to VP, and the two objects undergo parallel EF-movement to inner Specvs (the subject is later introduced to the outermost Specv). Due to parallel movement, the direct object is remerged first and immediately establishes Agree with the FQ (steps ② and ③ in (30)).

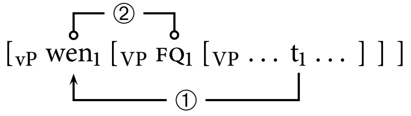

(30) [$_{\text{vP}}$ einen Professor$_2$ [$_{\text{VP}}$ FQ$_2$ [$_{\text{VP}}$ wem t$_2'$ …]]]

When the indirect object *wh*-phrase is remerged to Specv, it cannot enter into Agree with the FQ because the probe of the FQ has already been consumed (step ③ above). Bleeding is the result.

Having said this, it is obvious what happens if an object *wh*-phrase is remerged as the first argument to Specv but later moves to SpecC. In such a scenario, counter-bleeding arises. (31) illustrates:

(31) Wen$_1$ hat ein Professor alles$_1$ beleidigt?
who$_{acc}$ has a professor$_{nom}$ all insulted
"Who all did a professor insult?"

Suppose the derivation of (31) has already constructed vP. The direct *wh*-object has moved into an inner specifier of vP in order to remain PIC-accessible (see ① in (32)). Due to the ISC, the subject is introduced to an outer Specv. However, before this can happen, the *wh*-object enters into Agree with the FQ (step ②). As Agree is the syntactic precondition for semantic association, the *wh*-phrase can associate semantically with the FQ.


(32) [$_{\text{vP}}$ wen$_1$ [$_{\text{VP}}$ FQ$_1$ [$_{\text{VP}}$ … t$_1$ …]]]

Later, the subject is merged and the object *wh*-phrase moves (via SpecT) to SpecC (step ③ in (33)), thereby giving the surface impression of the indefinite subject intervening between the *wh*-phrase and the FQ. (For simplicity, we follow Grewendorf 1989 and Diesing 1992 in assuming that subject raising to SpecT is optional in German, but nothing hinges on this.)

③

(33) [$_{\text{CP}}$ wen$_1$ … [$_{v\text{P}}$ ein Professor$_2$ t$_1'$ [$_{\text{VP}}$ FQ$_1$ …]]]

This derives counter-bleeding. Due to the PIC, the object cannot move in one fell swoop to SpecC, thereby skipping the intermediate landing in Specv, a crucial precondition for opacity to arise. Of course, from a derivational point of view, there is nothing opaque about (31): FQ-association applies in a transparent manner, albeit at an intermediate derivational stage.[31]

4.2. Parasitic Gaps and Opacity

Next, consider opacity effects with PGs. To begin with, suppose that adjunct clauses containing PGs are adjoined to VP (cf. Nissenbaum 2000: 35-36). Suppose further that PGs come about by movement of a zero element OP from the position of the PG to SpecC of the adjunct clause (see Chomsky 1986, Nissenbaum 2000, among others). Similar to what was the case with FQs, we assume that semantic binding of a PG requires previous Agree in the syntax (see Assmann 2012). To this end, we propose that an OP associated with a PG bears a probe [*u*D] which must enter into Agree with an antecedent.[32] Once [*u*D] is eliminated, OP can no longer participate in Agree with another antecedent. Informally, we will speak of agreement between an antecedent and a PG, although technically, what is meant is an Agree relation between an antecedent and the OP associated with a PG. For the sake of brevity, we confine ourselves to PGs bound by scrambled elements, leaving aside PG-binding by *wh*-phrases ((9-b), section 2). This happens without loss of generality because, for our purposes, the latter constitute a subset of the former.

---

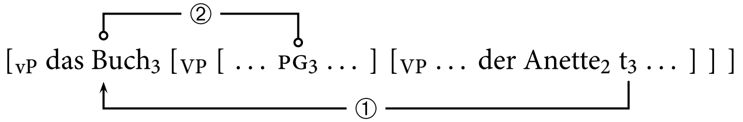
[31] According to Ko (2007), Korean exhibits almost identical asymmetries involving intervention with FQ-association. The analysis Ko (2007) presents is formulated in terms of cyclic linearization (Fox and Pesetsky 2005). For the most part, it can be transferred to German. Space limitations do not permit a detailed discussion at this point. Put briefly, we think that Ko's (2007) theory cannot account for all of the facts involving PGs (section 2.2), or at least not obviously so. In contrast, the present theory provides a unified analysis of the facts from PG-binding and FQ-association, which is motivated for German.

[32] Assmann (2012) derives OP from a lexical rule operating on the antecedent (see also Agbayani and Ochi 2007). This rule can also be held responsible for introducing [*u*D] on OP. Note that, similarly to what we assumed for FQs, [*u*D] must be c-commanded by the goal it agrees with (cf. sections 4.1, in particular footnote 27).

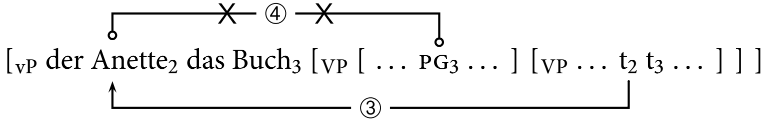
First consider (34), which illustrates that an intervening scrambled direct object blocks a scrambled indirect object from binding a PG.

(34) *wenn jemand der Anette$_2$ das Buch [anstatt PG$_2$ zu helfen]
if someone the Anette$_{dat}$ the book$_{acc}$ instead to help
wegnimmt
away takes
"if someone takes the book from Anette instead of helping her"

By assumption, the indirect object is merged higher within the VP than the direct object. Because the adjunct clause is left-adjoined to VP, both objects have to scramble to Specv in order to precede it. Scrambling is triggered by an EF on v, which attracts both objects in parallel. The direct object is remerged first and enters into Agree with the PG at once, see steps ① and ② in (35).

(35)


Agree with the direct object eliminates [*u*D] on the PG. Thus, the PG is no longer active and cannot establish Agree with the indirect object when it is remerged in the next step (③ in (36)). The indirect object simply re-enters the structure too late. This derives bleeding.

(36)


In principle, one would expect the PG in (34) to be semantically bound by the direct object. The corresponding reading, however, is blocked by requirements on the binder imposed by the embedded verb with respect to case and animacy (the direct object *das Buch* "the book" is inanimate and bears accusative; in contrast, the embedded verb *helfen* "to help" requires an animate object in the dative).

The interesting question is why the indirect object is still not able to bind the PG if scrambling of the direct object ends up in a position to the left of the indirect object (37).

(37) *wenn jemand das Buch der Anette$_2$ [ohne PG$_2$ zu vertrauen]
if someone the book$_{acc}$ the Anette$_{dat}$ without to trust
ausleiht
lends
"if someone lends Anette the book without trusting her"

Both objects in (37) must minimally target a Specv-position to the left of the adjunct clause. Since v is assigned only one EF per derivation, this EF must attract both objects, preserving their relative order. There is no other way the direct object can reach Specv, and therefore the steps ①, ②, and ③ of the derivation of (37) are identical to those in (35) and (36) above. The direct object moves first (more precisely: is remerged first) and immediately enters into Agree with the PG. The indirect object moves next, but it comes too late: The PG has already become inactive by then. Later, an EF on T in parallel attracts the direct object and the subject, which has been merged to the outermost Specv. Both arguments thus end up in SpecT, to the left of the indirect object (see steps ④ and ⑤ in (38)).

⑤
(38) [$_{\text{TP}}$ Hans$_4$ das Buch$_3$ [$_{\text{vP}}$ t$_4$ der Maria$_2$ t$_3'$ [… PG$_3$ …] …]]
④

Although, on the surface, the indirect object is closest to the PG, it cannot bind it because, derivationally, the direct object reaches the position relevant for binding first. This leads to counter-feeding. Finally note that movement of the direct object to SpecT in one fell swoop (fusing steps ① and ④ above into one movement step) is blocked by the PIC. The intermediate halt in Specv that leads to the local Agree relation between the direct object and the PG cannot be skipped. Binding of the PG by the direct object is forced and the ungrammaticality of (37) is thus explained.

A direct object that has scrambled to the left of an adjunct clause containing a PG can always bind the PG. This is obvious for cases where the direct object appears closest to the PG on the surface (see (12-a), section 2.2). But even if an indirect object shows up in between the direct object and the adjunct clause is PG-binding by the direct object possible (39).

(39) dass Hans das Buch$_2$ der Maria [ohne PG$_2$ durchzulesen]
that Hans the book$_{acc}$ the Maria$_{dat}$ without through to read
zurückgibt
back gives
"that Hans returns the book to Maria without reading it through"

In all relevant respects, the derivation of (39) is the same as the one of (37) above, but this time it results in counter-bleeding. The reason is that in (39) syntactic binding of the PG by the direct object also leads to semantic binding: there is no incompatibility between the requirements of the embedded verb and the direct object with respect to case and animacy (*durchlesen* "to read through" requires an inanimate object in the accusative, exactly the specification of the direct object *das Buch* "the book"). After Agree, the direct object scrambles to the left of the indirect object, rendering the structure opaque. Recall that subjects never act as interveners for PG-binding by an object. (40) is a case in point.

(40) dass das Buch$_2$ jemand [anstatt PG$_2$ wegzuwerfen] verschenkt
that the book$_{acc}$ jemand instead away to throw give away
"that someone gives the book away instead of throwing it away"

It follows immediately why this should be so. Although the object in (40) precedes the subject on the surface, it must have occupied a position in between the subject and the adjunct clause at some earlier derivational stage, due to the by now familiar interaction of PIC and ISC: the PIC enforces object movement to the edge of vP (see ① in (41)); the ISC ensures that the object occupies an inner specifier, below the external argument. From there, it binds the PG (step ② in (41)).

②
(41) [$_{vP}$ das Buch$_2$ [$_{VP}$ [… PG$_2$ …] [$_{VP}$ … t$_2$ …]]]
①

When the subject is merged to the outermost specifier of vP, the PG has already become inactive. Later, the object scrambles to SpecT, which makes it look as if the subject were closest to the PG and should therefore block PG-binding by the object: counter-bleeding. From the derivational perspective, the subject simply cannot reach a position appropriate for binding before the PG is deactivated by the object.

Finally, remember that a direct object does not interrupt PG-binding by an indirect object if the direct object itself binds another PG (42).

(42) wenn jemand der Anette$_2$ das Buch$_3$ [anstatt PG$_2$ PG$_3$ zu
if someone the Anette$_{dat}$ the book$_{acc}$ instead to
schenken] nur leiht
give only borrows
"if someone only borrows Anette the book instead of giving it to her as a present"

In section 3.2, we suggested that the MLC should be eliminated. From this, it now follows that the direct object in (42) can bind PG$_3$ across PG$_2$, rendering the former inactive. Under a theory that seeks to account for intervention effects with PG-binding in terms of the MLC (see Fanselow 1993; also Müller 1995: 264), it remains unclear why the direct object can skip PG$_2$, binding PG$_3$ instead.

The question arises as to why the direct object does not also enter into Agree with PG$_2$, thereby bleeding binding of PG$_2$ by the indirect object. After all, PG$_2$ also bears [*u*D], and, by assumption, uninterpretable features have to be eliminated as soon as possible once they have been introduced into the derivation (see section 4.1 and also section 5.4 below). At this point, we have no explanation for this and must therefore resort to the stipulation that an antecedent can bind only one PG.[33] Thus, the direct object exclusively binds PG$_3$, leaving PG$_2$ to be bound by the indirect object.[34]

5. Further Issues

5.1. Wh-in-Situ

In section 2.1, we illustrated counter-bleeding in the context of FQ-association. In fact, the analysis suggested there also predicts the existence of counter-feeding within the realm of FQs. The relevant scenario involves surface adjacency between a *wh*-phrase and a FQ that comes about by scrambling an

---

[33] If there are multiple PGs bound by a single antecedent, then more has to be said here. Ross (1967, 192) mentions potential cases but at the same time judges them as "less than felicitous".

[34] Theoretically, the direct object could as well have bound PG$_2$, leaving PG$_3$ to the indirect object. Such a binding in (42), however, is blocked due to a mismatch with respect to case and animacy.

indefinite away from in between the former two: Before the indefinite scrambles, it associates with the FQ and thus prevents the *wh*-phrase from doing so. The scenario presupposes that the *wh*-phrase does not undergo *wh*-movement at a later step, as is the case with *wh*-in-situ in multiple questions. As we will see shortly, the prediction that counter-feeding as described above exists is *not* borne out. A similar problem arises for counter-bleeding and *wh*-in-situ. In what follows, we will make a proposal as to why this should be so.

To begin with, (43) illustrates that *wh*-in-situ phrases can, in principle, associate with a FQ. (This also holds if the *wh*-phrase bears a grammatical function other than direct object.)

(43) Wann hat sie einem Professor wen$_2$ alles$_2$ vorgestellt?
when has she a professor$_{dat}$ who$_{acc}$ all introduced
"When did she introduce who all to a professor?"

Starting with counter-feeding and *wh*-in-situ, suppose a scenario where a *wh*-object is merged higher than its indefinite co-argument. If both arguments scramble to the left of *alles*, then association of the *wh*-phrase with *alles* is blocked because the indefinite is remerged first and thus eliminates [*u*INDEF] on the FQ. Suppose that, next, the indefinite scrambles to the left of the *wh*-phrase. Then, the surface looks as if it would feed association between the *wh*-phrase and the FQ. But in fact, the FQ has associated with the indefinite and the output is predicted to be ungrammatical. The facts are given in (44).

(44) a. Wann hat sie einen Professor$_1$ wem$_2$ t$_1$ alles$_2$ vorgestellt?
when has she a professor$_{acc}$ who$_{dat}$ all introduced
"When did she introduce a professor to who all"

b. Wann hat einem Professor$_1$ wer$_2$ t$_1$ alles$_2$ geholfen?
when has a professor$_{dat}$ who$_{nom}$ all helped
"Who all helped a professor and when?"

c. Wann hat einen Professor$_1$ wer$_2$ t$_1$ alles$_2$ verklagt?
when has a professor$_{acc}$ who$_{nom}$ all sued
"Who all sued a professor and when?"

The prediction is not borne out. (44-a-c) are on a par with (43). Thus, the theory under-generates.

Turning to counter-bleeding, first note that FQ-association with an in-situ *wh*-phrase can be blocked by an intervening indefinite (45). (Again, examples where the *wh*-phrase bears a grammatical function other than indirect object are omitted for the sake of brevity.)

(45) *Wann hat sie wem$_1$ einen Professor alles$_1$ vorgestellt?
when has she who$_{dat}$ a professor$_{acc}$ all introduced
"When did she introduce a professor to who all?"

Assume now a scenario where the intervening indefinite is merged higher than the *wh*-phrase. When both arguments scramble in parallel to the left of *alles*, the *wh*-phrase is remerged first and associates with the FQ. Later, the *wh*-phrase scrambles across the indefinite. The result should be well-formed because, in theory, it instantiates counter-bleeding: the surface position of the *wh*-phrase does not indicate transparently that it had the chance to associate with the FQ. Relevant examples are given in (46).

(46) a. ?*Wann hat sie wen$_1$ einem Professor t$_1$ alles$_1$ vorstellt?
when has she who$_{acc}$ a professor$_{dat}$ all introduced
"When did she introduce who all to a professor?"
b. ?*Wann hat wem$_1$ ein Professor t$_1$ alles$_1$ geholfen?
when has who$_{dat}$ a professor$_{nom}$ all helped
"When did a professor help who all?"
c. ?*Wann hat wen$_1$ ein Professor t$_1$ alles$_1$ erkannt?
when has who$_{acc}$ a professor$_{nom}$ all recognized
"When did a professor recognize who all?"

The grammaticality of (46-a-c) seems severly degraded. Therefore, the theory also over-generates.

There are at least two possible reactions to this . First, we could conclude that the theory presented in section 4 is on the wrong track. Second, we could (a) take further possible factors into account that explain the ungrammaticality of (46), and (b) modify the theory such that it derives the grammaticality of (44) without giving up the gist of the analysis. Given that the facts from section 2 are sufficiently robust and that the account proposed so far is conceptually attractive because it also provides an explanation for an asymmetry in another domain (section 4.2), we opt for the second possibility.

We begin with (46). What distinguishes these examples from the well-formed instances of counter-bleeding in (6) (section 2.1) is that in (46) the *wh*-phrase remains in a position it has reached by EF-driven scrambling. While, according to the present analysis, the *wh*-phrase in (6) must also undergo EF-driven movement in order to reach a position from where it can associate with the FQ, it does not stop there but undergoes subsequent *wh*-movement. Interestingly, it has been argued (see, e.g., Wiltschko 1997, Sauerland 1999) that *wh*-phrases that end up in scrambling positions in German are interpreted as D(iscourse)-linked in the sense of Pesetsky (1987). On the assumption that *which*-phrases in English are inherently D-linked, Pesetsky (1987, 107-108) describes this concept as follows: "When a speaker asks a question like 'Which book did you read?', the range of felicitous answers is limited by a set of books both speaker and hearer have in mind." Now, Reis (1992), drawing on a notion of definiteness proposed by Hawkins (1978), observes that *alles* requires its antecedent to denote an "open set" in the sense that "there is no anaphoric or deictic/situational link to an independently established antecedent set". It is our hunch that these two requirements lead to a semantic incompatibility in (46): *alles* must associate with an antecedent whose denotation is not tied to a situationally constrained set, yet scrambling of the *wh*-phrase (inducing D-linking) creates precisely such an interpretation for the *wh*-phrase. We therefore, once more, suggest that although the examples in (46) are syntactically well-formed, they do not receive a proper interpretation.

Turning to the examples in (44), we saw that they receive an analysis in terms of counter-feeding that, counter-factually, predicts them to be ungrammatical. If there were an alternative analysis such that the *wh*-phrase could associate with the FQ, the grammaticality of the examples in (44) would follow. This alternative must not be available for examples such as (45) and for those in (5) (section 2.1), for which an counter-feeding analysis makes the correct prediction. All this can be achieved by assuming that *alles* can appear in two positions: it can be adjoined to VP (as assumed so far); or it can be merged directly with a *wh*-phrase (similar to the original proposal in Reis 1992). That the second option is needed anyway is suggested by the fact that *alles* can undergo pied-piping together with the *wh*-phrase (47).[35]

---

[35] It is generally assumed that in German V/2 main-clauses only one constituent can precede the finite verb. Thus, *wen alles* in (47) must form a constituent. Assuming two merge positions for invariant *alles* is not without precedent: In their analysis of the FQ *beide* "both" in German, Reis and Vater (1980) distinguish two positions for floating *beide*, one forming a constituent with the associated argument, the other appearing in isolation. Similarly, Link (1974, 124, footnote 7) argues, also on the basis of facts from V/2 clauses, that there are two positions for variant *all(es)*.

(47) [Wen alles $]_2$ hat sie t_2 beleidigt?
who all has she insulted
"Who all did she insult?"

We would like to propose that the second option is what underlies an alternative derivation of the examples in (44). If the *wh*-phrase and FQ form a constituent, then Agree (and therefore semantic association) can apply between them right away. In the next step, the constituent consisting of *wh+alles* is merged to an argument position above the indefinite. Finally, the indefinite undergoes scrambling across *wh+alles*. (In other words, the traces left by scrambling in (44) should actually be not right-adjacent to the *wh*-phrase but right-adjacent to the FQ.) This derivation does not involve counter-feeding and is therefore predicted to result in grammaticality.

What remains to be explained is why a derivation based on merging the *wh*-phrase directly with the FQ is not available for (45) and the examples in (5). To this end, we propose that a *wh*-phrase cannot strand a FQ it has been merged with. Independent motivation for this proposal can be gained from the hypothesis that pied-piping is a last resort strategy (see Heck 2009 and references therein). To put it in a nutshell, the last resort analysis of pied-piping implies that the existence of a structure such as (47) owes to the non-existence of an alternative derivation (also based on merging the *wh*-phrase and the FQ) that involves stranding of *alles*. Since (47) is grammatical, there must be a ban against stranding *alles*.

5.2. Verb Classes

It is often assumed that the underlying order of object arguments in German is indirect object > direct object (see Lenerz 1977, Thiersch 1982, Webelhuth 1992: 194-199, among others). We have followed this view here. A more fine grained distinction is argued for in Haider (1992, 1993, 2010), where it is claimed that different verb classes project different relative orders of objects in German.[36] Thus, while verbs such as *geben* "give" or *vorstellen* "introduce" project the order indirect object > direct object, verbs such as *aussetzen* "to

---

[36] A more radical alternative can be found in Müller (1995, 2001): There, it is assumed that direct object > indirect object is the underlying order of objects in German throughout. Obviously, this is incompatible with the analysis presented in section 4.

expose" or *unterziehen* "to subject" are claimed to belong to a minor class of verbs in German that project the order direct object > indirect object.

Since we argued that the underlying order of arguments is preserved by multiple scrambling and that it can be detected by grammatical processes such as FQ-association and PG-binding, the same diagnostics can now be put to use to see whether different verbs impose different relative orders on their objects. If verbs such as *aussetzen*, *unterziehen*, etc. indeed induce the order direct object > indirect object, then this, combined with the present theory, predicts that with these verbs it is always the indirect object which can associate with a FQ or to bind a PG and not the direct object (provided both objects are in a position appropriate for association and binding to begin with).

Let us start by considering examples that involve PG-binding. We think that there is a detectable, albeit subtle, asymmetry in (48-a-d). In fact, however, the asymmetry favors PG-binding by the direct object over binding by the indirect object. This is at variance with the prediction mentioned above.

(48) a. ?weil er das Kind$_2$ dem Test [ohne PG$_2$ zu schonen]
because he the child$_{acc}$ the test$_{dat}$ without to spare
aussetzte
exposed
"because he exposed the child to the test instead of sparing her"

b. *?weil er das Kind dem Test$_3$ [ohne PG$_3$ zu vertrauen]
because he the child$_{acc}$ the test$_{dat}$ without to trust
aussetzte
exposed
"because he exposed the child to the test without trusting it (the test)"

c. ?weil er das Instrument$_2$ der Prüfung [ohne PG$_2$ zu
because he the instrument$_{acc}$ the test$_{dat}$ without to
schonen] unterzog
spare subjected
"because he subjected the instrument to the test without preventing it from damage"

d. *weil er das Instrument der Prüfung$_3$ [ohne PG$_3$ zu
because he the instrument$_{acc}$ the test$_{dat}$ without to
misstrauen] unterzog
distrust subjected
"because he subjected the instrument to the test without distrusting it"

Relevant examples that involve association with the FQ *alles* illustrate the same asymmetry, perhaps more clearly so, see (49).[37]

(49) a. Welche Beziehungen$_2$ hat er einer Belastung alles$_2$ ausgesetzt?
which relationships$_{acc}$ has he a strain$_{dat}$ all exposed
"Which relationships all did he strain?"
b.?*Welchen Belastungen$_2$ hat er eine Beziehung alles$_2$ ausgesetzt?
which strains$_{dat}$ has he a relationship$_{acc}$ all exposed
"Which strains all did he put on a relationship?"
c. Welche Instrumente$_2$ hat er einer Prüfung alles$_2$ unterzogen?
which instruments$_{acc}$ has he a test$_{dat}$ all subjected
"Which instruments all did he subject to a test?"
d. *Welchen Prüfungen$_2$ hat er ein Instrument alles$_2$ unterzogen?
which tests$_{dat}$ has he an instrument$_{acc}$ all subjected
"To which tests all did he subject an instrument?"

Thus, the diagnostics from PG-binding and FQ-association appear to indicate that even with verbs such as *unterziehen*, *aussetzen*, etc., the underlying order of objects in German is indirect object > direct object. As already noted, this is at variance with the claim put forward in Haider (1992, 1993, 2010). There are independent arguments in the literature that support this claim. One such argument is based on the observation (due to Höhle 1982) that maximal focus projection from an argument immediately preceding the verb on the surface is possible in German only if this argument is the underlying sister of the verb. Crucially, with verbs that belong to the minor class maximal focus projection is possible from the indirect object but not from the direct object. Another argument, going back to Frey (1993), relies on the observation that in German a quantifier Q_1 that is c-commanded by another quantifier Q_2 on the surface can scope over Q_2 if it c-commands a trace of Q_2. It turns out that a direct object quantifier can scope over an indirect object quantifier under the surface

---

[37] The dative marked argument of *aussetzen* and *unterziehen* is typically inanimate. But the simplex inanimate *wh*-phrase *was* "what" in German cannot serve as an argument that is dative marked by a verb (see Pittner 1996). For this reason, the examples (49-b) and (49-d) were chosen to involve *welch*-phrases ("which"-phrases). If German *welch*-phrases were inherently D-linked (as is often assumed for *which*-phrases in English), then they would be semantically incompatible with invariant *alles* (see section 5.1). However, Reis (1992) argues that, despite common believe, *welch*-phrases in German are not inherently D-linked. Hence, another account of the ill-formedness of (49-b) and (49-d) is needed.

order indirect object > direct object only if they are co-arguments of a verb belonging to the minor class. Taking these arguments seriously, the question arises as to how the claim made in Haider (1992, 1993, 2010) can be reconciled with the present findings.

To this end, we would like to invoke a proposal put forward in Meinunger (2000), where it is argued that the indirect objects of verbs belonging to the minor class are actually PPs headed by an empty preposition. Such PPs are merged lower than their direct object co-arguments, namely as the sister of the verb. In this position, the indirect object can project its focus on the whole clause (in line with Höhle's 1982 generalization); and when it moves away, it leaves behind a trace that can be c-commanded by a direct object quantifier, thus leading to scope inversion (in agreement with Frey's 1993 generalization). However, if the indirect object is embedded within a PP, it will not be able to bind a PG or to associate with a FQ, since it does not c-command the PG or the FQ. Therefore, with verbs belonging to the minor class it will always be the direct object that binds a PG or associates with a FQ, even if the underlying order projected by these verbs is direct object > indirect object.

5.3. Scrambling as a Transformation

In the preceding discussion, we presupposed that scrambling comes about by a movement transformation. Of course, this is not a new idea. It goes back at least as far as Bierwisch (1963, 100-101) and Ross (1967, 74-78)[38], and it has often been argued for since then (see Fanselow 1990, Giusti 1990, Webelhuth 1992: 164-178, Müller and Sternefeld 1994, Grewendorf and Sabel 1999). However, there are also approaches, for a variety of languages (including German), that analyze scrambling in terms of base generation (see Haider 1988, Fanselow 1993, 2001, 2003, Bayer and Kornfilt 1994, Kiss 1994, Neeleman 1994, Bošković and Takahashi 1998).

Some proposals that argue for a base generation approach critically discuss the traditional arguments put forward by proponents of the movement based approach to scrambling (see in particular Fanselow 1993, 2001 on German; cf. also Bailyn 2001 and Bošković 2004 for relevant discussion on Russian and Japanese). Sometimes these arguments are based on the assumption that

---

[38] Although Ross (1967) assumed scrambling to be a transformation, he did not consider it a movement transformation in the narrow sense but argued that it better be placed outside syntax in what he called the "stylistic component".

scrambled categories are associated with their underlying positions. Accordingly, the criticism often focuses on the tests supposed to provide evidence for such positions, trying to show that they are not decisive. Interestingly, the present arguments are based on the existence of *intermediate* positions (instead of underlying ones) because it is these positions that ultimately render opaque configurations transparent. As such, these arguments are not subject to the above mentioned criticism. But more importantly, we now briefly illustrate that, all other things equal, an approach to scrambling in terms of base generation faces problems when confronted with (some of) the opacity effects discussed in the present study.
First consider the case of simple bleeding with a FQ in (50-a).

(50) a. *Wem$_1$ hat sie einen Professor alles$_1$ vorgestellt?
who$_{dat}$ has she a professor$_{acc}$ all introduced
"Who all did she introduce a professor to?"
b. Wem$_1$ hat sie alles$_1$ einen Professor vorgestellt?
who$_{dat}$ has she all a professor$_{acc}$ introduced

Given that (50-a) is ungrammatical, a base-generation approach to scrambling could, as a first hypothesis, assume that indefinites that are base generated in between a *wh*-phrase and a FQ interrupt the association between the latter two. Provided the assumption that there is no scrambling transformation, the *wh*-phrase and the FQ in (50-b) must then be base generated adjacent to each other (and to the left of the indefinite) before *wh*-movement applies. Moreover, since the indirect and the direct object can appear in any order, it follows that it must also be possible to base generate an indefinite direct object to the left of an indirect *wh*-object to the left of a FQ. If *wh*-movement applies to such a configuration, then the result is again (50-a). To block this derivation of (50-a), the base generator could formulate the second hypothesis that movement of a *wh*-phrase associated with a FQ must not cross an indefinite.

But now consider the case of counter-bleeding in (51):

(51) Wen$_1$ hat ein Professor alles$_1$ beleidigt?
who$_{acc}$ has a professor$_{nom}$ all insulted
"Who all did a professor insult?"

In (51), the *wh*-phrase successfully associates with the FQ. On the one hand, it could have been generated adjacent to the FQ, in line with the base-generator's first hypothesis. This, however, is at variance with the second hypothesis because the *wh*-phrase must then move across the indefinite. If, on the other hand, the *wh*-phrase and the FQ are base generated separately (in agreement with the second hypothesis), with the indefinite in between them, then this is in conflict with the first hypothesis. To conclude, there is no way the base generator can account for both (50-a,b) and (51), at least not obviously so. The present study thus provides a novel argument for the idea that scrambling is a movement transformation.

5.4. Tucking-in

In section 4, we presented an analysis of ordering effects with multiple specifiers which distinguishes two cases. First, there are multiple specifiers that come about by a mixture of two operations: inner specifiers are created by Move, the outermost specifier is created by Merge. Second, there are multiple specifiers created by pure movement. Multiple specifiers of the mixed type were assumed to be the result of the ISC (based on the theory of EF insertion in Müller 2010, 2011). Instances of the pure type were argued to follow from the way the derivation handles multiple attraction by a single EF: first, the attracted categories are collected on a stack; then, they are remerged in the inverse order.

Richards (2001) provides an alternative mechanism to derive ordering effects in multiple specifiers of the pure type, which he dubs the theory of tucking-in. The gist of this theory is as follows. Suppose that α and β are attracted by the same probe on a head H. If α asymmetrically c-commands β, then the transderivational constraint Shortest Paths ((52); Chomsky 1995, Collins 1994, Nakamura 1998) requires that H first attracts α and then β:

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

If α asymmetrically c-commands β, then the path from α to SpecH is shorter than the path from β to SpecH. Therefore, the derivation where α moves first blocks the derivation where β moves first. In a second step, then, β undergoes movement and "tucks in" to the innermost specifier position γ, which is below the one occupied by α. Richards (2001) argues that tucking-in to the innermost specifier is also forced by Shortest Paths because the path from a position P to the innermost specifier of some head H is shorter than the path from P to an outer specifier of H. In the above scenario, β's movement path to the innermost specifier γ is shorter than β's movement path to an outer specifier position above α.

Order effects with multiple specifiers of the mixed type are only briefly touched upon by Richards (2001, 75). But it seems as if the tucking-in theory were also applicable to those. The reason is that in this theory the question as to whether movement targets an inner or an outer specifier does not depend on whether an already existing specifier β in the same specifier domain has been created by Move or by Merge. All that counts is that the path to a position below β is shorter than the path to a position above it. The only thing that needs to be added to the theory of tucking-in in order to derive the same order within mixed multiple specifiers as derived by the ISC is the assumption that Merge applies before Move (see Chomsky 1995), i.e., the specifier that comes about by Merge must be created first.

Since, representationally, multiple specifiers are the same, no matter whether they come about by pure movement or by a mixture of Move and Merge, it is natural to derive them by the same mechanism. In contrast, the theory on EF insertion that underlies the ISC has nothing to say about multiple specifiers of the pure type. Thus, the broader domain of application of Richards' (2001) tucking-in theory as compared to the one covered by the theory of EF insertion and the stack theory, respectively, when considered in isolation, speaks in favor of the theory of tucking-in.[39] Despite this, we have opted against the tucking-in theory for the following conceptual reasons.[40]

---

[39] This is not entirely correct because, as noted above, the theory of tucking-in must resort to the additional principle Merge before Move in order to derive the whole range of ordering effects in multiple specifiers. For the sake of argument, we grant here that this principle is well motivated.

[40] Empirically, the two theories make different predictions for mixed specifier domains where at least two specifiers are formed by Merge and one comes about by Move. In such a scenario, the tucking-in theory predicts that the movement in question targets the innermost specifier, due to Shortest Paths. The theory of EF insertion predicts that movement either targets the innermost specifier or a position in between two specifiers. The reason is that as long as the outermost specifier is yet to be merged, the head is still active, and therefore EF insertion can apply. The predictions are hard to test, though, because bona fide instances of multiple specifiers created by pure Merge are rather rare (if they exist at all).

First, tucking-in is incompatible with strict cyclicity: it involves movement to a position that forms a proper part of the current phrase marker, thus violating the SCC ((19) in section 3). Technically speaking, this violation is not effective for Richards (2001) because in this work a weaker notion of cyclicity is presupposed. The weaker notion is, of course, not violated by tucking-in. However, on purely conceptual grounds, strict cyclicity is to be preferred over weak cyclicity, following the general strategy that more restrictive theories should be preferred over less restrictive ones. As both the derivation of the ISC in (18) as well as the stack theory of parallel movement obey strict cyclicity, this is a first conceptual argument against tucking-in and in favor of the theory proposed here.[41]

Second, Richards' (2001) derivation of tucking-in relies on the transderivational constraint Shortest Paths.[42] Transderivational constraints are inherently more complex than local constraints in that they require the comparison of (parts of) derivations, which local constraints do not. Therefore, attempts have been made to replace transderivational constraints by local constraints (Chomsky 1995, Collins 1997, Frampton and Gutman 1999). It is not clear how the notion of Shortest Paths invoked in the derivation of tucking-in could be rephrased in terms of a local constraint, presumably because it comprises both one-to-many relations (between a head and the categories it attracts) and many-to-one relations (between various positions and a category that is supposed to move to one of them).[43] While we would not generally reject transderivational constraints as such, we believe that if an analysis can be shown to do without them, then it is to be preferred over one that cannot. The analysis in section 4 exclusively relies on local constraints.

Third, and perhaps most importantly, tucking-in does not fit into the strictly derivational account of opacity effects presented in section 4. To see this, reconsider the case of counter-bleeding as it shows up with PG-binding (12-a), here repeated as (53).

---

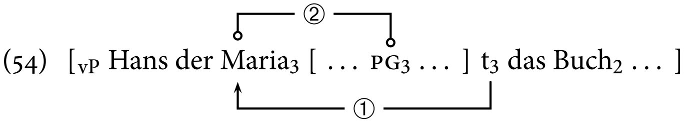
[41] Compare also Collins and Stabler (2011), who derive strict cyclicity as a theorem within their theory.

[42] The definition of Shortest Paths in Richards (2001, 98) differs from the one in (52), but it is still transderivational.

[43] Chomsky (1995) successfully replaces a weaker version of Shortest Paths, which is exclusively concerned with one-to-many relations, by the MLC. Crucially, this weaker version is not sufficient to fully derive tucking-in.

(53) dass Hans das Buch$_2$ der Maria [ohne PG$_2$ durchzulesen]
that Hans$_{nom}$ the book$_{acc}$ the Maria$_{dat}$ without through to read
zurückgibt
back gives
"that Hans returns the book to Maria without reading it through"

Given the background of the present analysis, a derivation of (53) based on tucking-in would involve an intermediate state where the indirect object has scrambled across the adjunct while the direct object still occupies its base position, waiting to tuck-in below the indirect object in the next step, see (54).



(54) [$_{vP}$ Hans der Maria$_3$ [… PG$_3$ …] t_3 das Buch$_2$ …]

In this configuration, nothing prevents Agree (and thus binding) to apply between the indirect object and the PG (see ② above). Consequently, binding of the PG by the direct object should be blocked: When the direct object undergoes tucking-in below the indirect object, the PG is already bound. Binding by the direct object, however, is exactly what is attested empirically in (53). In principle, the tucking-in theory could be complemented by an additional assumption to the effect that Agree does not apply until the current phrase (vP in (54)) has been completed, which would procrastinate Agree from being established with the PG until the direct object has undergone tucking-in. Once tucking-in has applied, one may argue, binding of the PG by the indirect object is blocked by the MLC because the direct object is closer to the PG than the indirect object.

Although this solution works technically, it runs against two important tenets of derivational syntax. First, it has been argued that Agree (and syntactic operations in general) should apply as early as possible once the feature that triggers the operation has been introduced (see section 4.1 for references). Procrastinating Agree in (54) is at variance with this assumption.[44] Second, Brody (2002) argues that the representational residue of derivational theories should be minimized. In the derivational account of opacity that we proposed, intervention effects as such do not exist; rather, they are simply a side effect of the way the derivation unfolds over time. For instance, binding of the PG

---

[44] Incidentally, the earliness requirement is also adopted by Richards (2001, 38-42), where it serves to derive the weaker version of cyclicity, on the basis of which tucking-in is derived.

by the indirect object in (53) is impossible simply because the direct object reaches a position from where it c-commands the PG and is thus able to bind it *before* the indirect object does. No reference to any particular representational relation between the two objects (such as intervention) is necessary to achieve this. In contrast, the alternative account based on tucking-in that procrastinates Agree makes reference to the MLC, which in turn refers to a particular representation in which the direct object intervenes between the indirect object and the PG. Thus, the present theory arguably reduces its representational residue relative to a comparable theory based on tucking-in. From this perspective, a theory without tucking-in seems more appropriate to account for order preservation effects within a derivational framework.

6. Conclusion

In this paper, we illustrated that syntactic relations between an associate and its antecedent are often opaque. The evidence came from association with floating quantifiers and binding of parasitic gaps in German. We argued that these instances of syntactic opacity and the asymmetries they involve can be given a derivational account within the probe-goal framework if (a) both the associate and its potential antecedents (the arguments of a verb) are merged in fixed positions, (b) multiple attraction of the potential antecedents by the same probe preserves their relative order, (c) association applies as soon as the structural context for its application is given, and (d) there is subsequent movement that renders the context of application opaque. To achieve this, we proposed that scrambling is triggered by EFs (subject to the ISC), and that order preservation is the result of collecting multiply attracted elements on a stack. As a result, the intervention effects with respect to antecedent and associate, which usually receive a representational treatment in terms of the MLC, were shown to follow without further ado in a strictly derivational fashion. Since the MLC introduces a representational residue, which should be minimized in derivational theories, and since it also poses various other problems (order preservation effects, anti-superiority, the analysis of multiple parasitic gaps), we argued that the MLC should be dispensed with. We made a proposal as to how the approach can be reconciled with apparent counter-evidence coming from *wh*-in-situ and from the claim that there is a minor class of verbs in German that project their objects in a non-standard order. It was argued that the analysis provides a novel argument to the effect that scrambling is to be analyzed as a transformation, and we finally suggested that the present account is suited better to account for order preservation effects within a derivational framework than the theory of tucking-in.

References

Adger, David (2003): *Core Syntax*. Oxford University Press, Oxford.

Agbayani, Brian and Masao Ochi (2007): Split Lexical Insertion in Parasitic Gap Constructions. Ms. University of California, Fresno and University of Osaka.

Anagnostopoulou, Elena (2003): *The Syntax of Ditransitives: Evidence from Clitics*. Studies in Generative Grammar, Mouton De Gruyter, Berlin.

Assmann, Anke (2012): Deriving Parasitic Gaps by Fission und Fusion. *In:* E. Boone, K. Linke and M. Schulpen, eds, *Proceedings of ConSOLE XIX*. Leiden, pp. 49–75.

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

Bailyn, John (2004): 'Generalized Inversion', *Natural Language and Linguistic Theory* 22, 1–50.

Baker, Mark (2008): *The Syntax of Agreement and Concord*. Cambridge University Press, Cambridge.

Bayer, Josef (1984): 'COMP in Bavarian Syntax', *The Linguistic Review* 3, 209–275.

Bayer, Josef and Jaklin Kornfilt (1994): Against Scrambling as an Instance of Move-*α*. *In:* N. Corver and H. Riemsdijk, eds, *Studies on Scrambling. Movement and Non-Movement Approaches to Free Word-Order Phenomena*. Mouton de Gruyter, Berlin, pp. 17–60.

Beck, Sigrid (1997): Wh-Constructions and Transparent Logical Form. PhD thesis, Universität Tübingen.

Beck, Sigrid (2006): 'Intervention Effects Follow from Focus Interpretation', *Natural Language Semantics* 14, 1–56.

Bennis, Hans and Teun Hoekstra (1984): 'Gaps and Parasitic Gaps', *The Linguistic Review* 4, 29–87.

Bierwisch, Manfred (1963): *Grammatik des Deutschen Verbs*. Studia Grammatica, Akademieverlag, Berlin.

Bobaljik, Jonathan (1995): Morphosyntax: The Syntax of Verbal Inflection. PhD thesis, MIT, Cambridge, Massachusetts.

Boeckx, Cedric (2003): *Islands and Chains*. John Benjamins, Amsterdam.

Boeckx, Cedric and Kleanthes K. Grohmann (2007): 'Putting Phases in Perspective', *Syntax* 10, 204–222.

Bošković, Željko (2002): 'A-Movement and the EPP', *Syntax* 5, 167–218.

Bošković, Željko (2004): 'Topicalization, Focalization, Lexical Insertion, and Scrambling', *Linguistic Inquiry* 35, 613–638.

Bošković, Željko (2007): 'Agree, Phases, and Intervention Effects', *Linguistic Analysis* 93, 54–96.

Bošković, Željko and Daiko Takahashi (1998): 'Scrambling and Last Resort', *Linguistic Inquiry* 29, 347–366.

Brody, Michael (2002): On the Status of Representations and Derivations. *In:* S. D. Epstein and T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 326–342.

Chomsky, Noam (1951): Morphophonemics of Modern Hebrew. Master's thesis, University of Pennsylvania, Philadelphia, Pennsylvania.

Chomsky, Noam (1973): Conditions on Transformations. *In:* S. Anderson and P. Kiparsky, eds, *A Festschrift for Morris Halle*. Holt, Reinhart and Winston, New York, pp. 232–286.

Chomsky, Noam (1982): *Some Concepts and Consequences of the Theory of Government and Binding*. MIT Press, Cambridge, Massachusetts.

Chomsky, Noam (1986): *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and J. Uriagereka, eds, *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. MIT Press, Cambridge, Massachusetts, pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. *In:* M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Massachusetts, pp. 1–52.

Chomsky, Noam (2007): Approaching UG from Below. *In:* U. Sauerland and H.-M. Gärtner, eds, *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Mouton de Gruyter, Berlin, pp. 1–30.

Chomsky, Noam (2008): On Phases. *In:* R. Freidin, C. P. Otero and M. L. Zubizarreta, eds, *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud*. MIT Press, Cambridge, Massachusetts, pp. 133–166.

Collins, Chris (1994): 'Economy of Derivation and the Generalized Proper Binding Condition', *Linguistic Inquiry* 25, 45–61.

Collins, Chris (1997): *Local Economy*. MIT Press, Cambridge, Mass.

Collins, Chris and Edward Stabler (2011): A Formalization of Minimalist Syntax. Ms., New York University and University of California Los Angeles.

Diesing, Molly (1992): *Indefinites*. MIT Press, Cambridge, Massachusetts.

Dikken, Marcel den and René Mulder (1991): 'Double Object Scrambling', *MIT Working Papers in Linguistics* 14, 67–82.

Engdahl, Elisabet (1983): 'Parasitic Gaps', *Linguistics & Philosophy* 6, 5–34.

Epstein, Samuel David (2007): 'On I(nternalist)-Functional Explanation in Minimalism', *Linguistic Analysis* 33, 20–53.

Fanselow, Gisbert (1990): Scrambling as NP-Movement. *In:* G. Grewendorf and W. Sternefeld, eds, *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 113–140.

Fanselow, Gisbert (1991): 'Minimale Syntax', *Groninger Arbeiten zur Germanistischen Linguistik* 32.

Fanselow, Gisbert (1993): 'Die Rückkehr der Basisgenerierer', *Groninger Arbeiten zur Germanistischen Linguistik* 36, 1–74.

Fanselow, Gisbert (2001): 'Features, θ-Roles, and Free Constituent Order', *Linguistic Inquiry* 32, 405–437.

Fanselow, Gisbert (2003): 'Free Constituent Order, A Minimalist Interface Account', *Folia Linguistica* 37, 191–231.

Fanselow, Gisbert and Denisa Lenertová (2011): 'Left Peripheral Focus: Mismatches between Syntax and Information Structure', *Natural Language and Linguistic Theory* 29, 169–209.

Felix, Sascha W. (1985): Parasitic Gaps in German. *In:* W. Abraham, ed., *Erklärende Syntax des Deutschen*. Narr, Tübingen, pp. 173–201.

Ferguson, K. Scott (1993): 'Notes on the Shortest Move Metric and Object Checking', *Harvard Working Papers in Linguistics* 3, 65–80.

Fitzpatrick, Justin (2006): The Syntactic and Semantic Roots of Floating Quantification. PhD thesis, MIT.

Fox, Danny and David Pesetsky (2005): 'Cyclic Linearization of Syntactic Structure', *Theoretical Linguistics* 31, 1–46.

Frampton, John and Sam Gutman (1999): 'Cyclic Computation', *Syntax* 2, 1–27.

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

Giusti, Giuliana (1990): 'Floating Quantifiers, Scrambling, and Configurationality', *Linguistic Inquiry* 21, 633–641.

Giusti, Giuliana (1991): The Syntax of Floating *alles* in German. *In:* W. Abraham, W. Kosmeijer and E. Reuland, eds, *Issues in Germanic Syntax*. Mouton de Gruyter, Berlin, pp. 327–350.

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

Grewendorf, Günther and Joachim Sabel (1999): 'Adjunction vs. Multiple Specifiers. On Scrambling in German and Japanese', *Natural Language and Linguistic Theory* 17, 1–65.

Haider, Hubert (1988): Θ-Tracking Systems – Evidence from German. *In:* L. Maràcz and P. Muysken, eds, *Configurationality*. Foris, Dordrecht, pp. 185–206.

Haider, Hubert (1992): Branching and Discharge. Arbeitspapiere des Sonderforschungsbereichs 340, Nr. 23, Universität Stuttgart.

Haider, Hubert (1993): *Deutsche Syntax – Generativ*. Narr, Tübingen.

Haider, Hubert (1997): Scrambling - Locality, Economy, and Directionality. *In:* S. Tonoike, ed., *Scrambling*. Kurosio Publishers, Tokio, pp. 61–91.

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

Haider, Hubert and Inger Rosengren (2003): 'Scrambling: Nontriggered Chain Formation in OV Languages', *Journal of Germanic Linguistics* 15, 203–267.

Hawkins, John (1978): *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. Croom Helm, London.

Heck, Fabian (2009): 'On Certain Properties of Pied-Piping', *Linguistic Inquiry* 40, 75–111.

Heck, Fabian and Gereon Müller (2003): 'Derivational Optimization of Wh-Movement', *Linguistic Analysis* 33, 97–148. Dynamic interfaces Part 1, Kleanthes Grohmann, ed.

Heck, Fabian and Gereon Müller (2007): Extremely Local Optimization. *In:* E. Bainbridge and B. Agbayani, eds, *Proccedings of WECOL 34*. University of Fresno, Fresno, CA, pp. 170–182.

Höhle, Tilman (1982): Explikationen für 'normale Betonung' und 'normale Wortstellung'. *In:* W. Abraham, ed., *Satzglieder im Deutschen*. Narr, Tübingen, pp. 75–153.

Hunter, Tim and Shiti Malhotra (2009): Reassessing the Argument for Tucking-In. Ms., University of Maryland.

Huybregts, Riny and Henk van Riemsdijk (1985): Parasitic Gaps and ATB. In: *Proceedings of NELS 15*. GLSA, Amherst, pp. 168–87.

Kathol, Andreas (2001): On the Nonexistence of True Parasitic Gaps in Standard German. *In:* P. W. Culicover and P. M. Postal, eds, *Parasitic Gaps*. MIT Press, Cambridge, Massachusetts, pp. 315–338.

Kenstowicz, Michael and Charles Kisseberth (1979): *Generative Phonology*. Academic Press, New York.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Kiss, Katalin É. (1994): Scrambling as the Base-Generation of Random Complement Order. *In:* N. Corver and H. Riemsdijk, eds, *Studies on Scrambling*. Mouton/de Gruyter, Berlin, pp. 221–256.

Kitahara, Hisatsugu (2002): Scrambling, Case and Interpretability. *In:* S. D. Epstein and T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 167–183.

Ko, Heejeong (2007): 'Asymmetries in Scrambling and Cyclic Linearization', *Linguistic Inquiry* 38, 49–83.

Koopman, Hilda (2006): Agreement Configurations: In Defense of Spec-Head. *In:* C. Boeckx, ed., *Agreement Systems*. John Benjamins, Amsterdam, pp. 159–199.

Koster, Jan (1978): *Locality Principles in Syntax*. Foris, Dordrecht.

Lasnik, Howard (1999): 'On Feature Strength: Three Minimalist Approaches to Overt Movement', *Linguistic Inquiry* 30, 197–217.

Lee, Young-Suk and Beatrice Santorini (1994): Towards Resolving Webelhuth's Paradox: Evidence from German and Korean. *In:* N. Corver and H. Riemsdijk, eds, *Studies on Scrambling*. Mouton de Gruyter, Berlin, pp. 257–300.

Lenerz, Jürgen (1977): *Zur Abfolge nominaler Satzglieder im Deutschen*. Narr, Tübingen.

Link, Godehard (1974): Quantoren-Floating im Deutschen. *In:* F. Kiefer and D. Perlmutter, eds, *Syntax und Generative Grammatik*. Athenaion, Frankfurt, pp. 105–127.

Lutz, Uli (2001): Studien zu Extraktion und Projektion im Deutschen. PhD thesis, Universität Tübingen.

Mahajan, Anoop (1990): The A/A-bar Distinction and Movement Theory. PhD thesis, MIT, Cambridge, Massachusetts.

McGinnis, Martha (1998): Locality in A-Movement. PhD thesis, MIT, Cambridge, Massachusetts.

Meinunger, André (2000): *Syntactic Aspects of Topic and Comment*. John Benjamins, Amsterdam.

Merchant, Jason (1996): Object Scrambling and Quantifier-Float in German. *In:* K. Kusumoto, ed., *Proceedings of NELS 26*. GLSA, Amherst, Massachusetts, pp. 179–193.

Miyagawa, Shigeru (2001): The EPP, Scrambling, and *Wh*-in-situ. *In:* M. Kenstowicz, ed., *Ken Hale: a Life in Language*. MIT Press, Cambridge, Massachusetts, pp. 293–338.

Müller, Gereon (1995): *A-bar Syntax: A Study in Movement Types*. Mouton de Gruyter, Berlin.

Müller, Gereon (2001): Order Preservation, Parallel Movement, and the Emergence of the Unmarked. *In:* G. Legendre, J. Grimshaw and S. Vikner, eds, *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Massachusetts, pp. 279–314.

Müller, Gereon (2004): Phrase Impenetrability and Wh-Intervention. *In:* A. Stepanov, G. Fanselow and R. Vogel, eds, *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 289–325.

Müller, Gereon (2010): 'On Deriving CED Effects from the PIC', *Linguistic Inquiry* 41, 35–82.

Müller, Gereon (2011): *Constraints on Displacement*. Benjamins, Amsterdam.

Müller, Gereon and Wolfgang Sternefeld (1994): Scrambling as A-bar Movement. *In:* N. Corver and H. v. Riemsdijk, eds, *Studies on Scrambling. Movement and Non-Movement Approaches to Free Word-Order Phenomena*. Mouton de Gruyter, Berlin, pp. 331–386.

Nakamura, Masanori (1998): Reference Set, MLC, and Parametrization. *In:* P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, eds, *Is the best good enough? Optimality and competition in syntax*. MIT Press, Cambridge, Massachusetts, pp. 291–313.

Neeleman, Ad (1994): Scrambling as a D-Structure Phenomenon. *In:* N. Corver and H. v. Riemsdijk, eds, *Studies on Scrambling. Movement and Non-Movement Approaches to Free Word-Order Phenomena*. Mouton de Gruyter, Berlin, pp. 387–430.

Nissenbaum, Jon (2000): Investigations of Covert Phrase Movement. PhD thesis, MIT, Cambridge, Massachusetts.

Pafel, Jürgen (1991): Zum relativen Quantorenskopus im Deutschen. Arbeitspapiere des Sonderforschungsbereichs 340, Nr. 5, Universität Tübingen.

Pesetsky, David (1987): Wh-in-Situ: Movement and Unselective Binding. *In:* E. Reuland and A. ter Meulen, eds, *The Representation of (In)definiteness*. MIT Press, Cambridge, Massachusetts, pp. 98–129.

Pesetsky, David (1989): Language Particular Processes and the Earliness Principle. Ms., MIT.

Pittner, Karin (1996): 'Zur morphologischen Defektivität des Pronomens *wer*', *Deutsch als Fremdsprache* 33, 73–77.

Reis, Marga (1992): The Category of Invariant *alles* in Wh-Clauses: On Syntactic Quantifiers vs. Quantifying Particles in German. *In:* R. Tracy, ed., *Who Climbs the Grammar Tree?*. Niemeyer, Tübingen, pp. 465–492.

Reis, Marga and Heinz Vater (1980): Beide. *In:* G. Brettschneider and C. Lehmann, eds, *Wege zur Universalienforschung*. Narr, Tübingen, pp. 365–391.

Richards, Marc (2004): Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax. PhD thesis, University of Cambridge.

Richards, Marc (2011): 'Deriving the Edge: What's in a Phase?', *Syntax* 14, 74–95.

Richards, Norvin (2001): *Movement in Language*. Oxford University Press, Oxford.

Riemsdijk, Henk van (1978): *A Case Study in Syntactic Markedness. The Binding Nature of Prepositional Phrases*. Foris, Dordrecht.

Rizzi, Luigi (1990): *Relativized Minimality*. MIT Press, Cambridge, Massachusetts.

Ross, John Robert (1967): Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, Massachusetts.

Safir, Ken (1987): 'The Anti-C-Command Condition on Parasitic Gaps', *Linguistic Inquiry* 18, 678–683.

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

Sells, Peter (2002): *Structure, Alignment, and Optimality in Swedish*. CSLI Publications, Stanford.

Sportiche, Dominique (1989): 'Le mouvement syntaxique: contraintes et paramètres', *Langages* 95, 35–80.

Stroik, Thomas (2009): *Locality in Minimalist Syntax*. MIT Press, Cambridge, Massachusetts.

Takahashi, Daiko (1994): Minimality of Movement. PhD thesis, University of Connecticut, Storrs.

Taraldsen, Tarald (1981): The Theoretical Interpretation of a Class of Marked Extractions. *In:* A. Belletti, L. Brandi and L. Rizzi, eds, *The Theory of Markedness in Generative Grammar*. Scuola Normale Superiore, Pisa, pp. 475–516.

Thiersch, Craig (1982): 'A Note on "Scrambling" and the Existence of VP', *Wiener Linguistische Gazette* 27/28, 83–95.

Unger, Christina (2010): A Computational Approach to the Syntax of Displacement and the Semantics of Scope. PhD thesis, Universiteit Utrecht.

Vikner, Sten (1989): 'Object Shift and Double Objects in Danish', *Comparative Germanic Syntax* 44, 119–141.

Vikner, Sten (1995): *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford University Press, Oxford.

Webelhuth, Gert (1992): *Principles and Parameters of Syntactic Saturation*. Oxford University Press, Oxford.

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

Wiltschko, Martina (1997): 'D-Linking, Scrambling, and Superiority in German', *Groninger Arbeiten zur Germanistischen Linguistik* 41, 107–142.

Zimmermann, Malte (2007): Quantifying Question Particles in German: Syntactic Effects on Interpretation. *In:* E. Puig-Waldmüller, ed., *Proceedings of Sinn und Bedeutung 11*. Universitat Ponpeu Fabra, Barcelona, pp. 627–641.

On Accelerating and Decelerating Movement: From Minimalist Preference Principles to Harmonic Serialism

Fabian Heck & Gereon Müller[*]

Abstract

Derivations in the minimalist programm (MP) frequently encounter competition between elementary operations. A way in MP to resolve such competitions is to assume preference principles, the most prominent being "Merge before Move" (Chomsky 2000). Assmann, Georgi et al. (2013) discuss competition between Agree and Move. There, either Agree must be procrastinated in favor of Move or the other way round. Procrastination, in turn, suggests constraint violability and thus an optimality theoretic account. While it is, in principle, possible to formulate an inviolable constraint such that it mimics the effects of two interacting violable constraints, such a move is arguably conceptually unattractive as it requires the constraint in question to be complex. Moreover, if it turns out that there are scenarios where the general preference expressed by the constraint is exceptionally reversed, then it has to be complicated further, shedding serious doubt on its plausibility. In this paper, we argue that scenarios of exceptional preference reversal do indeed exist. In optimality theory, they are captured straightforwardly by assuming a higher ranked constraint that brings about the change.

1. Introduction

Syntactic derivations in feature based theories such as the minimalist programm (Chomsky 1995, 2000, 2001) often involve competition between elementary operations. This means that there are derivational stages where more

---

[*] For comments and discussion, we would like to thank the audiences of the workshop on *Advances in Optimality Theory: Syntax and Semantics* (Johns Hopkins University, November 2012) and of the workshop on *Interaction of Syntactic Primitives* that formed part of the 35th DGfS conference (Universität Potdsdam, March 2013), and in particular Anke Assmann, Bob Frank, Doreen Georgi, Timo Klein, Geraldine Legendre, Mike Putnam, Jochen Trommer, and Philipp Weisser. The research documented here was supported by DFG grant MU 1444/8-1 (*Local Modelling of Non-Local Dependencies in Syntax*, Heck/Müller).

Rule Interaction in Grammar, 511–558
Fabian Heck & Anke Assmann (eds.)
Linguistische Arbeits Berichte 90, Universität Leipzig 2013

than one operation may in principle apply. Assuming a general earliness requirement (Pesetsky 1989, Chomsky 1995: 233, Lasnik 1999: 198), operations apply as soon as their context for application is present. But if there is no simultaneous rule application in grammar (see Epstein and Seely 2002; contra Pullum 1979, Chomsky 2008), then a conflict arises: More than one operation should apply immediately, yet only one of them can be executed at each step. Consequently, competition between operations can arise, which must be resolved by giving preference to (ranking) one or the other operation.

For instance, Chomsky (1995, 2000) observes that there are derivational stages where both Merge (external Merge, EM) and Move (internal Merge, IM) can in principle apply. Chomsky (1995, 2000) argues on the basis of contrasts such as (1) that there is a general preference to apply Merge before Move.

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

The embedded SpecT-position in (1) can be filled in two ways: Either *someone* moves or the expletive *there* is merged. If *there* is merged, then it can undergo movement to the matrix SpecT-position at some later step, deriving (1-a). If *someone* moves to the embedded SpecT first, then *there* will be merged into the higher SpecT later, deriving the ungrammatical (1-b). In order to block (1-b), Chomsky (2000) proposes the preference principle in (2).

(2) *Merge before Move*:
Suppose that the derivation has reached stage Σ_n, and Σ_{n+1} is a legitimate instance of Merge, and Σ'_{n+1} is a legitimate instance of Move. Then, Σ_{n+1} is to be preferred over Σ'_{n+1}.

The question arises as to whether the order of Merge and Move can be derived from more general assumptions about the make-up of these operations. Chomsky (2000) suggests that Merge is simpler than Move because Move might be Merge plus Agree, plus Pied Piping, etc., which should explain the preference in terms of economy. However, in contrast to this, Chomsky (2013) states that, if anything, Move should be simpler than Merge "since it requires vastly less search" because external Merge "must access the workspace of already generated objects and the lexicon". It is also worth noting that on the basis of Chomsky's (2000) assumptions about the complexity of Merge and Move (with Move emerging as less simple, i.e., more specific), the specificity-based preference principle for ordering operations postulated by Koutsoudas (1966, 1973) and Pullum (1979) (also see van Koppen 2005, Lahne 2012, and Georgi 2013 for similar more recent concepts) would in fact also predict a reverse Move before Merge outcoume. Perhaps the lack of an uncontroversial, obvious evaluation metric for ordering the two operations can be taken to indicate that both resolutions are in principle available in natural languages: Conflicts between elementary operations are resolved by ranking (giving preference to one of the two options), but there is no inherently fixed resolution strategy.[1]

Thus, sometimes the order of applying Merge and Move is underdetermined. The conflict can be resolved by ranking the requirements: The highest-ranked requirement is satisfied immediately; lower-ranked ones must remain unsatisfied at the current derivational step. Such unsatisfiability does not lead to a crash of the derivation and thus suggests an analysis in terms of violable constraints.

While (2) is (mildly) transderivational in nature, one may argue that it does not require violability of the constraint demanding the application of Move (which is procrastinated due to the preference principle): If the constraint does not require movement as such but rather the filling of some specifier position, then it can be equally well fulfilled by Merge. However, it has been argued that there is also competition between Agree and Move (see Assmann, Georgi et al. 2013), two operations for which it is less likely that their application is reducable to an identical trigger. Thus, in this case either Agree must be procrastinated in favor of Move or the other way round. Procrastination, in turn, presupposes constraint violability and thus suggests an optimality theoretic account. While it is, in principle, possible to formulate an inviolable constraint that mimics the effects of two interacting violable constraints, such a move is conceptually unattractive as it requires the constraint to be complex. What is more, if there are scenarios where the preference expressed by the complex constraint is exceptionally reversed, then the constraint must be further complicated, increasing the conceptual burden of the approach and thus rendering an alternative approach in terms of violable constraints more plausible.

---

[1] And perhaps there are also systematic conflicts and resolutions among different types of Move, as in intermediate vs. criterial movement steps; see Georgi (2013).

In section 2, we briefly report the analysis of Assmann, Georgi et al. (2013). There, it is suggested that resolving the competition between Move and Agree in morphologically ergative languages by giving preference to Move over Agree accounts for a restriction on ergative movement (on the TP cycle) in these languages. At the same time, the analysis explains the absence of a parallel restriction on accusative movement in morphologically accusative languages, where the competition between Move and Agree is resolved by giving Agree preference over Move.[2]

Based on this background, we then illustrate in section 3 that movement in accusative encoding systems (where Move is usually procrastinated in favor of Agree) is in some contexts *accelerated*, so that it applies before Agree. Further we illustrate in section 4 that movement in ergative encoding systems (where Move usually applies before Agree) is in some contexts *decelerated*, so that it applies after Agree. This accounts for a priori unexpected mobility restrictions on dative arguments in German, and for a priori unexpected movement options for ergative arguments in Chol, Basque, Avar, and Pitjantjatjara. The upshot will be that the effects illustrated in sections 3 and 4 can straightforwardly be derived in an optimality-theoretic approach (they signal the presence of more specific, higher-ranked constraints). However, as suggested above, they are less straightforwardly derivable in a more orthodox minimalist approach.

Finally note that the reasoning presented here presupposes an extremely local, derivational approach to optimization. In other words, the domain for optimization (conflict resolution) is the minimal derivational step (Epstein and Seely 2002, Heck and Müller 2007, 2013, McCarthy 2010). If the optimization domain is larger than the step-level, wrong empirical predictions are made.

2. Move vs. Agree: A Constraint on Ergative Movement

2.1. The Phenomenon

The starting point of Assmann, Georgi et al. (2013) is the observation that in many morphologically ergative languages ergative arguments (DP_{erg}) cannot undergo Ā-movement, i.e., they cannot undergo *wh*-movement, focussing, or relativization (see Campana 1992, Aldridge 2004, Stiebels 2006, Coon et al.

---

[2] The relevant rankings for ergative type languages and accusative type languages are independently motivated by the theory of argument encoding proposed in Müller (2009).

2011, Deal 2012). (3) briefly illustrates this for *wh*-movement in the Mayan language Kaqchickel.

(3) *Wh-movement of* DP_{erg} *vs.* DP_{abs} *in Kaqchikel (Mayan)*:

a. *achike n-Ø-u-löq' jun sik'iwuj?
Q INCOMPL-3SG.ABS-3SG.ERG-buy INDEF book
'Who buys a book?'

b. atux n-Ø-u-löq' a Carlos?
Q INCOMPL-3SG.ABS-3SG.ERG-buy CL Carlos
'What does Carlos buy?'

c. achike ri n-Ø-tze'en?
Q DET INCOMPL-3SG.ABS-laugh
'Who laughs?'

(3-a) involves *wh*-movement of an ergative subject, which is ungrammatical.[3] Nothing is wrong with *wh*-moving a subject per se, as the grammatical case of *wh*-movement of a absolutive marked subject in (3-c) illustrates. Finally, *wh*-movement of an absolutive marked object (3-b) is also impeccable.

Assmann, Georgi et al. (2013) propose an account of the restriction on Ā-movement of the ergative DP that is co-argument based. The leading idea is that there is nothing wrong with movement of the ergative marked external argument as such. Rather, movement of the ergative invariably leads to maraudage of the absolutive case provided for the internal argument, thus leaving the internal argument caseless, which leads to a violation of the case filter (Rouveret and Vergnaud 1980).

2.2. Theoretical Assumptions

To begin with, the clause structure in (4) in assumed in Assmann, Georgi et al. (2013).

(4) *Clause structure*:

$[_{\mathrm{CP}}\ \mathrm{C}\ [_{\mathrm{TP}}\ \mathrm{T}\ [_{\mathrm{vP}}\ \mathrm{DP}_{ext}\ [_{\mathrm{v'}}\ \mathrm{v}\ [_{\mathrm{VP}}\ \mathrm{V}\ \mathrm{DP}_{int}\ ]]]]]$

---

[3] Strictly speaking, argument DPs in Mayan languages do not bear overt case markers, but ergative and absolutive DPs trigger different kinds of agreement: DP_{ext} triggers ergative agreement whereas DP_{int} and the sole argument of an intransitive verb trigger absolutive agreement.

The internal argument DP_{int} is the complement of the verb. The external argument DP_{ext} is introduced as the specifier of the functional head v (Chomsky 1995, Kratzer 1996), which takes VP as its complement. There are two other functional heads above v, namely T and C.

Next, following Chomsky (1995), all operations are assumed to be feature-driven in Assmann, Georgi et al. (2013). Agree is triggered by probe features (5-a). Merge (external and internal) is triggered by subcategorization/edge features (Svenonius 1994, Sternefeld 2006, Chomsky 2007, 2008), see (5-b).

(5) *Two types of features that drive operations*:
  a. Probe features trigger Agree: [∗F∗].
  b. Subcategorization features/edge features trigger Merge: [•F•].

The definitions of the operations Merge, Move, and Agree (cf. Chomsky 2001), which are driven by these features, are given in (6)–(8), respectively. Note that some features may lack a value, which they must acquire by entering into Agree with another feature that bears a value (Chomsky 2000, 2001). A feature [F] that lacks a value is rendered as [F:□].

(6) *Merge*:
α can undergo merge with β, yielding [α β], if α bears a structure-building feature [•F•] and F is the label of β.

(7) *Move*:
Move is Merge, with β internal to α.

(8) *Agree*:
α agrees with β with respect to a feature bundle Γ iff (a) and (b) hold:
  a. α bears a probe feature [∗F∗] in Γ and may thereby provide the α-value for a matching goal feature [F] of β in Γ.
  b. α m-commands β.

Note that (8-b) permits an Agree relation between a head and its specifier. Incidentally, Assmann, Georgi et al. (2013) assume that Agree by a head H with its (innermost) specifier is not only possible but is actually preferred over Agree by H with any item bearing another structural relation towards H (see Chomsky 1986, 1995, Kayne 1989, Koopman 2006 for related proposals; Béjar and Řezáč 2009 express a similar idea with the bias inversed). This principle is dubbed the Specifier-Head Bias in Assmann, Georgi et al. (2013):

(9) Specifier-Head Bias (SHB):
Agree between (first) specifier and head is preferred to other instances of Agree.

To a certain extent, (9) replaces standard minimality conditions such as Relativized Minimality (Rizzi 1990) or the Minimal Link Condition (Fanselow 1991, Ferguson 1993, Chomsky 1995), though with a somewhat different empirical coverage. Müller (2004, 2011) argues that further effects usually accounted for by standard minimality conditions can be derived from the PIC (see below). At the same time, the SHB is compatible with equi-distance effects, which pose a problem for path-based, or closest c-command-based, definitions of minimality. It is therefore assumed in Assmann, Georgi et al. (2013) that minimality as such does not exist and that the effects traditionally attributed to it derive from independent principles (such as SHB and PIC).

The designated constraints in (10) and (11) ensure that Merge (incl. Move) and Agree must take place as soon as their context of application is present (Heck and Müller 2007, 2013).[4] This derives the earliness requirement for syntactic operations that was mentioned above.

(10) Agree Condition (AC):
Probes ([∗F∗]) participate in Agree.

(11) Merge Condition (MC):
Structure-building features ([•F•]) participate in Merge.

Every argument must receive structural case in the syntax (so as not to violate some form of the case filter). Structural case is assigned by the heads v and T under Agree. By standard assumption, T and v have valued case probe features that assign their value α to DPs with an unvalued case feature [case:□]. Assmann, Georgi et al. (2013) follow Murasugi (1992) (see also Jelinek 1993, Ura 2000, Müller 2009) in assuming that in ergative as well as in accusative

---

[4] In what follows, we will mainly be concerned with the interaction of Agree and Move, so for the most part, MC could just as well stand for Move Condition rather than for Merge Condition. Notwithstanding the considerations concerning possible orders of Merge and Move towards the end of section 1, assuming a more general Merge Condition comprising both operations might be argued to be conceptually preferable. More importantly in the present context, it also turns out to be crucial (at least for the data considered in this paper) when extraction options (i.e., phenomena related to Move) are closely tied to basic argument encoding patterns (i.e., phenomena related to Merge); see below.

languages T assigns the unmarked structural case (i.e., nominative = absolutive) and v assigns the marked structural case (i.e., ergative = accusative). In intransitive contexts only T is active, so the single argument receives the unmarked case.

More specifically, the assumption is that there is a single structural case feature [CASE]. This feature can have the two values ext(ernal) and int(ernal). The unmarked case (nominative/absolutive) is represented as the external case [CASE:*ext*] and the marked case (ergative/accusative) as the internal case [CASE:*int*]. Since T assigns unmarked external case and v assigns the marked internal case, these heads bear the following probe features:

(12) *The role of T and v in argument encoding*:
   a. T bears [*CASE:*ext**] that instantiates [CASE:*ext*] on DP.
   b. v bears [*CASE:*int**] that instantiates [CASE:*int*] on DP.

Turning to the issue of locality, Assmann, Georgi et al. (2013) suggest that movement that starts from within the vP-domain and targets SpecC must obligatorily make an intermediate movement step to SpecT. This is achieved by assuming that, generally, movement takes place successive-cyclically, from one XP edge domain to the next one higher up (see Sportiche 1989: 36, 45-47, Boeckx 2003: 16-25, Müller 2004, Chomsky 2005: 18, among others). Given the Phase Impenetrability Condition (PIC; Chomsky 2001) in (13) and the notion of edge in (14), this follows automatically if every XP is a phase.

(13) *Phase Impenetrability Condition (PIC; Chomsky 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.

(14) *Edge*:
   The edge of a head X comprises all specifiers of X (and adjuncts to XP).

Additionally, it must be ensured in a theory of syntax where all operations are feature-driven that intermediate steps of movement as required under the PIC are possible in the first place. A standard assumption here is that category neutral edge features ([•X•]) can be inserted on all intervening phase heads (Chomsky 2007, 2008). These edge features then trigger intermediate movement steps.

The following assumptions pertain to the activity of structural case features and the way that probe features and the goals they enter into Agree with are paired off. First, suppose that a structural case goal G can enter into Agree with a case probe P even if G has already acquired a value via Agree with another probe P′ at a previous stage of the derivation. That is, structural case features on arguments remain active throughout the derivation. This is explicitly stated in (15) (cf. Merchant 2006).

(15) *Activity of structural case features*:
Structural case features act as active goals.

Independent motivation for (15) comes from the existence of case stacking, as it exists in some of the world's languages (see Andrews 1996, Nordlinger 1998, Richards 2007, Merchant 2006).

Imagine now a situation where an argument with a structural case goal G enters into a Spec/head configuration with a functional head that bears a case probe P. Assume that P has not yet entered into Agree and that it is ultimately supposed to provide a value for another case goal G′. In this situation, P must enter into Agree with G (instead of G′) due to the Specifier-Head Bias, even though G has already acquired a case value while G′ has not. In this way, G *marauds* the functional head by taking away its case probe, which should normally be reserved for G′ (see Georgi et al. 2009 on maraudage; similar concepts are suggested in Chomsky 2001, Abels 2003, Anagnostopoulou 2005, Adger and Harbour 2007, Béjar and Řezáč 2009.)

The situation is abstractly depicted in (16). The configuration in (16-a) may involve Agree between [∗case:*ext*∗] on X and [case:*int*] on α or not. If Agree involves [case:*int*], this leads to a crash of the derivation because there remains an unchecked [case:□] on β. If, however, [∗case:*ext*∗] enters into Agree with [case:□] on β, then the derivation converges, which is sufficient to ensure grammaticality. The situation is different in (16-b). Here, α is in a Spec/head configuration with X. Thus, the SHB forces Agree between [∗case:*ext*∗] and [case:*int*], thereby leaving [case:□] on β without a value. It follows that (16-b) invariably leads to a crash.

(16) a. $[_{X'}\ X_{[*\text{case:ext}*]}\ [_{ZP} \ldots\ \alpha_{[\text{case:int}]} \ldots\ \beta_{[\text{case:}\Box]} \ldots\ ]]$
b. $[_{XP}\ \alpha_{[\text{case:int}]}\ [_{X'}\ X_{[*\text{case:ext}*]}\ [_{ZP} \ldots\ t_{\alpha} \ldots\ \beta_{[\text{case:}\Box]} \ldots\ ]]]$

Recall in this context Assmann, Georgi et al.'s (2013) assumption that there is no minimality condition. Given general activity of structural case goals, this is necessary to ensure convergence of at least one of the derivations based on (16-a). However, it is not sufficient. What is needed in addition is that both α and β are PIC-accessible to X in (16). This implies that the PIC is slightly less restrictive, as eventually proposed in Chomsky (2001), or that Agree operations can escape the PIC, as suggested by Bošković (2007), among others.

A question arises as to how multiple case agreement is possible. The presupposition behind (15) is that Agree of [CASE:*int*] on α with a conflicting [*CASE:*ext**] on X is harmless as such. α simply maintains its original feature value, which then accordingly surfaces in morphology; or α adds the new case feature, leading to case stacking (cf. Assmann, Edygarova et al. 2013). However, [*CASE:*ext**] on X is then discharged, and not available for further operations anymore.

Finally, Assmann, Georgi et al. (2013) adopt an idea put forward in Müller (2009) that ergative type languages vs. accusative type languages are distinguished by the relative ordering of Merge and Agree: The ranking MC ≫ AC derives ergative type encoding systems (by assigning the internal case of v to the external argument in Specv, due to the SHB which preferes Specv to CompV if both external and internal argument are present in the structure when AC needs to be satisfied), while the reversed ranking generates accusative type encoding systems (by assigning the internal case of v to the internal argument in the VP, which is the only DP requiring structural case that is present at the point of the derivation where AC must be satisfied under this ranking, with Merge of the external argument delayed).

In the following section, we introduce an optimality theoretic variant of the analysis of Assmann, Georgi et al. (2013), which our arguments in sections 3 and 4 will then be based upon.

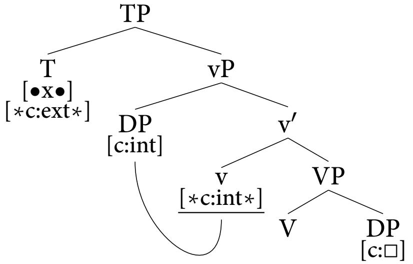
2.3. Analysis

2.3.1. *Displacement in Languages with Ergative Encoding Patterns*

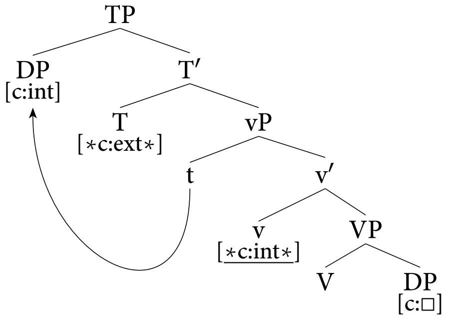
According to Müller (2009), the ordering conflict between Merge and Agree in a morphologically ergative language is resolved by the ranking MC ≫ AC. Imagine a situation where DP_{ext} is a *wh*-phrase that is supposed to undergo Ā-movement (*wh*-movement, relativization, focus movement) and ultimately show up in some SpecC-position. The details of deriving the ergative encoding system need not concern us here. It is sufficient to recall that v assigns internal case (= ergative). It follows that DP$_{ext}$ must have its case feature valued as ergative while it still resides in the m-command domain of v, i.e., upon completion of the vP. At this point, DP$_{int}$ still bears an unvalued case probe, awaiting valuation by T, see (17-a).[5]

Suppose now T is introduced into the structure. Given the PIC, DP$_{ext}$ needs to move from Specv to SpecT if it is to undergo subsequent movement to SpecC. Based on the null hypothesis that the ranking MC ≫ AC that leads to ergative type encoding systems on the vP-cycle is also maintained on the TP cycle, movement of DP$_{ext}$ (as an instance of internal Merge) will have to *precede* Agree of T with DP$_{int}$, which has not yet valued its case feature (as absolutive), see (17-b). The optimization of this derivational step is illustrated in tableau T_1.

(17) a. *Structure after T is merged:*



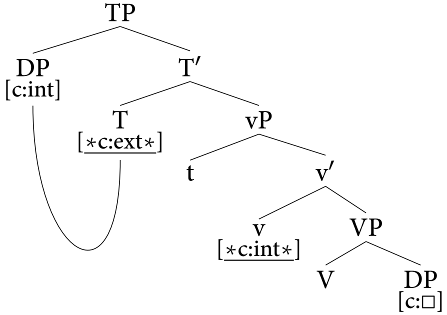
b. *MC ≫ AC triggers movement of DP$_{ext}$ first:*



---

[5] Here and henceforth, case probe features that have participated in Agree are signalled by underlining in trees; they are rendered here only so as to enhance perspicuity.

c. *Specifier-Head Bias triggers maraudage of T:*



Finally, given the SHB, DP_{ext} will next maraud T's case probe, see (17-c). The relevant optimization is given in tableau T_2. The internal argument DP will consequently remain without a valued case feature. Assuming that all DPs must have their case features valued eventually (and assuming that there is no such thing as a default case in a normal transitive clause where all arguments could in principle get their cases valued), the derivation will crash. In a nutshell, ergative movement is impossible because it applies too early, thereby bleeding absolutive case assignment to DP_{int}.

T_1: *Ergative movement, step 1: Move*

Input: $[_{\text{T}'}\ \text{T}_{[*\text{CASE}:\mathit{ext}*],[\bullet\text{X}\bullet]}\ [_{\text{vP}}\ \text{DP}_{[\text{CASE}:\mathit{int}]}\ [_{\text{v}'}\ \text{v} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots]]]]$	SHB	MC	AC
☞ O_1: $[_{\text{TP}}\ \text{DP}_{[\text{CASE}:\mathit{int}]}\ [_{\text{T}'}\ \text{T}_{[*\text{CASE}:\mathit{ext}*]}\ [_{\text{vP}}\ \text{t}\ [_{\text{v}'}\ \text{v} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots]]]]$			*
O_2: $[_{\text{T}'}\ \text{T}_{[\bullet\text{X}\bullet]}\ [_{\text{vP}}\ \text{DP}_{[\text{CASE}:\mathit{int}]}\ [_{\text{v}'}\ \text{v} \ldots \text{DP}_{[\text{CASE}:\mathit{int}]} \ldots]]]$		*!	

T_2: *Ergative movement, step 2: Agree (maraudage)*

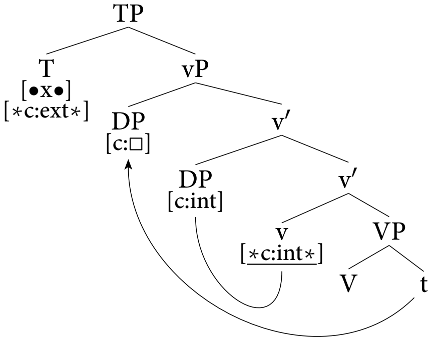
Input: $[_{\text{TP}}\ \text{DP}_{[\text{CASE}:\mathit{int}]}\ [_{\text{T}'}\ \text{T}_{[*\text{CASE}:\mathit{ext}*]}\ [_{\text{vP}}\ \text{t}\ [_{\text{v}'}\ \text{v} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots]]]]$	SHB	MC	AC
☞ O_1: $[_{\text{TP}}\ \text{DP}_{[\text{CASE}:\mathit{ext,int}]}\ [_{\text{T}'}\ \text{T}\ [_{\text{vP}}\ \text{t}\ [_{\text{v}'}\ \text{v} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots]]]]$			
O_2: $[_{\text{TP}}\ \text{DP}_{[\text{CASE}:\mathit{int}]}\ [_{\text{T}'}\ \text{T}\ [_{\text{vP}}\ \text{t}\ [_{\text{v}'}\ \text{v} \ldots \text{DP}_{[\text{CASE}:\mathit{ext}]} \ldots]]]]$	*!		

It is assumed here that a violation of the case filter eventually leads to a crash of the derivation. This means that an unvalued case feature represents an instance of Grimshaw's (1994) "no good output" approach to absolute ungrammaticality (or 'ineffability'): The optimal candidate is characterized by a property that gives rise to problems at the interfaces.

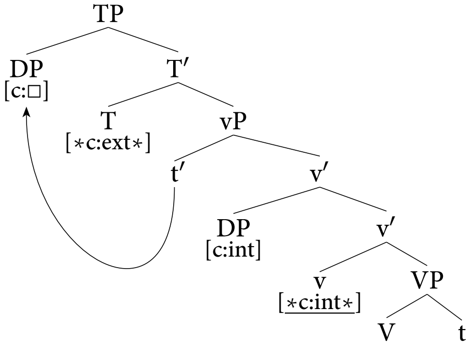
Next, consider the case where DP$_{int}$ undergoes Ā-movement. First, DP$_{ext}$ is merged in an inner Specv and DP$_{int}$ moves into an outer Specv, triggered by an edge feature inserted in v. These operations apply first, given the ranking MC ≫ AC. After this, DP$_{ext}$ enters into Agree with the case probe on v, thereby receiving ergative case (18-a).[6]

(18) *Legitimate movement of DP$_{abs}$:*

a. Structure after T is merged



b. Merge before Agree triggers movement of DP$_{abs}$ first

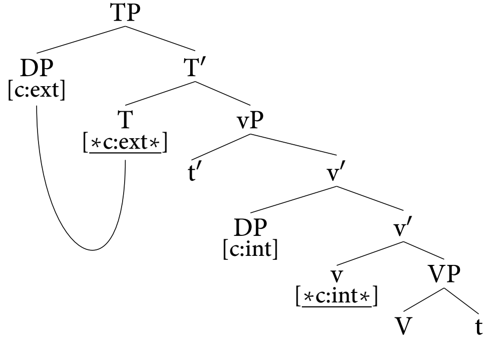


---

[6] The SHB blocks Agree between DP$_{int}$ and v. This raises the question as to what prevents DP$_{int}$ from occupying the innermost specifier (leaving the outer specifier for DP$_{ext}$) and thus receiving internal case. Such a derivation would wrongly lead to an accusative encoding pattern. To block it, Assmann, Georgi et al. (2013) assume a preference for Merge over Move.

If DP$_{int}$ is to remain accessible for further movement (to SpecC), it first has to raise to SpecT. MC ≫ AC forces this intermediate movement step to apply before Agree values absolutive case on DP$_{int}$, see (18-b). Tableau T3 illustrates the optimization. Finally, the case probe on T enters into Agree with the case feature on DP$_{int}$, valuing the latter as absolutive (18-c). As DP$_{ext}$ has already received its case value on the vP-level, the derivation converges. The optimization is shown in tableau T4.

(18) c. Finally, Agree with T ensures external case of DP$_{abs}$; no maraudage



T3: Absolutive movement, step 1: Move

Input: [$_{\text{T}'}$ T$_{[*\text{CASE}:\textit{ext}*],[\bullet\text{X}\bullet]}$ [$_{\text{vP}}$ DP$_{[\text{CASE}:\square]}$ … …[$_{\text{v}'}$ DP$_{[\text{CASE}:\textit{int}]}$ [$_{\text{v}'}$ v …t …]]]]	SHB	MC	AC
☞ O1: [$_{\text{TP}}$ DP$_{[\text{CASE}:\square]}$ [$_{\text{T}'}$ T$_{[*\text{CASE}:\textit{ext}*]}$ [$_{\text{vP}}$ t′ … …[$_{\text{v}'}$ DP$_{[\text{CASE}:\textit{int}]}$ [$_{\text{v}'}$ v …t …]]]]]			*
O2: [$_{\text{T}'}$ T$_{[\bullet\text{X}\bullet]}$ [$_{\text{vP}}$ DP$_{[\text{CASE}:\textit{ext}]}$ … …[$_{\text{v}'}$ DP$_{[\text{CASE}:\textit{int}]}$ [$_{\text{v}'}$ v …t …]]]]		*!	

T4: Absolutive movement, step 2: Agree (with SpecT)

Input: [$_{\text{TP}}$ DP$_{[\text{CASE}:\square]}$ [$_{\text{T}'}$ T$_{[*\text{CASE}:\textit{ext}*]}$ [$_{\text{vP}}$ t′ … …[$_{\text{v}'}$ DP$_{[\text{CASE}:\textit{int}]}$ [$_{\text{v}'}$ v …t …]]]]]	SHB	MC	AC
☞ O1: [$_{\text{TP}}$ DP$_{[\text{CASE}:\textit{ext}]}$ [$_{\text{T}'}$ T [$_{\text{vP}}$ t′ … …[$_{\text{v}'}$ DP$_{[\text{CASE}:\textit{int}]}$ [$_{\text{v}'}$ v …t …]]]]]			
O2: [$_{\text{TP}}$ DP$_{[\text{CASE}:\square]}$ [$_{\text{T}'}$ T [$_{\text{vP}}$ t′ … …[$_{\text{v}'}$ DP$_{[\text{CASE}:\textit{ext/int}]}$ [$_{\text{v}'}$ v …t …]]]]]	*!		

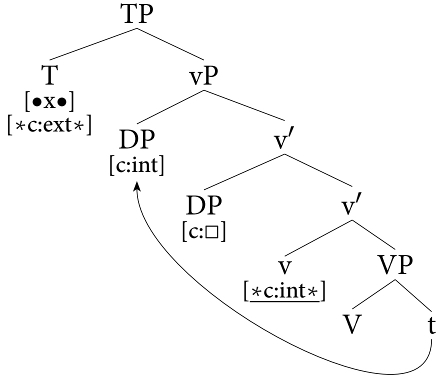
2.3.2. *Displacement in Languages with Accusative Encoding Patterns*

We now illustrate how the approach in Assmann, Georgi et al. (2013) accounts for the absence of a parallel restriction on movement of the accusative argument in morphologically accusative type languages.

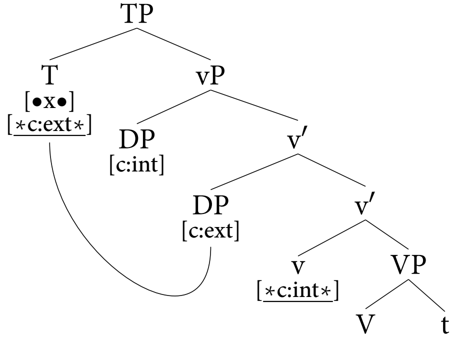
According to Müller (2009), the ranking in accusative type languages is AC ≫ MC. This ranking, giving rise to an accusative pattern in the first place (on the vP cycle), is also active on the TP cycle. Thus, in a derivation where DP_{int} is supposed to undergo extraction, it will target an outer specifier of vP after its case feature has been valued accusative by the probe on v (19-a). Once T is merged, AC ≫ MC ensures that case on DP_{ext} gets valued nominative *before* DP_{int} moves on to SpecT, see (19-b) and tableau T5.

(19) *Legitimate movement of* DP_{acc}:

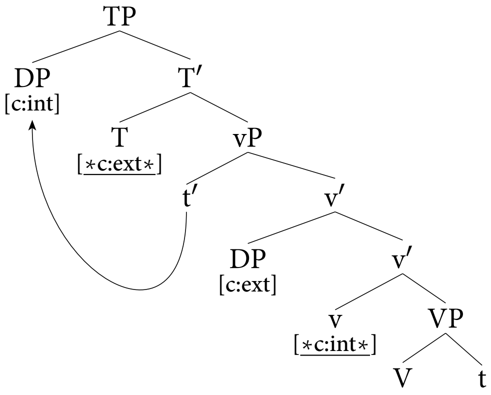
a. Structure after T is merged



b. No maraudage: AC ≫ MC triggers case valuation of DP_{ext} next



c. Finally, movement of DP_{int} takes place to SpecT



Finally, in the last step DP_{int} moves on to SpecT to satisfy an edge feature on the T-head. This is shown in (19-c) and tableau T_6, respectively.

T_5: Accusative Movement, step 1: Agree

Input: $[_{T'}\ T_{[*\text{CASE}:ext*],[\bullet X\bullet]}\ [_{vP}\ DP_{[\text{CASE}:int]} \dots$ $\dots [_{v'}\ DP_{[\text{CASE}:\square]}\ [_{v'}\ v \dots t \dots]]]]$	SHB	AC	MC
O1: $[_{TP}\ DP_{[\text{CASE}:int]}\ [_{T'}\ T_{[*\text{CASE}:ext*]}\ [_{vP}\ t' \dots$ $\dots [_{v'}\ DP_{[\text{CASE}:\square]}\ [_{v'}\ v \dots t \dots]]]]]$		*!	
☞ O2: $[_{T'}\ T_{[\bullet X\bullet]}\ [_{vP}\ DP_{[\text{CASE}:int]} \dots$ $\dots [_{v'}\ DP_{[\text{CASE}:ext]}\ [_{v'}\ v \dots t \dots]]]]$			*

T_6: Accusative Movement, step 2: Move

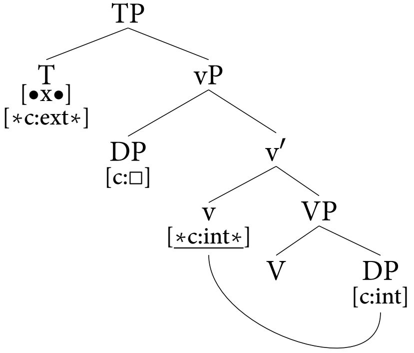
Input: $[_{T'}\ T_{[\bullet X\bullet]}\ [_{vP}\ DP_{[\text{CASE}:int]} \dots$ $\dots [_{v'}\ DP_{[\text{CASE}:ext]}\ [_{v'}\ v \dots t \dots]]]]$	SHB	AC	MC
☞ O1: $[_{TP}\ DP_{[\text{CASE}:int]}\ [_{T'}\ T\ [_{vP}\ t' \dots$ $\dots [_{v'}\ DP_{[\text{CASE}:ext]}\ [_{v'}\ v \dots t \dots]]]]]$			
☞ O2: $[_{TP}\ DP_{[\text{CASE}:ext]}\ [_{T'}\ T\ [_{vP}\ DP_{[\text{CASE}:int]} \dots$ $\dots [_{v'}\ t'\ [_{v'}\ v \dots t \dots]]]]$			

Nothing so far rules out O_2 in T_6. However, because of the PIC, only DP_{ext} can move on. Eventually, this leads to unchecked operator features on the attracting head and DP_{int}, and thus to a crash of the derivation.

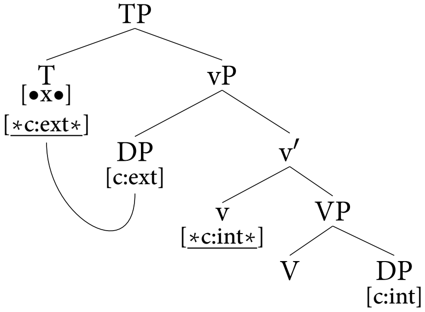
Finally consider a derivation involving Ā-movement of DP_{ext}. Similarly to movement of DP_{int} in ergative type systems, there is no problem for movement of DP_{ext} in accusative type systems because DP_{int} has already been assigned case when DP_{ext} moves. As (20-a) shows, the case feature of DP_{int} is valued as accusative already within vP. When T has been merged, AC ≫ MC dictates valuation of the case feature of DP_{ext} to apply before movement of DP_{ext}, see (20-b) and tableau T_7. Finally, when all cases have been valued, DP_{ext} moves to SpecT, see (20-c) and tableau T_8. From there, it can move on to SpecC.

(20) *Legitimate movement of* DP_{ext}:

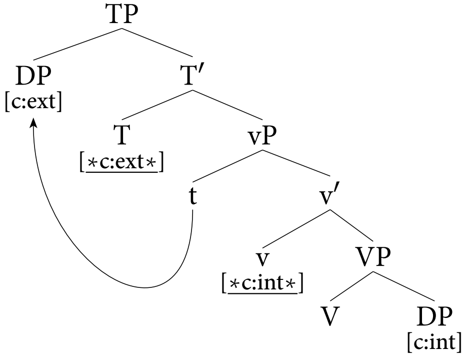
a. Structure after T is merged



b. AC ≫ MC triggers valuation of DP_{ext} next



c. Finally, movement of DP$_{ext}$ takes place to SpecT



T7: Nominative movement, step 1: Agree

Input: $[_{T'}\ T_{[*\text{CASE}:\mathit{ext}*],[\bullet X\bullet]}\ [_{vP}\ DP_{[\text{CASE}:\square]} \ldots \ldots [_{v'}\ v \ldots DP_{[\text{CASE}:\mathit{int}]} \ldots]]]$	SHB	AC	MC
O$_1$: $[_{TP}\ DP_{[\text{CASE}:\square]}\ [_{T'}\ T_{[*\text{CASE}:\mathit{ext}*]}\ [_{vP}\ t \ldots \ldots [_{v'}\ v \ldots DP_{[\text{CASE}:\mathit{int}]} \ldots]]]]$		*!	
☞ O$_2$: $[_{T'}\ T_{[\bullet X\bullet]}\ [_{vP}\ DP_{[\text{CASE}:\mathit{ext}]} \ldots \ldots [_{v'}\ v \ldots DP_{[\text{CASE}:\mathit{int}]} \ldots]]]$			*

T8: Nominative movement, step 2: Move

Input: $[_{T'}\ T_{[\bullet X\bullet]}\ [_{vP}\ DP_{[\text{CASE}:\mathit{ext}]} \ldots \ldots [_{v'}\ v \ldots DP_{[\text{CASE}:\mathit{int}]} \ldots]]]$	SHB	AC	MC
☞ O$_1$: $[_{TP}\ DP_{[\text{CASE}:\mathit{ext}]}\ [_{T'}\ T_{[\bullet X\bullet]}\ [_{vP}\ t \ldots \ldots [_{v'}\ v \ldots DP_{[\text{CASE}:\mathit{int}]} \ldots]]]]$			

2.4. Extremely Local vs. Less Local Optimization

It is crucial for the analysis that optimization applies to the single derivational step. If the optimization domain is not the derivational step but rather comprises phrases (phases, clauses, sentences), then a wrong prediction is made for accusative contexts: Maraudage would be expected to arise, and thus one would expect (at least some) morphologically accusative languages to exhibit a restriction on Ā-movement of the accusative argument, parallel to the restriction that shows up in many morphologically ergative languages. To wit, if optimization applies at the phrase level, then the order of operations induced by the accusative type ranking AC ≫ MC is lost: The optimal TP will always have its specifier filled by DP_{int} before DP_{ext} has been assigned case by T, and thus SHB will force Agree between $[*\text{CASE}:ext*]$ on T and DP_{int}, and make case assignment to DP_{ext} impossible. This is shown in tableau T9.

T9: TP optimization under AC ≫ MC ("accusative") ranking: wrong result

Input: $\text{T}_{[*\text{CASE}:ext*],[\bullet\text{X}\bullet]} \oplus [_{\text{vP}}\ \text{DP}_{[\text{CASE}:int]}\ [_{\text{v}'}\ \text{DP}_{[\text{CASE}:\square]} \ldots$ $\ldots [_{\text{v}'}\ \text{v} \ldots \text{t} \ldots]]]$	SHB	AC	MC
O_1: $[_{\text{TP}}\ \text{DP}_{[\text{CASE}:int]}\ [_{\text{T}'}\ \text{T}\ [_{\text{vP}}\ \text{t}'\ [_{\text{v}'}\ \text{DP}_{[\text{CASE}:ext]} \ldots$ $\ldots [_{\text{v}'}\ \text{v} \ldots \text{t} \ldots]]]]]$	*!		
☛ O_2: $[_{\text{TP}}\ \text{DP}_{[\text{CASE}:ext/int]}\ [_{\text{T}'}\ \text{T}\ [_{\text{vP}}\ \text{t}'\ [_{\text{v}'}\ \text{DP}_{[\text{CASE}:\square]} \ldots$ $\ldots [_{\text{v}'}\ \text{v} \ldots \text{t} \ldots]]]]]$			

Finally, note that the derivation with extraction of the accusative includes a stage that represents an interesting case of opacity, namely counter-bleeding (Chomsky 1951, 1975, Kiparsky 1976): When the moved accusative DP_{int} occupies SpecT, one would expect it to maraud T's case probe, thereby bleeding nominative case valuation of DP_{ext}. However, no such bleeding takes place. The reason is, of course, that nominative case valuation already took place at a previous step in the derivation. The interesting aspect of this instance of counter-bleeding is that it cannot be accounted for representationally by postulating abstract items (like traces). It therefore provides a good argument in favor of a derivational grammar.

Having presented an optimality theoretic version of the analysis proposed in Assmann, Georgi et al. (2013), we are now in a position to move on to the central argument of the present paper. So far, the empirical evidence and theoretical analyses are compatible both with postulating (parametrized) preference principles like *Merge before Agree* and *Agree before Merge* (as in Assmann, Georgi et al. 2013), and with postulating local optimization involving parametrized rankings of violable AC and MC constraints (as in our reconstruction in this section). In what follows, we are going to propose that the respective ranking established for ergative type languages and accusative type languages can be overwritten in particular contexts. We suggest that in both cases this happens in order to satisfy a higher ranked, more specific requirement. This is exactly what one would expect under an optimality-theoretic account, but it comes as a surprise under a preference principle-based analysis.

3. Accelerating Move: A Constraint on Dative Movement in Accusative Systems

The first argument concerns the reversal of the general preference for Agree over Move. It is based on a restriction against movement of dative arguments out of ECM-complements in German. The idea is that movement of the dative applies too early, namely before accusative case agreement can apply, thereby creating problems for the co-argument of the dative.

3.1. Data

It is a long-standing observation in the literature on German syntax that extraction of dative arguments out of ECM-complements leads to ungrammaticality (see Höhle 1978: 56-57, Thiersch 1978: 168-169, Fanselow 1986: 4, Grewendorf 1989: 150, Fanselow 1990: 121). This is illustrated for different contexts and movement types in (21)–(23).

(21) *Scrambling and pronoun movement of a DP$_{dat}$ object from ECM complements:*

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

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

c. *weil mir$_1$ niemand [$_{XP}$ Karl t$_1$ helfen ließ]
because me$_{dat}$ no-one$_{nom}$ Karl$_{acc}$ help let

(22) *Wh-movement and topicalization of a DP$_{dat}$ object from ECM complements:*

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

b. *Dem Lehrer$_1$ sah/ließ Karl [$_{XP}$ den Jungen t$_1$ helfen]
the teacher$_{dat}$ saw/let Karl$_{nom}$ the boy$_{acc}$ help

(23) *Movement of DP$_{dat}$ from ECM complements with double object constructions:*

a. *Wem$_1$ ließ/sah Karl [$_{XP}$ den Jungen t$_1$ das Buch geben]?
whom$_{dat}$ let/saw Karl the boy$_{acc}$ the book$_{acc}$ give

b. *dass keiner dieser Frau$_1$ [$_{XP}$ den Jungen t$_1$ das Buch
that no-one$_{nom}$ this woman$_{dat}$ the boy$_{acc}$ the book$_{acc}$
geben] ließ/sah
give let/saw

c. *dass er ihm$_1$ [$_{XP}$ den Jungen t$_1$ das Buch geben] ließ/sah
that he$_{nom}$ him$_{dat}$ the boy$_{acc}$ the book$_{acc}$ give let/saw

It is clear that the ungrammaticality of the previous examples cannot be attributed to the extraction of DP$_{dat}$ as such: In other contexts, movement of a dative argument produces impeccable results, see (24). (24-a) involves wh-movement from a finite clause (embedded by a bridge verb); (24-b) is a case of topicalization from a non-restructuring (i.e., fully clausal) infinitive; and (24-c) instantiates scrambling from a restructuring infinitive (that we here assume to be a vP).

(24) *Legitimate movement of DP$_{dat}$ in other contexts*:

a. Wem$_1$ meint sie [$_{CP}$ dass wir t$_1$ das Buch geben sollten]?
whom$_{dat}$ thinks that we the book give should

b. Diesem Plan$_1$ habe ich abgelehnt [$_{CP}$ PRO t$_1$ meine
this plan$_{dat}$ have I rejected my
Unterstützung zu geben]
support to give

c. dass ihm$_1$ der Fritz [$_{vP}$ t$_1$ das Buch zu geben] versuchte
that him$_{dat}$ the Fritz the book to give tried

Here, we would like to put forward the hypothesis that this restriction has the same source as the ban on ergative movement in morphologically ergative languages: In all of the cases (21)–(23), the dative argument moves too early, and thus marauds the matrix v's [∗CASE:int∗] feature. This ultimately precludes accusative case assignment to the ECM subject, which consequently leads to a crash of the derivation.

But there is a complication: German is an accusative language and therefore exhibits the ranking AC ≫ MC, which would normally order case assignment of v to the embedded DP$_{ext}$ before an intermediate movement step of the dative DP to matrix Specv. In order to overcome this problem, we propose that movement of the dative DP is exceptionally *accelerated* by a higher-ranked constraint in this particular context. The constraint in question is one that regulates proper and improper movement.

Traditionally, the notion of improper movement is meant to cover instances of a composite movement that decomposes into smaller movements that apply in a particular order, each targeting positions of different types. A classical case is movement that first targets a SpecC-position and then a SpecT-position (called super-raising), as illustrated for English in (25-b).

(25) *Raising vs. Super-Raising in English*:
  a. Mary$_1$ seems [$_{TP}$ t$_1$ to like John]
  b. *Mary$_1$ seems [$_{CP}$ t$_1'$ that t$_1$ likes John]

In contrast, movement from one SpecT-position to another, as in (25-a), is unproblematic (hence an instance of proper movement). A case of improper movement from German involves movement to SpecC followed by movement to a scrambling position, presumably a specifier of vP, resulting in long-distance scrambling (26-b). In contrast, movement to an outer Specv from within the VP of the same clause is unproblematic in German (26-a).

(26) *Long-Distance Scrambling in German*:
  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

Interestingly, the assumption that the PIC forces movement to SpecC to proceed via Specv creates a representation in the context of non-clause bound movement that superficially bears the hallmark of improper movement as it arises in long-distance scrambling: Compare (27-a,b), which both involve a local movement step from SpecC to Specv.

(27) *A dilemma for improper movement, given the PIC*:

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

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

Thus, any theory of improper movement which exclusively concentrates on this local step faces the problem of how to account for the difference in grammaticality between (28-a) and (28-b); see Neeleman and van de Koot (2010, 346-347), Bader (2011, ch. 5), and Müller (2013, sect. 2.6.3).

In response to this problem, a new version of a standard theory of improper movement is proposed in Müller (2013). In what follows, we would like to suggest that this approach provides the constraint that is responsible for the early movement of the dative in German ECM-contexts.

3.2. Assumptions

3.2.1. *Improper Movement*

The assumptions about improper movement made in Müller (2013) are the following. First, it is assumed that edge features are defective copies of categorial features of phase heads. When an edge feature attracts some category, it values a movement-related feature on this category. In this way, successive cyclic movement triggered by edge features creates a list on the moved item that records aspects of the derivational history of its movement. The information on the list is deleted when information of the same type is encountered in the course of the movement. Finally, there is a constraint to the effect that if the moved item reaches a criterial landing site, then the functional sequence of categories (*f-seq*: C-T-v-V) must be respected on the list containing the history of the movement steps performed by the item so far (cf. Williams 1974, 2003). This constraint will be called the WILLIAMS CYCLE (an explicit formulation of the WILLIAMS CYCLE will be given in section 3.2.2).

To illustrate the mechanics of this, consider the contrast between legitimate long *wh*-movement in (28) and illegitimate long-distance scrambling in (29).

(28) *Legitimate long-distance wh-movement*:
What$_2$ do you think [$_{\text{CP}}$ C [$_{\text{TP}}$ she$_1$ T [$_{\text{vP}}$ t$_1$ v [$_{\text{VP}}$ said t$_2$]]]]?
[$_{\text{CP}}$ what$_{[\text{wh:}\boxed{\text{CTvV}\cancel{\text{CTvV}}}]}$ [$_{\text{C}'}$ C [$_{\text{TP}}$ you think she said]]] (✓f-seq)

By assumption, movement of the *wh*-phrase in (28) proceeds through the specifiers of all phrases on the path to the matrix SpecC. In the embedded clause, the *wh*-phrase collects categorial information of all intervening phrase boundaries encountered there, resulting in the partial list C-T-v-V. Movement within the matrix clause creates the same sequence on the list again, leading to successive deletion of each of the elements on the list collected in the embedded clause. When the *wh*-phrase reaches its criterial position, the matrix SpecC, the list exclusively contains the categorial information collected within the matrix clause: C-T-v-V. Since this sequence is conform with f-seq, the Williams Cycle is satisfied.

In principle, the derivation of long-distance scrambling proceeds along the same lines. The difference, however, is that the categorial information collected in the embedded clause is not fully mirrored by the information collected in the matrix clause. As a consequence, when the criterial Specv-position is reached not all of the items on the list stemming from the embedded clause have been deleted (29), and the remaining list thus does not conform to f-seq.[7] As a consequence, the Williams Cycle is violated and ungrammaticality results.

(29) *Illegitimate long-distance scrambling:*
*dass Karl das Buch glaubt [$_{\text{CP}}$ dass keiner t$_1$ liest]
that Karl$_{nom}$ the book$_{acc}$ thinks that no-one$_{nom}$ reads
[$_{\text{CP}}$ [$_{\text{C}}$ dass] [$_{\text{TP}}$ Karl [$_{\text{T}'}$ [$_{\text{vP}}$ [$_{\text{v}'}$ [$_{\text{DP}}$ das Buch]$_{[\Sigma:\boxed{\text{vVCT}\cancel{\text{vV}}}]}$ [$_{\text{v}'}$ [$_{\text{v}'}$ glaubt dass keiner liest]]]] T]]] (*f-seq → *Williams Cycle → crash)

---

[7] It is assumed here that scrambling is ultimately triggered by a feature Σ on v, see Grewendorf and Sabel (1999), Sauerland (1999).

3.2.2. *Exceptional Case Marking in German*

German ECM complements lack typical properties associated with TP (or CP) (von Stechow and Sternefeld 1988, Fanselow 1991, Wurmbrand 2001). For instance, they do not host any separate temporal specification, there is obligatory wide scope for negation, and there is a systematic absence of *zu* ('to'), see (30-a-c). The conclusion from this is that XP in the examples (21)–(23) is vP.

(30) *German ECM complements are vPs*:
a. *Wir sehen [$_{\text{vP}}$ ihn den Diener erschossen haben]
we see him_{acc} the servant_{acc} shot have
b. Wir lassen [$_{\text{vP}}$ den Diener den Mann nicht schlagen]
we let the servant_{acc} the man_{acc} not hit
'We do not make/allow the servant (to) hit the man.'
c. *Wir hören [$_{\text{vP}}$ ihn zu schnarchen]
we hear him_{acc} to snore

3.3. Analysis

Against this background, consider now the analysis of extraction of a dative-marked argument out of ECM complements in German. In this configuration, matrix v has a dual role (see (31-a)): First, it assigns accusative case to the ECM subject (i.e., it bears [∗CASE:*int*∗]), and secondly, it has an (edge) feature to effect the (intermediate) movement step. Thus, the derivation faces the familiar conflict between Agree and Move on the vP cycle.

Under the accusative type ranking AC ≫ MC, this conflict is expected to be resolved by giving preference to Agree over Move. But note now that a moved dative DP originating in the embedded ECM complement (and having been assigned lexical case there by V) has a chance on the matrix vP-cycle to immediately remedy temporary f-seq violations on the feature list of its movement-related feature, and to thereby satisfy the WILLIAMS CYCLE (WC) quickly.[8] The specific version of WC that is required for this to happen is given in (31). It is formulated such that a temporary violation can be initiated

---

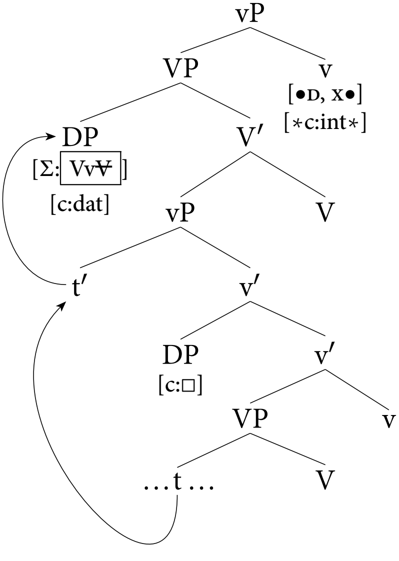
[8] For this to be the case, it is crucial that ECM complements in German are vPs, not TPs (or V clusters, for that matter).

without violating WC, which must be possible given that WC is ranked high (above AC).[9]

(31) WILLIAMS CYCLE (WC):
If categorial information on a list of a movement-related feature does not conform to f-seq (C-T-v-V) in the input, it must conform to f-seq in the output.

Assuming that WC in (31) outranks AC in German, movement of the dative DP to Specv will have to *precede* case assignment by v to the embedded DP_{ext} in German (32-b). The competition is shown in tableau T_{10}. Together with the SHB, this gives rise to maraudage of v's case feature [*CASE:*int**], see (32-c) and tableau T_{11}, and the derivation will ultimately crash because the embedded DP_{ext}'s case feature remains permanently unvalued.

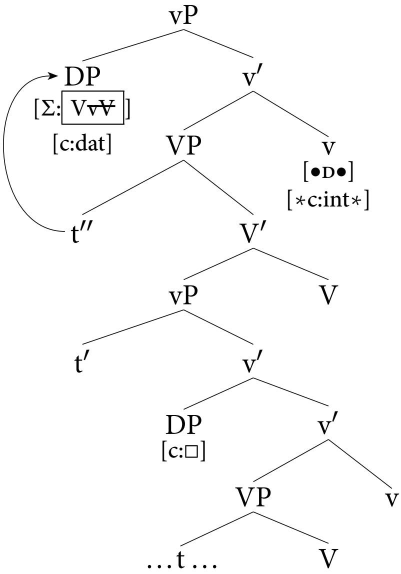
(32) *Illegitimate movement of* DP_{dat} *from ECM complements:*
   a. Structure after matrix v is merged; DP_{dat} almost satisfies f-seq



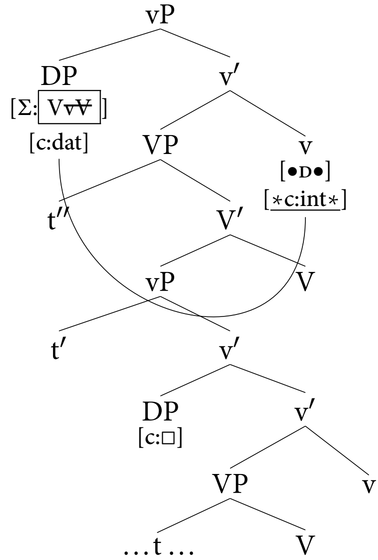
---

[9] Note that there is a certain similarity with anti-faithfulness constraints in phonology here; see Alderete (2001). Also see Baković and Wilson (2000) on targeted constraints.

b. WC ≫ AC ≫ MC triggers movement of DP$_{dat}$ to Specv



c. SHB triggers maraudage of v



T_{10}: Dative movement in ECM contexts, step 1: Move

Input: $\text{DP}_{ext} \oplus [_{\text{v}'}\ [_{\text{VP}}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{vP}} \ldots \text{DP}_{[\text{CASE}:\square]} \ldots \ldots \text{v}\ ]]\ \text{v}_{[*\text{CASE}:int*],[\bullet \text{X} \bullet],[\bullet \text{D} \bullet]}]$	SHB	WC	AC	MC
☞ O_1: $[_{\text{v}'}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{v}'}\ [_{\text{VP}}\ t''\ [_{\text{vP}} \ldots \text{DP}_{[\text{CASE}:\square]} \ldots \ldots \text{v}\ ]]\ \text{v}_{[*\text{CASE}:int*],[\bullet \text{D} \bullet]}]]$			*	*
O_2: $[_{\text{v}'}\ [_{\text{VP}}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{vP}} \ldots \text{DP}_{[\text{CASE}:int]} \ldots \ldots \text{v}\ ]]\ \text{v}_{[\bullet \text{X} \bullet],[\bullet \text{D} \bullet]}]$		*!		**
O_3: $[_{\text{vP}}\ \text{DP}_{ext}\ [_{\text{v}'}\ [_{\text{VP}}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{vP}} \ldots \text{DP}_{[\text{CASE}:\square]} \ldots \ldots \text{v}\ ]]\ \text{v}_{[*\text{CASE}:int*],[\bullet \text{X} \bullet]}]]$		*!	*	*

T_{11}: Dative movement in ECM contexts, step 2: Agree (maraudage)

Input: $\text{DP}_{ext} \oplus [_{\text{v}'}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{v}'}\ [_{\text{VP}}\ t''\ [_{\text{vP}} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots \text{v}\ ]]\ \text{v}_{[*\text{CASE}:int*],[\bullet \text{D} \bullet]}]]$	SHB	WC	AC	MC
☞ O_1: $[_{\text{v}'}\ \text{DP}_{[\text{CASE}:int/dat]}\ [_{\text{v}'}\ [_{\text{VP}}\ t''\ [_{\text{vP}} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots \text{v}\ ]]\ \text{v}_{[\bullet \text{D} \bullet]}]]$				*
O_2: $[_{\text{v}'}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{v}'}\ [_{\text{VP}}\ t''\ [_{\text{vP}} \ldots \ldots \text{DP}_{[\text{CASE}:int]} \ldots \text{v}\ ]]\ \text{v}_{[\bullet \text{D} \bullet]}]]$	*!			*
O_3: $[_{\text{vP}}\ \text{DP}_{ext}\ [_{\text{v}'}\ \text{DP}_{[\text{CASE}:dat]}\ [_{\text{v}'}\ [_{\text{VP}}\ t''\ [_{\text{vP}} \ldots \ldots \text{DP}_{[\text{CASE}:\square]} \ldots \text{v}\ ]]\ \text{v}_{[*\text{CASE}:int*]}]]]$			*!	

3.4. Extremely Local vs. Less Local Optimization

This time, assuming larger optimization domains like the phrase does not make a wrong prediction: At the vP phrase level, WC, MC and AC are all satisfied, and SHB will continue to pick a maraudage output. Of course, the analysis is nevertheless also compatible with an approach where optimization applies at the step-level, as illustrated.

3.5. Consequences

The approach makes at least one interesting additional prediction, but it also raises various questions. In what follows, these issuess are briefly addressed.

First, if there is no embedded DP_{ext} in what is otherwise the same construction, then the prediction is that movement of the dative DP should be fine because there is no external co-argument DP that could violate the case filter after early (WC-driven) movement of the dative DP from the ECM infinitive to the matrix Specv position. This prediction is borne out. Consider the so-called *lassen*-passive construction in (33), where the external argument of the embedded infinitive is demoted exactly as in standard passive constructions (including the option of realizing it as a PP, not indicated here) even though no morphological reflex of passive is present; see Höhle (1978), among many others.

(33) *DP$_{dat}$ movement where an embedded DP$_{ext}$ is not present:*

a. dass keiner [$_{DP}$ dieser Frau]$_1$ gestern/gerne [$_{XP}$ t$_1$
that no-one$_{nom}$ this woman$_{dat}$ yesterday/gladly
helfen ließ]
help let

b. dass er [$_{DP}$ ihm]$_1$ gestern/ungern [$_{XP}$ t$_1$ helfen ließ]
that he$_{nom}$ him$_{dat}$ yesterday/reluctantly help let

c. Wem$_1$ ließ Karl gestern/ungern [$_{XP}$ t$_1$ helfen] ?
whom$_{dat}$ let Karl$_{nom}$ yesterday/reluctantly help

d. Dem Lehrer$_1$ ließ Karl gestern/ungern [$_{XP}$ t$_1$
the teacher$_{dat}$ let Karl$_{nom}$ yesterday/reluctantly
helfen]
help

This effect is fully parallel to the one identified in Assmann, Georgi et al. (2013) with respect to legitimate ergative movement in the absence of an internal argument DP that requires structural (i.e., absolutive) case.

Second, there is the question as to why extraction of dative DPs becomes possible again if the predicate of the ECM complement is an unaccusative verb. The problem here is that the embedded DP$_{int}$ gets case from the matrix v and does not block dative movement; see (34) (from Fanselow 1990).

(34) *DP$_{dat}$ movement is fine in unaccusative contexts:*

a. *dass mir$_1$ niemand [$_{XP}$ Karl t$_1$ helfen] ließ/saw
that me$_{dat}$ no-one$_{nom}$ Karl$_{dat}$ help let/saw

b. dass mir$_1$ niemand [$_{XP}$ t$_1$ ein Unglück zustoßen] ließ/sah
that me$_{dat}$ no-one$_{nom}$ an accident happen to let/saw

This problem is solved if there is no vP with unaccusative predicates, pace Legate (2003). Under this assumption, WC does not force early DP$_{dat}$ movement because there is no improper f-seq when DP$_{dat}$ enters the matrix VP domain; hence, there is no maraudage.

Third, one may wonder what happens if DPs with other cases undergo extraction from ECM complements. It turns out that DP$_{acc}$ can undergo such movement easily (see (35), (36)). In contrast, DP$_{gen}$ movement is arguably much more restricted, see (37).

(35) *Scrambling of a DP$_{acc}$ object from ECM complements:*

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

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

(36) *Pronoun movement of a DP$_{acc}$ object from ECM complements:*

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

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

(37) *Movement of a DP$_{gen}$ object from ECM complements.*

a. Karl sieht/lässt den Jungen der Toten gedenken
Karl$_{nom}$ sees/lets the boy$_{acc}$ the dead$_{gen}$ commemorate

b. ?*dass derer/der Toten keiner den Jungen gedenken
that they$_{gen}$/the dead$_{gen}$ no-one$_{nom}$ the boy$_{acc}$ commemorate
sieht/lässt
sees/lets

c. ?*Der Toten$_1$ sieht/lässt Karl den Jungen gedenken
the dead$_{gen}$ sees/lets Karl$_{nom}$ the boy$_{acc}$ commemorate

This would follow without further ado if maraudage is blocked if exactly the same case is involved; and genitive and accusative are sufficiently different.

4. Decelerating Move: Mobility of Lexical/Oblique Arguments in Ergative Systems

The second argument deals with a reversal of the general preference for Move over Agree. Starting point is the observation that not all morphologically ergative languages exhibit the ban against extraction of the ergative subject.

4.1. Data

In some morphologically ergative languages, the ergative extracts freely and without any special morphology (such as the agent focus morphology encountered in many Mayan languages). Below, this is illustrated for Chol (Mayan), Basque (isolate), Avar (Nakh-Dagestanian), and Pitjantjatjara (Pama-Nyungan).

(38) *Wh-movement of DP$_{erg}$ in Chol* (Coon 2010: 226, Coon et al. 2011):
a. Maxki$_{1/2}$ tyi y-il-ä (t$_1$) aj-Maria (t$_2$)?
who ASP A3-see-DTV DET-Maria
'Who saw Maria?' / 'Who did Maria see?'
b. Maxki$_1$ tyi y-il-ä t$_1$ a-wakax?
who PRFV A 3-see-DTV A 2-COW
'Who saw your cow?'

(39) *Wh-movement of DP$_{erg}$ in Basque* (Hualde and Ortiz de Urbina 2003):
Nork$_1$ lagunduko die t$_1$ sure lagunei?
who.ERG help.FUT AUX your friends.DAT
'Who will help your friends?'

(40) *Relativization of DP$_{erg}$ in Avar* (Polinsky et al. 2011):
[Ø.ERG$_1$ t$_1$ ʕoloqana-y yas repetici-yal-de y-ač:-un
unmarried-II girl.ABS rehearsal-OBL-LOC II-bring-GER
y-ač’-ara-y] artistka$_1$ bercina-y y-igo.
II-come-PRTC-II actress.ABS beautiful-II II-AUX
'The actress that brought the young girl to the rehearsal is pretty.'

(41) *Relativization of DP$_{erg}$ in Pitjantjatjara* (Bowe 1990: 101):
Wati panya [Ø.ERG$_1$ t$_1$ wa<u>r</u>u atu-ntja-lu] ngayu-nya u-ngu.
man ANAPH wood chop-INF-ERG 1.SG-ACC give-PAST
'The man who chops wood gave me some.'

The idea of the analysis will be that extraction of the ergative subject is an option in these languages because Agree exceptionally applies before Merge on the TP-cycle despite the ergative ranking MC ≫ AC.

4.2. Assumptions

We propose that the possibility of moving the ergative argument in these languages is due the nature of the ergative case involved. More precisely, we would like to suggest that ergative case in Chol, Avar, Basque, and Pitjantjatjara is not structural but lexical (see Nash 1996, Alexiadou 2001, Woolford 1997, 2001, Legate 2008 for related claims). However, our overall argument here presupposes the theory of argument encoding put forward in Müller (2009), which is based on the idea that ergative type encoding systems involve a marked case [*CASE:*int**] on v, which is spelled out as ergative case.

In order to reconcile these two views, we make the following assumptions. The case probe [*CASE:*int**] representing internal structural case is in fact composed of the two subfeatures [–OBL] and [+GOV]. (Similarly, [*CASE:*ext**] is actually composed of [–OBL] and [–GOV].) Here, [±GOV] maintains the external/internal distinction, and [–OBL] indicates that the cases associated with T and v are structural (non-oblique). Such a decomposition of case features is first and foremost motivated by morphological considerations relating to syncretism: This way, natural classes of cases can be defined by referring to underspecified case information on morphological case exponents (e.g., [–OBL] captures the natural class of structural cases – nominative and accusative in accusative systems, and absolutive and ergative in ergative systems; [+GOV] captures the natural class of accusative/ergative, dative, and other governed cases; and so on); cf. Bierwisch (1967), Wiese (1999), and much recent work in Distributed Morphology.

In what follows, we will make use of case decomposition in the syntax. Accordingly, DP arguments bear unvalued variants of these subfeatures: [OBL:□] and [GOV:□]. A DP is valued with ergative (or, for that matter, accusative) case if its case subfeatures are valued [CASE:–OBL,+GOV]. In morphologically ergative languages with a *structural* ergative, these two case subfeatures are located on v. From there, they compositionally value case on DP_{ext}. Thus, here everything still works exactly as laid out above – the fine structure of the case feature may be relevant in morphology, but is in fact invisible in syntax. For languages with a *lexical* ergative, we assume that v only bears [–OBL] while [+GOV] is located on V; i.e., internal case is split between a [*CASE:−OBL*] probe on v and a [*CASE:+GOV*] probe on V. This reflects the hypothesis that lexical ergative is assigned by V in interaction with v. In addition, we assume that the two subfeatures [−OBL] and [+GOV] involved in lexical case assignment differ with respect to the structural conditions they require for entering into Agree: For [−OBL], m-command is sufficient (see (8)); but the [+GOV] feature on V that makes the composite case lexical is discharged under a stricter locality condition: It must *c-command* the goal that it is supposed to establish Agree with. This corresponds to the observation that "pure" lexical case assignment (e.g., a lexically assigned genitive in German) typically ends up on the lowest argument DP of a predicate (see Fanselow 2001).[10]

Next, we propose that arguments with partially valued case are inactive in the sense that they are invisible for structure-building features triggering Merge (cf. Richards 2008; also cf. Chomsky's 2001 *Activity Condition*). As a consequence, inactive elements cannot undergo movement. This is explicitly stated by the constraint in (42).

(42) ACTIVITY CONDITION (ACTC):
Inactive elements cannot undergo movement.

Thus, DPs that have at least one but not yet all of their case features valued must first finish valuation before they can undergo movement. In contrast, DPs that have their case features completey valued – or completely unvalued – are active.[11] The guiding hypothesis here is that there is a general *contiguity* requirement for syntactic operations: An operation consisting of several subparts must be fully completed once it has begun before the affected item can be accessed by other operations (i.e., qualify as active).

Finally, given that lexical (i.e., V-based) [+GOV] on V can only be assigned under c-command, and given that v-V does not c-command Specv, something

---

[10] No such strict c-command requirement holds for [+GOV] on v, where it is not a lexical case feature. This presupposes that [+GOV] on v (structural case) and [+GOV] on V (lexical case) can be distinguished accordingly. Given the minimal contextual difference (part of v vs. part of V), this would seem to be unproblematic.

[11] There is evidence suggesting that a DP that has not received any case value so far must not qualify as inactive under these assumptions, and can accordingly undergo movement. First, this is required by classical approaches to case-driven raising; second, it is in fact required for the derivation of constructions involving absolutive movement in ergative systems under present assumptions; recall the derivation in (18).

needs to be said about how [+GOV] can eventually be assigned to an external argument DP in Specv. The obvious conclusion would seem to be that this is effected by v-V-to-T movement; T c-commands Specv. We will assume that this is indeed correct. If so, there are basically two options how the required head movement operation can be brought about. First, it might be that v-V-to-T movement must independently exist in a language to make lexical ergative assignment possible; this would imply that a lexical ergative results from the conspiracy of two independent parameter settings (viz., (a) [+GOV] on V (not v) and (b) obligatory v-V-to-T movement). Second, it might be that the need for V to get rid of its lexical case subfeature [+GOV] may directly trigger v-V-to-T movement, as an instance of repair-driven movement (see Heck and Müller 2000), and irrespective of any general parameter setting for the head movement operation. Since nothing hinges on this, we will not choose between the two options in what follows; we tentatively adopt the first option for reasons of exposition alone.[12]

4.3. Analysis

Imagine a scenario where DP$_{ext}$ is supposed to undergo Ā-movement in an ergative system where the ergative is lexical. Given MC ≫ AC, DP$_{ext}$ is merged before v can trigger Agree; the same ranking may also be assumed to trigger V-to-v movement early. After being merged in Specv, DP$_{ext}$'s case feature is partially valued by [*CASE:–OBL*] on v (due to the SHB), yielding [CASE:–OBL,□]. However, DP$_{ext}$ is not in the c-command domain of v-V, so the remaining (lexical case) probe [*CASE:+GOV*] cannot participate in Agree at this point. As a consequence, DP$_{ext}$ is inactive when T is merged; see (43-a). Due to the ranking ACTC ≫ MC ≫ AC, the inactive DP$_{ext}$ now cannot immediately move to SpecT once T has been introduced into the structure, despite the presence of the "ergative" ranking MC ≫ AC. However, this ranking successfully triggers head movement of v-V to T; see (43-b). At this point,

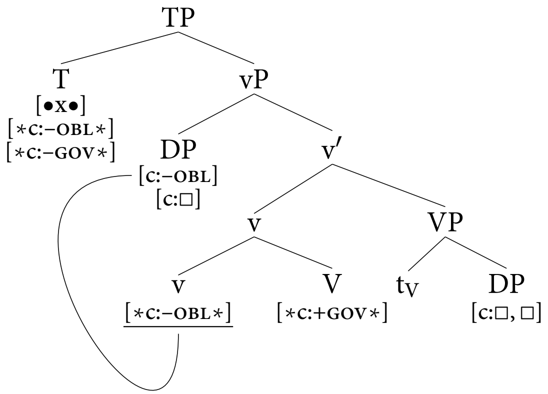
---

[12] Another issue that must be clarified in this context but is orthogonal to our main concerns is how head movement of V to v, and subsequently of v to T, can result in proper c-command by V (of Specv, as required for lexical case valuation, but also of its own trace). A standard solution to this problem is to minimally relax the locality condition on c-command, such that if a head α is adjoined to another head β, α c-commands whatever β c-commands (see Baker 1988). Alternatively, following Roberts (2010), we may assume that complete copying of a feature set derives the effects of head movement (without actual movement taking place). The copying operation would then also comprise the case feature [+GOV] on V.

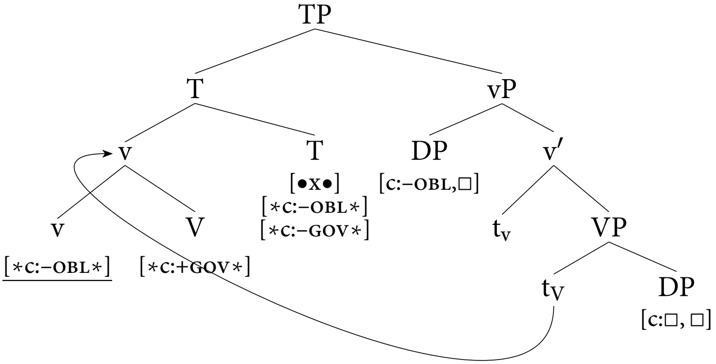
there are two case probes in v-V-T: On the one hand, there is [∗CASE:–OBL,–GOV∗] (i.e., [∗CASE:*ext*∗]) on T, and on the other hand, there is the partial probe [∗CASE:+GOV∗] on V. There are four possibilities as to what can happen next. T's case probe may undergo Agree with DP$_{int}$, T's case probe may undergo Agree with DP$_{ext}$, V's partial case probe may undergo Agree with DP$_{int}$, or V's partial case probe may undergo Agree with DP$_{ext}$. The present system does not distinguish between these four options; they all satisfy ActC, violate MC once, and violate AC once. (In contrast, movement of DP$_{ext}$ would fatally violate ActC.) Suppose that the first option is chosen: [∗CASE:–OBL,–GOV∗] on T undergoes Agree with DP$_{int}$, valuing the latter's case feature, as in (43-c). Then, in the next step, [∗CASE:+GOV∗] on V values DP$_{ext}$, which must still be in situ, given ActC; see (43-d). However, after this final Agree operation, DP$_{ext}$ is active again, and it can and must finally undergo the intermediate movement step to SpecT; cf. (43-e). Of course, this means that movement of DP$_{ext}$ comes too late to maraude T's absolutive case feature for DP$_{int}$. Hence, extraction of the lexically ergative-marked DP$_{ext}$ does not lead to ungrammaticality.

(43) *Legitimate movement of DP$_{erg}$ if the ergative is lexical*

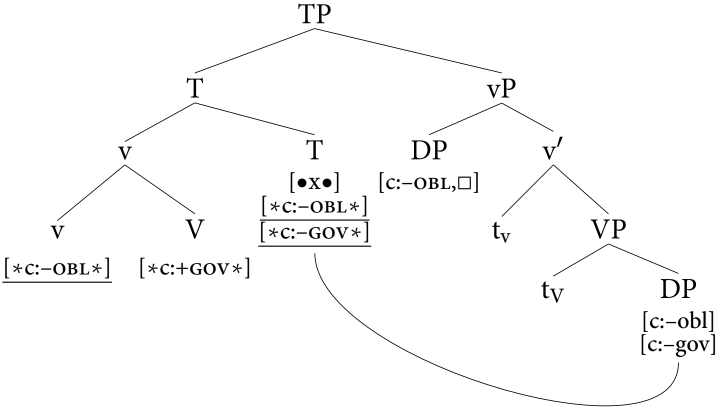
a. Structure after T is merged; partial case assignment to DP$_{ext}$



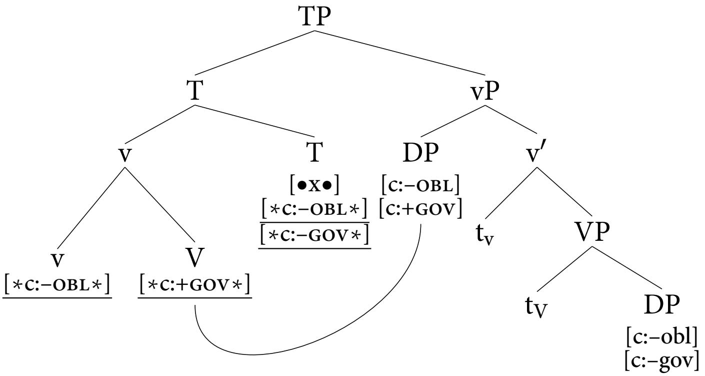
b. ActC ≫ MC ≫ AC blocks movement of DP$_{ext}$; permits movement of v-V



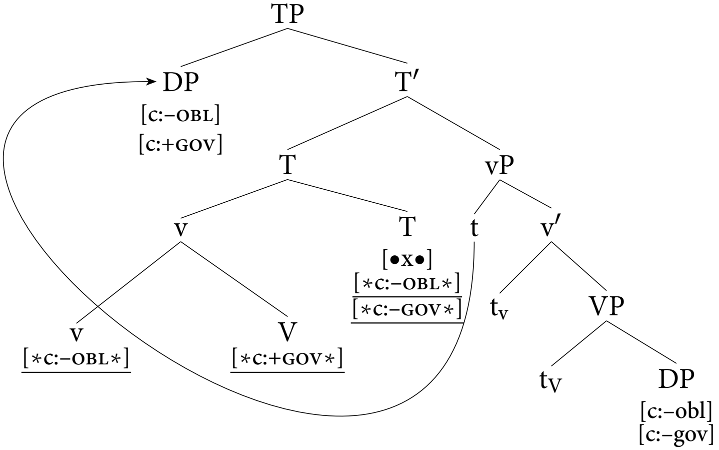
c. ActC ≫ MC ≫ AC blocks movement of DP$_{ext}$; permits Agree(T, DP$_{int}$)



d. Agree(V, DP$_{ext}$) renders DP$_{ext}$ active



e. DP_{ext} finally moves but comes too late to effect maraudage



The relevant competitions are given in tableaux T_{12}, T_{13}, and T_{14}. Note that there are four locally optimal continuations O_2–O_5 (that all carry out an Agree operation) in T_{12} (which illustrates the crucial step from (43-b) to (43-c)), in addition to O_1, which executes movement of DP_{ext} and thereby fatally violates ActC. However, of these four optimal outputs only O_2 (where T undergoes Agree with DP_{int}) will eventually lead to a well-formed output: In O_3 and O_5, DP_{ext} gets its case valued (by T and V, respectively), which means that it becomes active and will have to move in the next step, thereby marauding case features required for DP_{int}. Similarly, O_4 will invariably lead to a crash because DP_{int} undergoes Agree with V here, and will therefore never acquire a fully specified case feature. (Alternatively, DP_{ext} will fail to do so if DP_{int} marauds T's features as well; note that we assume that a case probe cannot distribute its decomposed feature values over different goals.)

Assuming a continuation with O_2, tableau T_{13} shows that the situation is still such that DP_{ext} cannot move without fatally violating ActC.

Finally, tableau T_{14} illustrates the trivial final competition on the TP cycle: DP_{ext} is now active, and movement can finally be carried out.

4.4. Extremely Local vs. Less Local Optimization

As with the very option of accusative movement under present assumptions, an argument for extremely local optimization emerges in the case of deceler-

T$_{12}$: *Lexical ergative movement, step 1: Agree(T,DP$_{int}$)*

Input: $[_{\text{TP}}$ v-V$_{[*\text{C:+GOV}*]}$-T$_{[*\text{C:–OBL,–GOV}*],[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL},\square]}$ … DP$_{[\text{C:}\square,\,\square]}$]]	ActC	MC	AC
O$_1$: $[_{\text{TP}}$ DP$_{[\text{C:–OBL},\square]}$ v-V$_{[*\text{C:+GOV}*]}$-T$_{[*\text{C:–OBL,–GOV}*]}$ $[_{\text{vP}}$ t … DP$_{[\text{C:}\square]}$]]	*!		**
☞ O$_2$: $[_{\text{TP}}$ v-V$_{[*\text{C:+GOV}*]}$-T$_{[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL},\square]}$ … DP$_{[\text{C:–OBL,–GOV}]}$]]		*	*
☞ O$_3$: $[_{\text{TP}}$ v-V$_{[*\text{C:+GOV}*]}$-T$_{[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL,–OBL,–GOV}]}$ … DP$_{[\text{C:}\square,\,\square]}$]]		*	*
☞ O$_4$: $[_{\text{TP}}$ v-V-T$_{[*\text{C:–OBL,–GOV}*],[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL},\square]}$ … DP$_{[\text{C:}\square\text{,+GOV}]}$]]		*	*
☞ O$_5$: $[_{\text{TP}}$ v-V-T$_{[*\text{C:–OBL,–GOV}*],[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL,+GOV}]}$ … DP$_{[\text{C:}\square,\,\square]}$]]		*	*

T$_{13}$: *Lexical ergative movement, step 2: Agree(V,DP$_{ext}$)*

Input: $[_{\text{TP}}$ v-V$_{[*\text{C:+GOV}*]}$-T$_{[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL},\square]}$ … DP$_{[\text{C:–OBL,–GOV}]}$]]	ActC	MC	AC
O$_1$: $[_{\text{TP}}$ DP$_{[\text{C:–OBL},\square]}$ v-V$_{[*\text{C:+GOV}*]}$-T $[_{\text{vP}}$ t … DP$_{[\text{C:–OBL,–GOV}]}$]]	*!		*
☞O$_2$: $[_{\text{TP}}$ v-V-T$_{[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL,+GOV}]}$ … DP$_{[\text{C:–OBL,–GOV}]}$]]		*	

T$_{14}$: *Lexical ergative movement, step 3: Move(T,DP$_{ext}$)*

Input: $[_{\text{TP}}$ v-V-T$_{[\bullet\text{x}\bullet]}$ $[_{\text{vP}}$ DP$_{[\text{C:–OBL,+GOV}]}$ … DP$_{[\text{C:–OBL,–GOV}]}$]]	ActC	MC	AC
☞O$_1$: $[_{\text{TP}}$ DP$_{[\text{C:–OBL,+GOV}]}$ v-V-T $[_{\text{vP}}$ t … DP$_{[\text{C:–OBL,–GOV}]}$]]			

ating ergative movement if the ergative is lexical, and it does so for essentially the same reason: If the whole TP (or an even larger domain) is considered, the ban on ergative movement that follows from the ranking MC ≫ AC cannot be circumvented anymore. ActC is of course fulfilled at the TP level; and with

DP_{ext} in SpecT, there should be maraudage of T's case features by DP_{ext}, given the SHB. As before, there is thus a counter-bleeding effect that is unproblematic if one derivational step is considered after the other, but that creates problems for more representational approaches where the relevant distinctions are lost. This is shown in tableau T_{15}.

T_{15}: *TP optimization under* ACTC ≫ *MC* ≫ *AC ranking: wrong result*

Input: $\text{T}_{[*\text{C:}-\text{OBL},-\text{GOV}*],[\bullet\text{X}\bullet]}$ ⊕ $[_{\text{vP}}\ \text{DP}_{[\text{C:}-\text{OBL},\square]}\ \text{v-V}_{[*\text{C:}+\text{GOV}*]} \ldots \text{DP}_{[\text{C:}\square,\square]}\ ]]$	ACTC	SHB	MC	AC
O_1: $[_{\text{TP}}\ \text{DP}_{[\text{C:}-\text{OBL},+\text{GOV}]}\ \text{v-V-T}$ $[_{\text{vP}}\ \text{t} \ldots \text{DP}_{[\text{C:}-\text{OBL},-\text{GOV}]}\ ]]$		*!		
☛ O_2: $[_{\text{TP}}\ \text{DP}_{[\text{C:}-\text{OBL},+\text{GOV},-\text{OBL},-\text{GOV}]}\ \text{v-V-T}$ $[_{\text{vP}}\ \text{t} \ldots \text{DP}_{[\text{C:}\square,\square]}\ ]]$				

5. Conclusion and Outlook

In this paper, we have tried to defend four related claims.

First, it seems to be a fact that given standard minimalist assumptions about structure-building, competition between Merge (Move) and Agree can arise in the derivation. In particular, such a situation occurs when the head of a phrase has to carry out more than one operation. On the vP cycle, this is the case with a v that introduces an external argument DP and assigns structural case (v[∗CASE:_int_∗],[•D•]); on the TP cycle this is the case with a T that triggers an intermediate movement step via an edge feature and also assigns structural case (T[∗CASE:_ext_∗],[•X•]).

Second, the conflicts that arise between Merge (Move) and Agree can and must be resolved in one way or the other. We have argued that there may be no intrinsic, fixed way of resolution; rather, the empirical evidence suggests that how conflicts are resolved is a matter of parametrization. This can be implemented either by invoking parametrized preference principles (of the type of the Merge before Move principle in Chomsky 1995, 2001); or by postulating constraint ranking. Assuming (for concreteness) the latter, we have seen that MC ≫ AC on the vP cycle gives rise to an ergative encoding system whereas the reverse resolution strategy following from AC ≫ MC on the vP cycle predicts an accusative encoding system; and that a ranking MC ≫ AC on the

TP cycle accounts for the immobility of DPs bearing structural ergative case (because these items move too early, bringing about maraudage of T's case feature) whereas a reverse ranking AC ≫ MC on the TP cycle correctly predicts movement of DPs bearing structural accusative case to be possible (because these items move late, thereby avoiding maraudage of T's case features).

Third, and most importantly, we have argued that the ban on dative movement from ECM contexts in German, and the option of lexical ergative movement in Chol, Avar, Basque, and Pitjantjatjara, can be accounted for straightforwardly if it is assumed that there can be an acceleration of movement in accusative systems that normally give preference to Agree over Merge (Move) in the case of conflict if this is forced by an independent factor; and that there can also be a deceleration of movement in ergative systems that normally give preference to Merge (Move) over Agree in the case of conflict if this is forced by an independent factor. These situations are fully expected under an optimality-theoretic approach, but less so under a more orthodox minimalist approach employing (parametrized) preference principles. Thus, a ranking WC ≫ AC ≫ MC on the vP cycle correctly predicts dative movement from German ECM constructions to be impossible (because it comes too early), and a ranking ActC ≫ MC ≫ AC on the TP cycle correctly predicts lexically marked ergative subjects to be mobile in Chol, Avar, Basque, and Pitjantjatjara. In contrast, a modified preference principle like "Agree before Move unless satisfaction of the Williams Cycle demands otherwise" does not per se look like a plausible candidate for a constraint of grammar; and the same goes for a modified preference principle like "Move before Agree unless satisfaction of the Activity Condition demands otherwise".[13]

Fourth and finally, the analyses presented in this paper provide evidence for extremely local serial optimization in syntax, and against less local optimization procedures (including ones where the whole sentence is subject to a single, parallel optimization): If the domain is larger than the derivational step, then (i) AC ≫ MC on the vP cycle does not derive accusative encoding systems; (ii) AC ≫ MC on the TP cycle wrongly blocks accusative movement; and (iii) ActC ≫ MC ≫ AC on the TP cycle cannot circumvent the ban on

---

[13] That said, it might eventually not be impossible to save the preference principle-based approach, by postulating that only *convergent* steps are considered, and further assuming that violations of WC and ActC lead to non-convergence. It is far from obvious, however, that a simple notion of convergence can be devised that covers all relevant contexts in a natural way; see Sternefeld (1996) for related discussion.

ergative movement. The reasonings here rely on standard arguments based on opacity of rule interaction in generative grammar.

From a more general point of view, the present study can be seen as an attempt to sketch the outlines of a new approach to an empirical domain that received a lot of attention in earlier work in the Principles and Parameters tradition but has arguably been given much less attention in more recent minimalist approaches, viz., asymmetries between types of categories with respect to their extractability. It has often been observed that some kinds of linguistic expressions are less mobile than others in the sense that they may not cross domains that are transparent for other items. Such asymmetries have been noted for objects vs. subjects, for arguments vs. adjuncts, for referential vs. non-referential phrases, for items that have an "address" vs. others that don't (see Manzini 1992), and so on. Standardly, these kinds of asymmetries were captured by imposing appropriate constraints on empty categories that are assumed to be left behind by displacement operations (cf., e.g., Chomsky's 1981 Empty Category Principle (ECP) for traces, or the different constraints for trace vs. *pro* in Cinque 1990). However, such options do not exist anymore under minimalist assumptions according to which all constraints are either principles of efficient computation or imposed by the interfaces (see Chomsky 2001, 2008). Furthermore, traces – as special items enriching the syntactic ontology for which designated constraints can be formulated – have come to be widely regarded as suspect from a minimalist viewpoint.[14]

Taken together, this means that there is a gap in current minimalist approaches to syntax: It is a priori unclear how asymmetries between moved items can be accounted for. The present approach can be viewed as a program for filling this gap. The basic premise is that if some items are less mobile than others, this must be so because their movement may lead to problems elsewhere (i.e., in domains not directly related to the movement operation), either for themselves or for other items in the clause. We have argued that movement of certain items (α) may create problems for other, sufficiently similar items (β). Thus, by pursuing this program, we end up with a *relational*, co-argument-based approach to displacement (α cannot move in the presence of β because α-movement creates problems for β-licensing) of the type that has sometimes

---

[14] This may be so because displacement does not leave a reflex in the original position to begin with; see Epstein and Seely (2002), Unger (2010), Müller (2011) for some options; or because a multidominance approach is adopted; see Gärtner (2002), Starke (2001), Abels (2004), Frampton (2004), among others.

been suggested for case assignment (α is assigned x-case in the presence of β; see Marantz 1991, Bittner and Hale 1996, Wunderlich 1997, Stiebels 2000, McFadden 2004). More specifically, a common pattern emerges that captures the legitimate and illegitimate instances of movement discussed in the present paper: On the one hand, movement of some category α that takes place early on a given cycle brings with it the danger of maraudage of features that would be needed for the licensing of some other category β, and may thereby lead to ungrammaticality; this holds for DPs that bear structural ergative case and for dative DPs that have a chance to immediately remedy a temporary improper movement configuration.[15] On the other hand, movement of some category α that takes place late on a given cycle will more likely be able to circumvent maraudage effects for some other category β, and will therefore more often lead to grammaticality; this holds for DPs that bear structural accusative case and for ergative DPs where the ergative is lexical and the DP in question is therefore not yet active (hence, not yet accessible by movement).[16] Overall, then, a simple generalization emerges:

Good things come to those who wait.

References

Abels, Klaus (2003): Successive Cyclicity, Anti-Locality, and Adposition Stranding. PhD thesis, University of Connecticut, Storrs, Connecticut.

Abels, Klaus (2004): Right Node Raising: Ellipsis or Across the Board Movement?. *In:* K. Moulton and M. Wolf, eds, *Proceedings of NELS 34*. GLSA, Amherst, Mass., pp. 45–60.

---

[15] This latter case shows inherent myopia (see Collins 1997), i.e., the strict lack of look-ahead of the current approach: By trying to correct a problem that can in principle be tolerated (viz., a temporary improper movement configuration), much greater damage is done, and the derivation eventually crashes.

[16] Needless to say, the new approach to asymmetries between moved items raises many new questions that will have to be addressed before it can be regarded as viable. From an empirical point of view, obvious challenges that need to be faced include *that*-trace effects, Left Branch Condition effects (i.e., the immobility of pre-nominal genitive DPs), and the stronger restrictions on adjunct movement. From a conceptual point of view, the availability of re-ranking among the relevant constrains remains an open issue; for instance, the fact that ActC and WC outrank the more basic constraints AC and MC in all cases addressed above does not seem to be accidental – but if it isn't, the task is to make this follow in some principled way that is compatible with basic optimality-theoretic tenets.

Adger, David and Daniel Harbour (2007): 'Syntax and Syncretisms of the Person Case Constraint', *Syntax* 10, 2–37.

Alderete, John (2001): *Morphologically Governed Accent in Optimality Theory*. Routledge, New York.

Aldridge, Edith (2004): Ergativity and Word Order in Austronesian Languages. PhD thesis, Cornell University.

Alexiadou, Artemis (2001): *Functional Structure in Nominals*. Benjamins, Amsterdam.

Anagnostopoulou, Elena (2005): Strong and Weak Person Restrictions. A Feature Checking Analysis. *In:* L. Heggie and F. Ordoñez, eds, *Clitics and Affix Combinations*. Benjamins, Amsterdam, pp. 199–235.

Andrews, Avery (1996): 'Semantic Case-Stacking and Inside-Out Unification', *Australian Journal of Linguistics* 16, 1–54.

Assmann, Anke, Svetlana Edygarova, Doreen Georgi, Timo Klein and Philipp Weisser (2013): Possessor Case in Udmurt: Multiple Case Assignment Feeds Postsyntactic Fusion. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 23-63.

Assmann, Anke, Doreen Georgi, Fabian Heck, Gereon Müller and Philipp Weisser (2013): Ergatives Move too Early: On an Instance of Opacity in Syntax. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 363-412.

Bader, Monika (2011): Constraining Operations. A Phase-Based Account of Improper Movement and Anaphoric Binding. PhD thesis, CASTL, Tromsø University., Tromsø.

Baker, Mark (1988): *Incorporation. A Theory of Grammatical Function Changing*. University of Chicago Press, Chicago.

Baković, Eric and Colin Wilson (2000): Transparency, Strict Locality, and Targeted Constraints. *In:* R. Billerey and B. D. Lillehaugen, eds, *WCCFL 19 Proceedings*. Cascadilla Press, Somerville, MA, pp. 43–56.

Béjar, Susana and Milan Řezáč (2009): 'Cyclic Agree', *Linguistic Inquiry* 40, 35–73.

Bierwisch, Manfred (1967): Syntactic Features in Morphology: General Problems of So-Called Pronominal Inflection in German. In: *To Honor Roman Jakobson*. Mouton, The Hague/Paris, pp. 239–270.

Bittner, Maria and Ken Hale (1996): 'The Structural Determination of Case and Agreement', *Linguistic Inquiry* pp. 1–68.

Boeckx, Cedric (2003): *Islands and Chains. Resumption as Stranding*. Benjamins, Amsterdam.

Bošković, Željko (2007): 'Agree, Phases, and Intervention Effects', *Linguistic Analysis* 33, 54–96.

Bowe, Heather (1990): *Categories, Constituents and Constituent Order in Pitjantjatjara: An Aboriginal Language of Australia*. Routledge, New York.

Campana, Mark (1992): A Movement Theory of Ergativity. PhD thesis, McGill University.

Chomsky, Noam (1951): Morphophonemics of Modern Hebrew. Master's thesis, University of Pennsylvania.

Chomsky, Noam (1975): *The Logical Structure of Linguistic Theory*. Plenum Press, New York.

Chomsky, Noam (1981): *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam (1986): *Barriers*. MIT Press, Cambridge, Mass.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. *In:* R. Martin, D. Michaels and J. Uriagereka, eds, *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. *In:* M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Chomsky, Noam (2005): 'Three Factors in Language Design', *Linguistic Inquiry* **36**, 1–22.

Chomsky, Noam (2007): Approaching UG from Below. *In:* U. Sauerland and H.-M. Gärtner, eds, *Interfaces + Recursion = Language?*. Mouton de Gruyter, Berlin, pp. 1–31.

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

Chomsky, Noam (2013): 'Problems of Projection', *Lingua* . To appear.

Cinque, Guglielmo (1990): *Types of A-bar Dependencies*. MIT Press, Cambridge, Mass.

Collins, Chris (1997): *Local Economy*. MIT Press, Cambridge, Mass.

Coon, Jessica (2010): Complementation in Chol (Mayan). A Theory of Split Ergativity. PhD thesis, MIT, Cambridge, Mass.

Coon, Jessica, Pedro Mateo Pedro and Omer Preminger (2011): The Role of Case in A-Bar Extraction Asymmetries: Evidence from Mayan. Ms., McGill University.

Deal, Amy Rose (2012): Ergativity. Ms., University of California, Santa Cruz. To appear in *Syntax. An International Handbook of Contemporary Research*, ed. Tibor Kiss & Artemis Alexiadou. Berlin: De Gruyter.

Epstein, Samuel David and T. Daniel Seely (2002): Rule Applications as Cycles in a Level-Free Syntax. *In:* S. D. Epstein and T. D. Seely, eds, *Derivation and Explanation in the Minimalist Program*. Blackwell, Oxford, pp. 65–89.

Fanselow, Gisbert (1986): Scrambling, Anaphern und Barrieren. GGS-Vortrag, Universität Regensburg.

Fanselow, Gisbert (1990): Scrambling as NP-Movement. *In:* G. Grewendorf and W. Sternefeld, eds, *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 113–140.

Fanselow, Gisbert (1991): Minimale Syntax. Habilitation thesis, Universität Passau.

Fanselow, Gisbert (2001): Optimal Exceptions. *In:* B. Stiebels and D. Wunderlich, eds, *The Lexicon in Focus*. Akademie Verlag, Berlin, pp. 173–209.

Ferguson, K. Scott (1993): 'Notes on the Shortest Move Metric and Object Checking', *Harvard Working Papers in Linguistics* 3, 65–80.

Frampton, John (2004): Copies, Traces, Occurrences, and All That. Ms., Northeastern University.

Gärtner, Hans-Martin (2002): *Generalized Transformations and Beyond*. Akademie-Verlag, Berlin.

Georgi, Doreen (2013): Opaque Interaction of Internal Merge and Agree. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 413–461.

Georgi, Doreen, Fabian Heck and Gereon Müller (2009): Maraudage. Ms., Universität Leipzig. (Talk at the Repairs Workshop of the DGfS conference 2009, Osnabrück, and at the Potsdam/Leipzig Workshop on Movement and Morphology, Leucorea, April 2009).

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

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

Grimshaw, Jane (1994): Heads and Optimality. Ms., Rutgers University. (Handout of talk, Universität Stuttgart).

Heck, Fabian and Gereon Müller (2000): Repair-Driven Movement and the Local Optimization of Derivations. Ms., Universität Stuttgart and IDS Mannheim.

Heck, Fabian and Gereon Müller (2007): Extremely Local Optimization. *In:* E. Brainbridge and B. Agbayani, eds, *Proceedings of the 26th WECOL*. California State University, Fresno, pp. 170–183. [P5].

Heck, Fabian and Gereon Müller (2013): Extremely Local Optimization. *In:* H. Broekhuis and R. Vogel, eds, *Linguistic Derivations and Filtering*. Equinox, Sheffield, pp. 136–165.

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

Hualde, José Ignacio and Jon Ortiz de Urbina, eds (2003): *A Grammar of Basque*. Mouton de Gruyter, Berlin.

Jelinek, Eloise (1993): Ergative 'Splits' and Argument Type. *In:* J. Bobaljik and C. Phillips, eds, *Papers on Case and Agreement I*. Vol. 18 of *MIT Working Papers in Linguistics*, MITWPL, MIT: Cambridge, Mass., pp. 15–42.

Kayne, Richard (1989): Facets of Participle Agreement in Romance. *In:* P. Benincà, ed., *Dialect Variation and the Theory of Grammar*. Foris, Dordrecht, pp. 85–104.

Kiparsky, Paul (1976): Abstractness, Opacity, and Global Rules. *In:* A. Koutsoudas, ed., *The Ordering and Application of Rules*. Mouton.

Koopman, Hilda (2006): Agreement Konfigurations. *In:* C. Boeckx, ed., *Agreement Systems*. Benjamins, Amsterdam, pp. 159–201.

Koppen, Marjo van (2005): One Probe – Two Goals. Aspects of Agreement in Dutch Dialects. PhD thesis, Leiden University.

Koutsoudas, Andreas (1966): *Writing Transformational Grammars: An Introduction*. McGrah-Hill Book Company, New York.

Koutsoudas, Andreas (1973): Unordered Rule Hypotheses. Ms., Indiana University Linguistics Club.

Kratzer, Angelika (1996): Severing the External Argument from its Verb. *In:* J. Rooryck and L. Zaring, eds, *Phrase Structure and the Lexicon*. Kluwer, Dordrecht, pp. 109–137.

Lahne, Antje (2012): Specificity-Driven Syntactic Derivation. *In:* M. Uribe-Etxebarria and V. Valmala, eds, *Ways of Structure Building*. Oxford University Press, Oxford, pp. 271–296.

Lasnik, Howard (1999): 'On Feature Strength: Three Minimalist Approaches to Overt Movement', *Linguistic Inquiry* 30, 197–217.

Legate, Julie Anne (2003): 'Some Interface Properties of the Phase', *Linguistic Inquiry* 34, 506–516.

Legate, Julie Anne (2008): 'Morphological and Abstract Case', *Linguistic Inquiry* 39(1), 55–101.

Manzini, Rita (1992): *Locality. A Theory and Some of Its Empirical Consequences*. MIT Press, Cambridge, Mass.

Marantz, Alec (1991): Case and Licensing. *In:* German Westphal, B. Ao and H.-R. Chae, eds, *Proceedings of the Eigth Eastern States Conference on Linguistics*. University of Maryland, pp. 234–253.

McCarthy, John (2010): 'An Introduction to Harmonic Serialism', *Language and Linguistics Compass* 4, 1001–1018.

McFadden, Thomas (2004): The Position of Morphological Case in the Derivation: A Study on the Syntax-Morphology Interface. PhD thesis, University of Pennsylvania.

Merchant, Jason (2006): Polyvalent Case, Geometric Hierarchies, and Split Ergativity. *In:* J. Bunting, S. Desai, R. Peachey, C. Straughn and Z. Tomkova, eds, *Proceedings of the 42nd Annual Meeting of the Chicago Linguistics Society*. Chicago Linguistics Society, Chicago, pp. 47–67.

Müller, Gereon (2004): Phrase Impenetrability and Wh-Intervention. *In:* A. Stepanov, G. Fanselow and R. Vogel, eds, *Minimality Effects in Syntax*. Mouton de Gruyter, Berlin, pp. 289–325.

Müller, Gereon (2009): Ergativity, Accusativity, and the Order of Merge and Agree. *In:* K. K. Grohmann, ed., *Explorations of Phase Theory. Features and Arguments*. Mouton de Gruyter, Berlin, pp. 269–308.

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

Müller, Gereon (2013): A Local Reformulation of the Williams Cycle. *In:* F. Heck and A. Assmann, eds, *Rule Interaction in Grammar. Volume 90 of Linguistische Arbeitsberichte*. Institut für Linguistik: Universität Leipzig, pp. 247–299.

Murasugi, Kumiko (1992): Crossing and Nested Paths. PhD thesis, MIT, Cambridge, Mass.

Nash, Léa (1996): The Internal Ergative Subject Hypothesis. *In:* K. Kusumoto, ed., *Proceedings of NELS 26*. GLSA, Amherst, Mass, pp. 195–209.

Neeleman, Ad and Hans van de Koot (2010): 'A Local Encoding of Syntactic Dependencies and its Consequences for the Theory of Movement', *Syntax* 13, 331–372.

Nordlinger, Rachel (1998): *Constructive Case: Evidence from Australian Languages*. CSLI, Stanford.

Pesetsky, David (1989): Language-Particular Processes and the Earliness Principle. Ms., MIT, Cambridge, Mass.

Polinsky, Maria, Carlos Gallo, Peter Graff and Ekaterina Kravtchenko (2011): 'Subject Preference and Ergativity', *Lingua* pp. 267–277.

Pullum, Geoffrey (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.

Richards, Marc (2008): Quirky Expletives. *In:* G. H. H. Roberta d'Alessandro and S. Fischer, eds, *Agreement Restrictions*. Mouton de Gruyter, pp. 181–213. [P5].

Richards, Norvin (2007): Lardil "Case Stacking" and the Structural/Inherent Case Distinction. Ms., MIT.

Rizzi, Luigi (1990): *Relativized Minimality*. MIT Press, Cambridge, Mass.

Roberts, Ian (2010): *Agreement and Head Movement: Clitics, Incorporation and Defective Goals*. MIT Press, Cambridge, Mass.

Rouveret, Alain and Jean-Roger Vergnaud (1980): 'Specifying Reference to the Subject', *Linguistic Inquiry* 11, 97–202.

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.

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

Sternefeld, Wolfgang (1996): Comparing Reference Sets. *In:* C. Wilder, H.-M. Gärtner and M. Bierwisch, eds, *The Role of Economy Principles in Linguistic Theory*. Akademie Verlag, Berlin, pp. 81–114.

Sternefeld, Wolfgang (2006): *Syntax*. Stauffenburg, Tübingen. Two volumes.

Stiebels, Barbara (2000): Linker Inventories, Linking Splits and Lexical Economy. *In:* B. Stiebels and D. Wunderlich, eds, *Lexicon in Focus*. Akademie-Verlag, Berlin, pp. 211–245.

Stiebels, Barbara (2006): 'Agent Focus in Mayan Languages', *Natural Language and Linguistic Theory* 24, 501–570.

Svenonius, Peter (1994): 'C-Selection as Feature-Checking', *Studia Linguistica* 48, 133–155.

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

Unger, Christina (2010): A Computational Approach to the Syntax of Displacement and the Semantics of Scope. PhD thesis, Universiteit Utrecht, LOT.

Ura, Hiroyuki (2000): *Checking Theory and Grammatical Functions in Universal Grammar*. Oxford University Press, New York.

Wiese, Bernd (1999): 'Unterspezifizierte Paradigmen. Form und Funktion in der pronominalen Deklination', *Linguistik Online* 4. (www.linguistik-online.de/3_99).

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

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

Woolford, Ellen (1997): 'Four-Way Case Systems: Ergative, Nominative, Objective and Accusative', *Natural Language and Linguistic Theory* 15, 181–227.

Woolford, Ellen (2001): Case Patterns. *In:* G. Legendre, J. Grimshaw and S. Vikner, eds, *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 509–543.

Wunderlich, Dieter (1997): 'Cause and the Structure of Verbs', *Linguistic Inquiry* 27, 27–68.

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