Workshop TAG+5, Paris, 25-27 May 2000

Scrambling in German and the non-locality of local TDGs

Laura Kallmeyer

SFB 441, University of Tübingen
lk@sfs.nphil.uni-tuebingen.de

Abstract

Existing analyses of German scrambling phenomena within TAG-related formalisms all use non-local variants of TAG. However, there are good reasons to prefer local grammars, in particular with respect to the use of the derivation structure for semantics. Therefore this paper proposes to use local TDGs, a TAG-variant generating tree descriptions that shows a local derivation structure. However the construction of minimal trees for the derived tree descriptions is not subject to any locality constraint. This provides just the amount of non-locality needed for an adequate analysis of scrambling. To illustrate this a local TDG for some German scrambling data is presented.

1. Introduction

Scrambling in German poses a problem for most grammar formalisms. Neither Tree Adjoining Grammar (TAG, Joshi *et al.*, 1975) nor even linear context-free rewriting systems (LCFRS, Weir, 1988) are powerful enough to deal with scrambling and the free word order in German (see Becker *et al.*, 1992). (Becker *et al.*, 1991) propose a scrambling analysis with non-local multicomponent TAG (MCTAG, Weir, 1988), and (Rambow & Lee, 1994; Rambow, 1994) propose the use of vector TAG (V-TAG). These formalisms are both non-local in the sense that when adding a new element of the grammar in a derivation step, this element is not attached to one single previously added element of the grammar.

There are however good reasons to prefer a local grammar. Firstly, locality often restricts the parsing complexity, and local grammars often generate only semilinear languages. (Though some non-local formalisms (lexicalized V-TAG for instance) also can be shown to be polynomially parsable.) Secondly, in a local grammar, the derivation structure might reflect a dependency structure based on which semantic representations can be built (as for TAGs in Joshi & Vijay-Shanker, 1999; Kallmeyer & Joshi, 1999). In a non-local grammar, the derivation structure does not directly determine a suitable dependency structure. In some formalisms, it is possible to identify parts of elementary structures that are relevant for the dependency structure (e.g. in D-Tree Grammars, Rambow *et al.*, 1995, the relevant part is the part of a d-tree that is substituted in a subsertion operation). But there is not one single structure that records the complete derivation and that is a suitable dependency structure.

As an alternative, I propose to use local Tree Description Grammars (local TDG, Kallmeyer, 1997; Kallmeyer, 1999). Local TDGs generate tree descriptions with a local derivation process. They have a context-free derivation structure and generate only semilinear languages. The descriptions generated by local TDGs allow an underspecification of the dominance relation, and the construction of so-called minimal trees for these descriptions is not subject to locality constraints. This limited amount of non-locality allows to deal with scrambling, as illustrated by a local TDG for some German scrambling and extraposition data.

2. Scrambling: The data

The paper accounts for data like word order variations of (1), taken from (Rambow, 1994).

(1) Weil niemand das Fahrrad zu reparieren zu versuchen verspricht
because nobody the$_{acc}$ bike$_{acc}$ to repair to try promises
because nobody promises to try to repair the bike

Assuming that each NP precedes its verb, we get 30 word orders when combining scrambling with extraposition. According to Rambow, 6 of them are clearly not acceptable. The other 24 also show differences with respect to the judgment, but in principle it should be possible to generate them all. The word orders without extraposition and their judgments are shown in (2). Word orders that are ruled out occur with extraposition of *reparieren* as in (3).

(2) a. ok Weil niemand das Fahrrad zu reparieren zu versuchen verspricht
b. ? Weil das Fahrrad niemand zu reparieren zu versuchen verspricht
c. ok Weil das Fahrrad zu reparieren niemand zu versuchen verspricht
d. ? Weil das Fahrrad zu reparieren zu versuchen niemand verspricht

(3) a. * Weil zu versuchen das Fahrrad niemand zu reparieren verspricht
b. * Weil das Fahrrad zu versuchen niemand zu reparieren verspricht
c. * Weil zu versuchen niemand das Fahrrad zu reparieren verspricht
d. * Weil niemand zu versuchen das Fahrrad verspricht zu reparieren
e. * Weil zu versuchen niemand das Fahrrad verspricht zu reparieren
f. * Weil zu versuchen das Fahrrad niemand verspricht zu reparieren

I will also consider more than two levels of embedding as in (4).

(4) weil das Fahrrad niemand glaubt zu reparieren zu versuchen versprechen zu müssen
because the$_{acc}$ bike$_{acc}$ nobody thinks to repair to try promise to need
because nobody thinks it necessary to promise to try to repair the bike

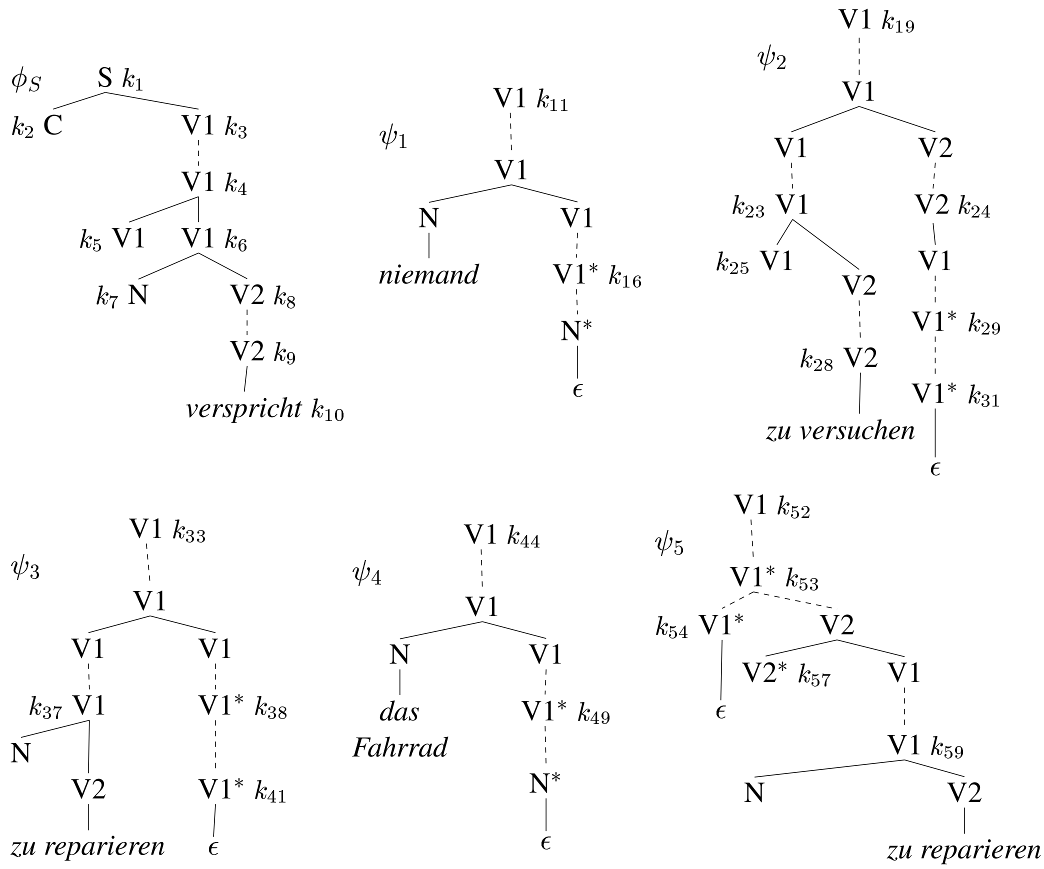
3. A local TDG for scrambling

Local TDGs consist of tree descriptions (*elementary descriptions*) and a *start description*. The tree descriptions are negation and disjunction free formulas in a quantifier-free first order logic. The logic allows to express relations between node names k_1, k_2 such as immediate dominance $k_1 \lhd k_2$, dominance (reflexive transitive closure of $\lhd$) $k_1 \lhd^* k_2$, linear precedence $k_1 \prec k_2$ and equality $k_1 \approx k_2$. Furthermore, nodes are supposed to be labelled by terminals or by atomic feature structures. δ denotes the labeling function, $\delta(k) \approx t$ signifies that k has a terminal label t, and $a(\delta(k)) \approx v$ signifies that k is labelled by a feature structure containing the attribute value pair $\langle a, v \rangle$. Roughly, tree descriptions in a local TDG are fully specified (sub)tree descriptions that are connected by dominance relations.[1] In elementary descriptions, some node names are *marked*; this is important for the derivation. In the graphical representations, marked names are equipped with an asterisk.

(5) shows a local TDG for some scrambling data with $\phi_S = k_1 \lhd k_2 \wedge k_1 \lhd k_3 \wedge k_2 \prec k_3 \wedge k_3 \lhd^* k_4 \wedge \ldots \wedge cat(\delta(k_1)) \approx S \wedge \ldots$ etc. (dotted edges represent dominance relations). Conjuncts as $k_3 \lhd^* k_4$ in ϕ_S not entailed by the rest of the formula are called *strong dominance*.

[1]Some of the conditions holding for descriptions in a local TDG are left aside here. For a formal definition of local TDGs see (Kallmeyer, 1999, Chapter 4).

(5)



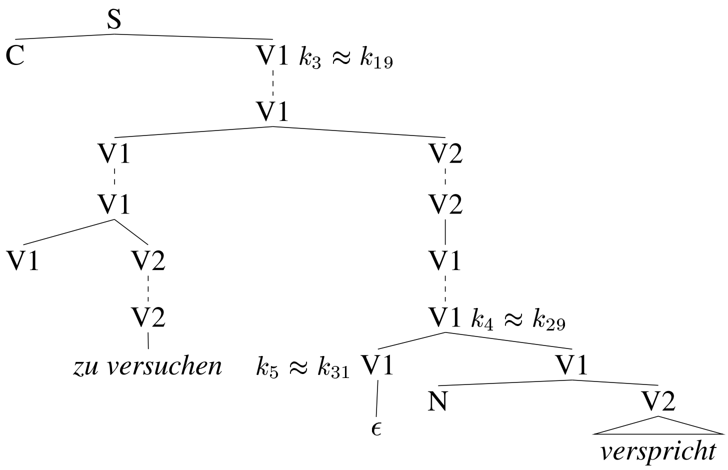
The labels V1 and V2 distinguish between VPs not allowing extraposed material to attach (V1) and VPs that allow this (V2). ψ_5 is an elementary description used for an extraposed clause. In the following we will see how the descriptions in (5) combine with each other.

4. Local derivation and underspecification

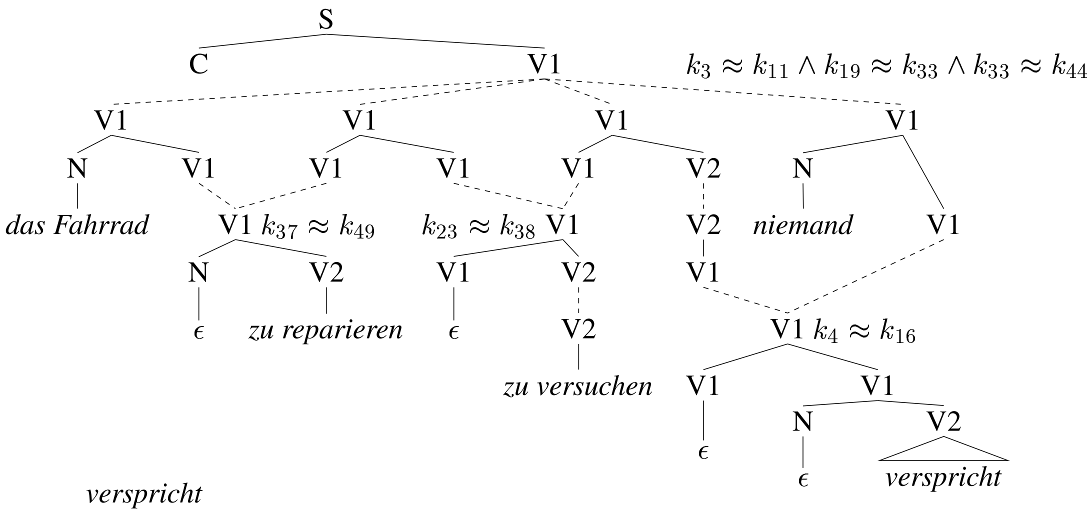
Derivations start with ϕ_S. In each step, an old ϕ_1 and an elementary ψ are combined to obtain a new ϕ_2. ϕ_2 can be viewed as a conjunction of ϕ_1, ψ and new formulas $k \approx k'$ where k is a name from ϕ_1 and k' a name from ψ. This derivation step must be such that

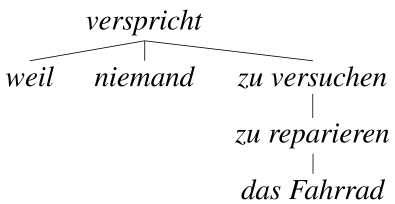
1. for a node name k_ψ in ψ, there is a new equivalence iff k_ψ is either marked or minimal (dominated by no other name, e.g. k_1 in ϕ_S in (5)),
2. a marked or minimal k' in ψ that is not a leaf name (i.e. dominates other names) but does not dominate any other marked name becomes equivalent to a leaf name in ϕ_1
3. the names k from ϕ_1 used for new equivalences are part of one single elementary or start description, the *derivation description* of this step (first locality condition),
4. for each marked name k_ψ in ψ with a parent, there is a strong dominance $k_1 \lhd^* k_2$ in ϕ_1 such that $k_2 \approx k_\psi$ is added and the subdescription between k_ψ and the next marked or minimal name dominating k_ψ is dominated by k_1 (second locality condition),
5. and the result ϕ_2 is maximally underspecified.

The 3. and 4. condition express the locality of the derivations. They are comparable to the locality constraint on derivations in set-local MCTAG. In fact, for each set-local MCTAG, an equivalent local TDG can be constructed in a straight-forward way (see Kallmeyer, 1999).
As a sample derivation step consider adding ψ_2 to ϕ_S in (5) which leads to $\phi_S \wedge \psi_1 \wedge k_3 \approx k_{19} \wedge k_4 \approx k_{29} \wedge k_5 \approx k_{31}$ in (6).

(6) 

If a marked name has no parent, an underspecification of the dominance can occur. The fifth condition then ensures that the most general solution is generated. E.g., adding ψ_1, ψ_3 and ψ_4 to (6) with derivation descriptions ϕ_S, ψ_2 and ψ_2 respectively gives (7). The derivation structure of (7), shown in (8), is the correct dependency structure.

(7) 

(8) 

Descriptions generated by a local TDG G denote infinitely many trees. The *tree language* of G contains "minimal" trees of these descriptions. A *minimal* tree γ of a description ϕ satisfies ϕ in such a way that all subtrees of heigth 1 of γ are described exactly once in ϕ. The minimal trees of (7) yield the strings in (2).

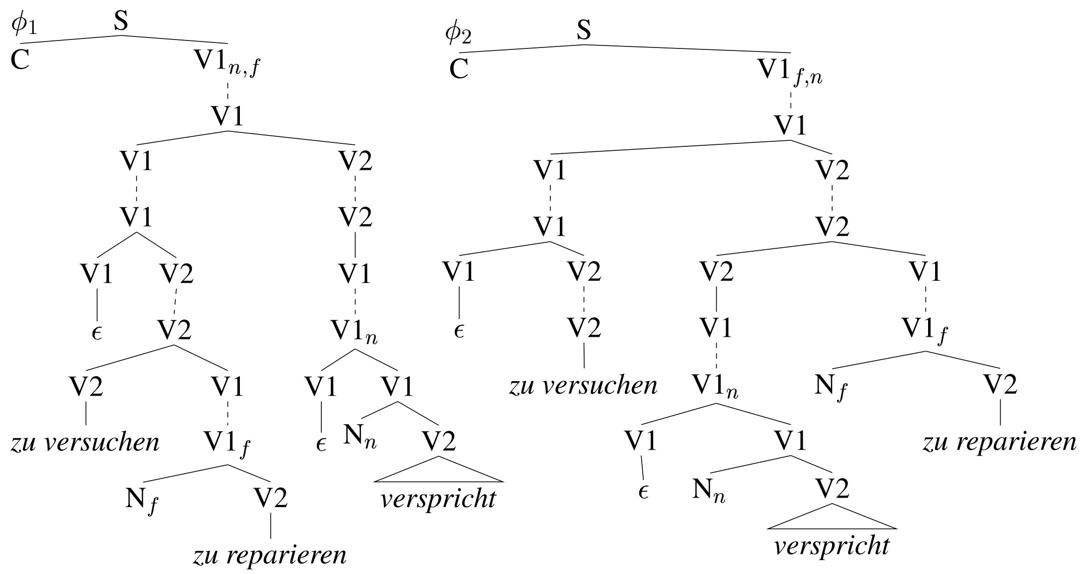
The possibility of underspecification increases the expressive power of local TDGs beyond LCFRS. However, despite this additional power, it is possible to find a context-free derivation grammar and thereby to show that the languages generated by local TDGs are semilinear.

5. Scrambling and extraposition

In ψ_2 there are two attachment sites (label V2) for extraposed clauses, k_{28} and k_{24}. This accounts for the different cases of extraposing *zu reparieren* (i) only past *zu versuchen* and (ii) past *zu versuchen* and *verspricht*. For extraposed VPs, elementary descriptions like ψ_5 for *zu reparieren* in (5) are needed. Adding ψ_5 to (6) with derivation description ψ_2 either leads to ϕ_1 or to ϕ_2 in (9). The subscripts $_n$ and $_f$ mark the names chosen for new equivalences when adding ψ_1 and

ψ_4 for *niemand* and *das Fahrrad* respectively. With ϕ_1, *niemand* is either left of all verbs or between *zu reparieren* and *verspricht*, which excludes (3)a., b. and c. With ϕ_2, *das Fahrrad* is either between *verspricht* and *zu reparieren* or left of all verbs. This excludes (3)d., e. and f.

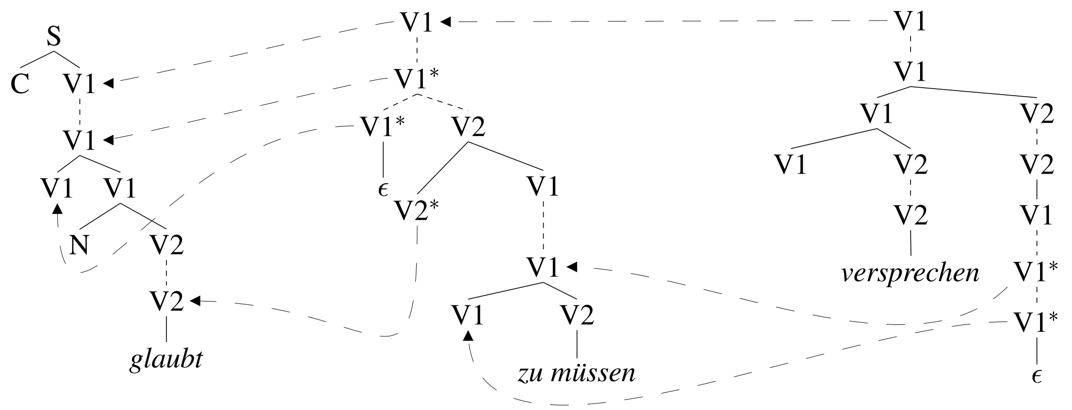
(9)



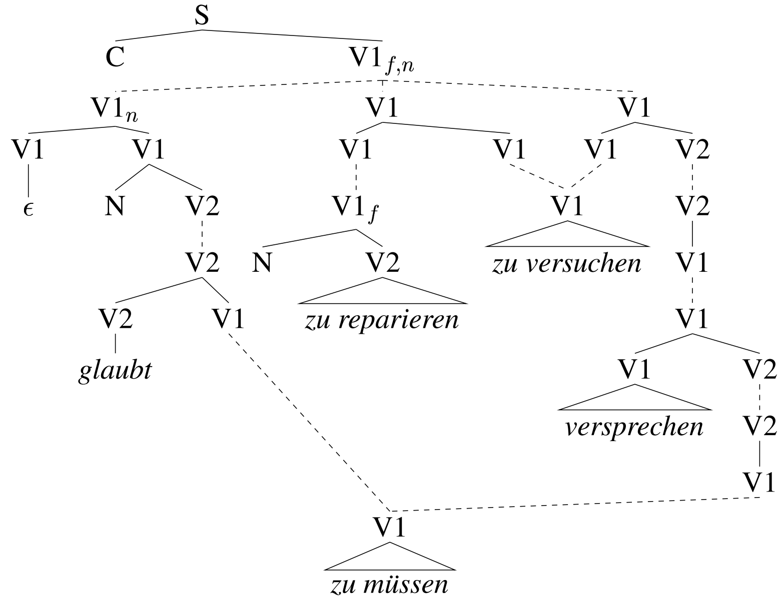
6. More than two levels of embedding

So far, we have considered only examples with up to two levels of embedding. Next, I will consider the analysis of (4), a sentence with four levels of embedding.

(10)



(11)



First, elementary descriptions for *glaubt*, *zu müssen* and *versprechen* are put together as sketched in (10). Then ψ_2 and ψ_3 from (5) for *zu versuchen* and *zu reparieren* are added which leads to (11). Further adding ψ_1 and ψ_4 gives a description that is such that in the minimal trees, *glaubt* is left of *zu müssen*, *zu reparieren* is left of *zu versuchen* which is left of *versprechen*, and *versprechen* is left of *zu müssen*. Furthermore, *niemand* is left of *glaubt* and *das Fahrrad* is left of *zu reparieren*. One of the minimal trees yields (4).

7. Conclusion

This paper addresses the problem that on the one hand, long-distance scrambling in German seems to be non-local in a limited way. On the other hand, there are good reasons to prefer a grammar with a local derivation process that leads to an appropriate dependency structure. I have proposed local TDGs as an alternative to other formalisms previously used to deal with scrambling. Local TDGs have the desired locality property but allow underspecification of the dominance relation. The construction of minimal trees is not subject of any locality constraint. Therefore, local TDGs show a very limited amount of "non-locality", which gives sufficient expressive power to account for scrambling phenomena. This was illustrated by a local TDG analysis of some German data.

Acknowledgments

I would like to thank two referees of this paper for their very valuable comments, which helped to improve the content of this paper.

References

BECKER T., JOSHI A. K. & RAMBOW O. (1991). Long-distance scrambling and tree adjoining grammars. In *Proceedings of ACL-Europe*.

BECKER T., RAMBOW O. & NIV M. (1992). *The Derivationel Generative Power of Formal Systems or Scrambling is Beyond LCFRS*. Technical Report IRCS-92-38, University of Pennsylvania.

JOSHI A. K., LEVY L. S. & TAKAHASHI M. (1975). Tree Adjunct Grammars. *Journal of Computer and System Science*, **10**, 136–163.

JOSHI A. K. & VIJAY-SHANKER K. (1999). Compositional Semantics with Lexicalized Tree-Adjoining Grammar (LTAG): How Much Underspecification is Necessary? In H. C. BLUNT & E. G. C. THIJSSE, Eds., *Proceedings ot the Third International Workshop on Computational Semantics (IWCS-3)*.

KALLMEYER L. (1997). Local Tree Description Grammars. In *Proceedings of the Fifth Meeting on Mathematics of Language, DFKI Research Report*, p. 77–84.

KALLMEYER L. (1999). *Tree Description Grammars and Underspecified Representations*. PhD thesis, Universität Tübingen. Technical Report IRCS-99-08, University of Pennsylvania, Philadelphia.

KALLMEYER L. & JOSHI A. K. (1999). Factoring Predicate Argument and Scope Semantics: Underspecified Semantics with LTAG. In P. DEKKER, Ed., *12th Amsterdam Colloquium. Proceedings*.

RAMBOW O. (1994). *Formal and Computational Aspects of Natural Language Syntax*. PhD thesis, University of Pennsylvania.

RAMBOW O. & LEE Y.-S. (1994). Word order variation and Tree-Adjoining Grammars. *Computational Intelligence*, **10** (4 p.), 386–400.

RAMBOW O., VIJAY-SHANKER K. & WEIR D. (1995). D-Tree Grammars. In *Proceedings of ACL*.

WEIR D. J. (1988). *Characterizing mildly context-sensitive grammar formalisms*. PhD thesis, University of Pennsylvania.