FG 2007: The 12th conference on Formal Grammar Dublin, Ireland August 4-5, 2007

Organizing Committee:
Laura Kallmeyer Paola Monachesi
Gerald Penn Giorgio Satta

CENTER FOR THE STUDY
OF LANGUAGE
AND INFORMATION

July 10, 2007

1

An MCTAG with Tuples for Coherent Constructions in German

TIMM LICHTE †

Abstract

This paper introduces the notion of tree tuples to MCTAG, an extension of Tree Adjoining Grammar (TAG). Using tuples and node sharing we can provide an empirically broad and linguistically sound analysis of coherent constructions and scrambling in German, without the compulsive use of traces or the additional descriptive means in former MCTAG approaches.

Keywords TREE ADJOINING GRAMMAR, SHARED NODES, TREE TUPLES, COHERENT CONSTRUCTIONS, SCRAMBLING, GERMAN

1.1 Introduction

Coherent constructions in German (see 1.2 for examples) pose a well studied challenge for their representation in grammar frameworks such as HPSG, CCG, or TAG. This is due to two properties of coherent constructions: (1) the interaction of the different subcategorization frames of the involved verbs, and (2) the free permutability of their depending constituents in the sentence, also known as *scrambling*.

Depending on the kind of grammar framework and the way, subcategorization frames are implemented, there exist different representational perspectives on coherent constructions, that can be boiled down to two paradigms (Rambow, 2003), namely sharing one subcategorization frame ("incorporation analysis") and sharing positions in a sentence ("syntactic analysis"). The first one is found, e.g., in argument

†Thanks to Laura Kallmeyer, Wolfgang Maier, Anders Søgaard, and the four reviewers.

FG-2007.
Organizing Committee:, Laura Kallmeyer, Paola Monachesi, Gerald Penn, Giorgio Satta.

raising for HPSG (Hinrichs and Nakazawa, 1994) and functional composition for CCG (Steedman, 1984). The second perspective is propagated in TAG approaches, and the one we will deal with in this paper.

Since standard TAG has to be extended in order to cover coherent constructions and scrambling data from German (under a certain analysis, see 1.3), several extensions of TAG have been proposed. Among them, Vector MCTAG (V-TAG, Rambow, 1994) and Multicomponent TAG with shared nodes (SN-MCTAG, Kallmeyer, 2005) have been shown to provide means that are powerful enough, and still to be computationally tractable. However, the extensions reveal both virtues and drawbacks: V-TAG does not impose traces, yet needs additional mechanisms in order to tame the very powerful (and computationally unappealing) core of the formalism; SN-MCTAG, on the other hand, is in fact computationally appealing without changing the face of the TAG formalism as substantially as V-TAG. However, it has to make use of traces, even for elements that, in a standard movement-based theory, would not be considered as moved (Kallmeyer, 2005, § 3.1). This contribution aims at introducing TT-MCTAG, a formalism that exhibits the virtues of V-TAG and SN-MCTAG, without carrying the mentioned drawbacks.

1.2 Scrambling in Coherent Constructions

The term *scrambling* (Ross, 1967) originally stems from transformational syntax theories and there denotes the movement of arguments and adjuncts from their base position within a certain syntactic domain, i.e., the *Mittelfeld* for German (Frey, 1993). In contrast, we use the term scrambling descriptively, having in mind a flexible word order without underlying base configuration as demonstrated in the following two (verb-final) sentences from German:[1]

(1) a. dass [Peter] heute [den Kühlschrank] repariert hat
that Peter$_{nom}$ today the fridge$_{acc}$ repaired has
'that Peter has repaired the fridge today'

b. dass [den Kühlschrank] heute [Peter] repariert hat
that the fridge$_{acc}$ today Peter$_{nom}$ repaired has

Both sentences carry the same denotation, but vary with respect to the position of the subject *Peter* and the direct object *den Kühlschrank* ('the fridge').[2]

[1] Potential differences in acceptability are omitted for the time being.

[2] There is even more variability, since also the modifier *heute* ('today') can participate, which leads to six differing word orders, still with the same denotation.

Scrambling is also available in *coherent constructions*, for which Gunnar Bech established the standard descriptive analysis and terminology (Bech, 1955, Kiss, 1995). In coherent constructions, all verbs, say V_0 to V_n, are connected in a subcategorization chain, where V_i subcategorizes for (or "governs") a non-finite verb V_{i+1}. In the example in (2), the finite complex verb *versprochen hat* governs the non-finitive status of *zu reparieren.* Furthermore, the verbs V_0 to V_n in general appear next to each other and make up the verb complex.

(2) dass [den Kühlschrank] Peter [zu reparieren] versprochen hat
that the fridge Peter to repair promised has
'that Peter has promised to repair the fridge'

Without providing an explanation such as movement up the subcategorization chain ("long-distance scrambling", Rambow, 1994), Bech posits a joint access to the Mittelfeld by the participating verbs: constituents depending syntactically/semantically on V_0 to V_n may be placed in the Mittelfeld in any order.

In this paper, we will not be concerned with the structure of the verb complex. We will, however, pay attention to the so called *third construction* (den Besten and Rutten, 1989), of which an instance is given in (3):

(3) dass [den Kühlschrank] Peter versprochen hat [zu reparieren]
that the fridge Peter promised has to repair
'that Peter has promised to repair the fridge'

The extraposition of the clause *den Kühlschrank zu reparieren* is partial: the non-finite verb *zu reparieren* is extraposed, while its complement *den Kühlschrank* remains in the Mittelfeld. The mirror image of partial extraposition, where the *zu*-infinitive is fronted in verb-second clauses, is also briefly touched in section 1.7.

Having pictured the kind of data that we want to model, we now turn to a description of the grammar formalism TT-MCTAG. We will do so gradually, starting with a description of TAG in the next section, of which TT-MCTAG is an extension.

1.3 Tree Adjoining Grammars and their Limitations

A Tree Adjoining Grammar (TAG, Joshi et al., 1975) consists of a set of *elementary trees* that can be combined via *substitution* (replacing one leaf with a new tree) and *adjunction* (replacing an internal node with a new tree). Elementary trees that are able to adjoin are called *auxiliary trees.* They have exactly one non-terminal leaf which is called

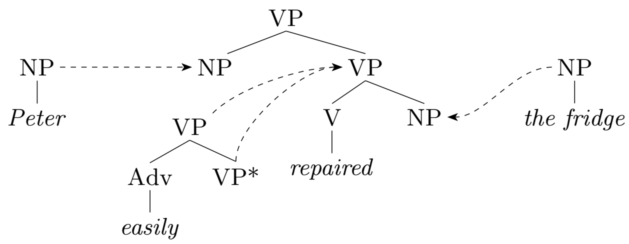

FIGURE 1 Sample TAG derivation of "Peter easily repaired the fridge".

the *footnode* and which is commonly decorated by an asterisk (*). In contrast, elementary trees which do not adjoin and do not have a footnode are called *initial trees*. Auxiliary and initial trees may contain one or more non-terminal leaves that allow (and in fact force) substitution. A simple TAG fragment is given in Fig. 1. Here, the transitive verb *repaired* receives its nominal complements through substitution at the NP-leaves, whereas the adverbial modifier *easily* is adjoined to the internal VP-node of the tree of *repaired*.[3]

V-TAG, SN-MCTAG and TT-MCTAG fall back on the notion of *multicomponent TAG* (MCTAG, Joshi, 1987, Weir, 1988). An MCTAG additionally lets one declare tree sets consisting of elementary trees, meaning two things: firstly, using a tree set implicates using all the trees belonging to it synchronously; secondly, the attachment (i.e. adjunction or substitution) of the trees of a tree set can be restricted with respect to the place of attachment: if the trees of a tree set are attached to the same elementary tree, the MCTAG is called *tree-local*; if they are attached to the same tree set, the MCTAG is called *set-local*; otherwise (i.e. without attachment restriction) the MCTAG is called *non-local*. Tree-local and set-local MCTAG are polynomially parsable (the former are even strongly equivalent to simple TAG) while non-local MCTAG are NP-complete (Rambow and Satta, 1992).

As Becker et al. (1991) point out, coherent constructions that cannot be generated by a CFG or TAG are easily found, if the co-occurrence constraint holds that complements and verbs are introduced in the same elementary tree. In contrast, both V-TAG and SN-MCTAG offer a linguistically appropriate expressive power for modeling coherent constructions and will be examined in the following section.

[3]Note that all elementary trees in the sample have a lexical item as terminal. TAGs that have this further property are called *lexicalized* TAG (LTAG). We will use the term TAG in the sense of LTAG.

1.4 Extensions of TAG: V-TAG and SN-MCTAG

1.4.1 V-TAG

At its core, a V-TAG (Vector MCTAG, Rambow, 1994) resembles a non-local MCTAG. However, the trees of a tree set may be used non-synchronously, such that computational tractability is met again. Moreover, in order to express locality restrictions, dominance links and integrity constraints are added. Due to non-locality, a verb and its complements can be represented as separate elementary trees in a tree set, as shown in Fig. 2. Dominance links, indicated by dashed arrows, impose a dominance relation between the connected nodes in the derived tree. A node with an integrity constraint is decorated with a triangle, which imposes that no dominance link runs across it. Integrity constraints thus act as barriers for scrambling and topicalization.

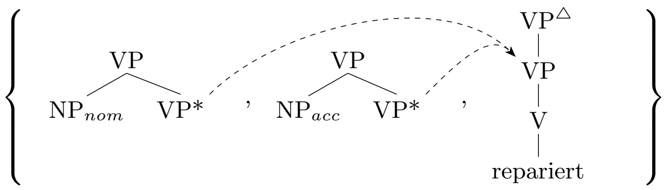

FIGURE 2 V-TAG tree set for the finite verb *repariert* ('repairs') and its complements, a nominative and an accusative NP.

1.4.2 SN-MCTAG

In contrast to V-TAG, the formalism of SN-MCTAG (Multicomponent Tree-Adjoining Grammar with Shared Nodes) (Kallmeyer, 2005) rests on tree-local MCTAG. It relaxes the notion of tree-locality (instead of restricting non-locality) by allowing certain derivation trees not permitted in standard MCTAG. The idea of node sharing is the following: other than standard MCTAG (and TAG), after adjunction at a node n, n is seen to be part of both the adjoining tree and the destination tree. Thus, when using a tree set $\{\beta_1, ..., \beta_n\}$, its elementary trees may also attach to different trees, as long as these trees are in a node sharing relation ($\mathcal{SN}_D$) to the same elementary tree, say δ. The node sharing relation holds for $< \delta, \beta_i >$, iff β_i is immediately attached to δ, or β_i is adjoined to the root node of some tree δ' for which the node sharing relation $< \delta, \delta' >$ holds. Kallmeyer (2005) proves that a restricted variant of SN-MCTAG is polynomially tractable.[4]

[4]This restricted SN-MCTAG, RSN-MCTAG with fixed arity, requires (1) at least one tree of a tree set to attach directly to the δ-tree and (2) the number of inter-

Node sharing allows to state constraints for scrambling (and topicalization) more naturally in terms of TAG, namely by the means of elementary tree structure and derivation tree structure: substitution establishes a barrier for movement, while adjunction widens the domain of locality . As already mentioned, however, SN-MCTAG is well suited for the modeling of word order variation in German, only if a base word order, movement and traces are among the linguistic assumptions. The valency of a verb is expressed through the base position of its complements in its elementary tree. Any deviation has to be generated by tree sets, consisting of the moved constituent and its trace. Fig. 3 contains an example. Note that dashed arrows here again indicate attachment operations.

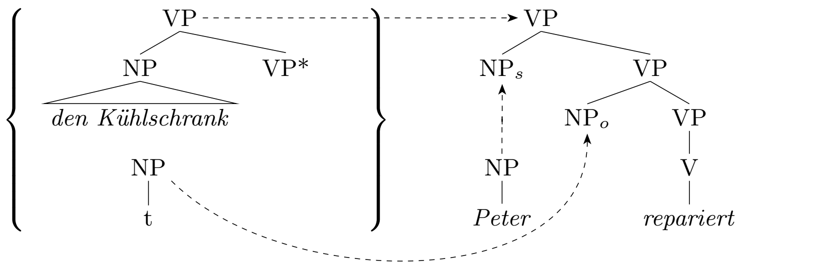

FIGURE 3 A tree-local MCTAG analysis of the German sentence "*[dass] den Kühlschrank Peter repariert*" ('[that] Peter repairs the fridge'). The object noun *den Kühlschrank* ('the fridge') is fronted, leaving a trace in its base position behind the subject NP.

1.5 TT-MCTAG: SN-MCTAG with Tree Tuples

As mentioned in the last section, SN-MCTAG forces the grammar writer to use traces, even in the case of free word order languages, while with V-TAG the grammar writer has to use rather TAG-alien dominance links and integrity constraints. To overcome these drawbacks, we introduce a new structured object of multiple elementary trees, namely *tree tuples*. A tree tuple consists of two components: a lexicalized elementary tree δ (the destination tree), and a tree set $\{\beta_1, ...\beta_n\}$, written as $< \delta, \{\beta_1, ...\beta_n\} >$. The meaning of such a tree tuple is the following: during derivation, the trees from the tree set have to attach to the elementary tree δ, or to a tree that is in a node sharing relation to δ. Formally, $< \delta, \beta_1 >, ..., < \delta, \beta_n > \in \mathcal{SN}_D$. Other than tree sets in SN-MCTAG, however, the tree set of a tree tuple need not be re-

secting tree sets to have a fixed maximum.

solved synchronously. We can now encode the valency of the German transitive verb *reparieren* ('to repair') in the way presented in Fig. 4.

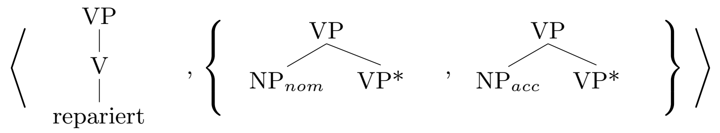

FIGURE 4 Tree tuple for the verb *repariert* ('repairs'). The destination tree is the verb and its projection, and the accompanied tree set contains the complements of the verb.

As Søgaard et al. (2007) show, the computational complexity of unrestricted TT-MCTAG is rather unappealing: the universal recognition problem of TT-MCTAG is NP-complete. Furthermore the MIX language can be generated, which is considered not to be mildly context sensitive. In order to retain polynomial time complexity, Søgaard et al. suggest to restrict (1) the size of the inner tree set, and (2) the number of concurrently active tree tuples to some fixed maximum.

1.6 An Analysis for Coherent Constructions

In this section, we present an analysis of a sample of German coherent constructions by means of TT-MCTAG. Because scrambling is intimately connected with coherent constructions in German, we also give an analysis for this phenomenon. We concentrate on *zu*-infinitivals, since they allow also for partial extraposition ("third construction"). The following analysis for coherent constructions is essentially the same for bare infinitivals.

First, we turn to a standard case of coherence: a subordinate clause with verb-final structure, which was already presented in (2), repeated here as (4):

(4) dass den Kühlschrank Peter zu reparieren versprochen hat.

While they can freely permutate in the Mittelfeld, the object noun *den Kühlschrank* comes from the non-finite verb *zu reparieren*, and the subject noun *Peter* is introduced by the finite (complex) verb *versprochen hat.* The tree tuples that license this clause and all its scrambled relatives are depicted in Fig. 5. Note that we are not assuming *zu reparieren* to have a PRO subject, and we omit the initial trees for the nouns for lack of space.

Under the coherent analysis, the non-finite *zu reparieren* is inserted into the verbal complex dominated by *versprochen hat.* The tree of *zu reparieren* is therefore of (bottom) category V, as well as the footnode of the tree of *versprochen hat*, that adjoins into *zu reparieren.* Crucially, the complement tree of the *zu reparieren* tuple cannot adjoin to the destination tree, because the destination tree does not offer an appropriate node for adjunction. Instead, the tree of *versprochen hat* offers a VP node via node sharing. The VP node thus acts as a marker of the Mittelfeld.

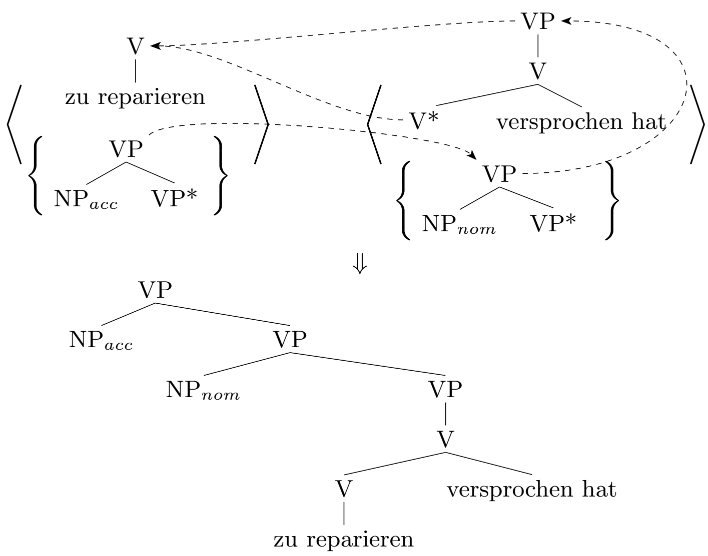

FIGURE 5 Tree tuples and the derived tree for a (verb-final) coherent construction.

In contrast to bare infinitives, *zu*-infinitives can be partially extraposed as in (3), repeated here in (5) with the extra adverb *heute* ('today'):

(5) dass den Kühlschrank Peter versprochen hat, heute zu reparieren

The complements and adjuncts of the non-finite *zu reparieren* may appear in an extraposed position together with the verb, such as *heute*, or in the Mittelfeld of the governing verb, such as *den Kühlschrank.* We respect these facts, as the tuples for both *zu reparieren* and *versprochen hat* are modified in Fig. 6. Firstly, the root node of the *zu reparieren* tree has now the category VP, such that the object noun, or any adjunct, can directly adjoin to it. In fact, this is also the tree tuple for the incoherently used *zu reparieren.* Secondly, the tree for *versprochen hat* adjoins to a VP-node on the right of the lexical anchor *versprochen.*

Thereby, it also makes the upper VP-node available for complements or adjuncts of the extraposed *zu*-infinitive.

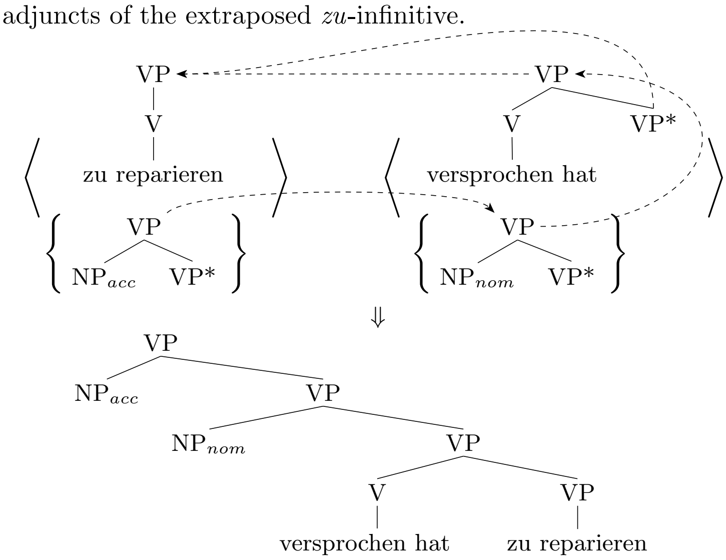

FIGURE 6 Tree tuples and the derived tree for (verb-final) partial extraposition.

1.7 Beyond Node Sharing?

In the last section, tree tuples were shown, that, in combination with node sharing, account for coherent constructed verbs and all the scrambled configurations of their complements and adjuncts. However, we were only considering verb-final clauses. If we broaden the view and also consider verb-second clauses (clauses that have the finite verb in second position, whereas non-finite parts of the verb complex appear in verb-final position), we encounter at least two configurations that challenge the node sharing approach somewhat.

The first one applies for rather standard verb-second clauses as in (6):

(6) Heute hat [den Kühlschrank] Peter [zu reparieren] versprochen.
today has the fridge Peter to repair promised
'Today Peter has promised to repair the fridge.'

Under the assumption that the finite temporal auxiliary *hat* in verb-second position is introduced together with the main verb *versprochen* in a destination tree, the corresponding tree tuples might look like in

Fig. 7. This analysis, however, is not supported by node sharing, since the nominal object of *zu reparieren* has no access to the lower VP-node (representing the Mittelfeld), but only to the root node of the *hat-versprochen* tree. One way to make the lower VP-node accessible would be to separate *hat* from *versprochen*, and to let *hat* adjoin to the VP-node of *versprochen*. Another option (that maintains the flexibility for the grammar writer) is to relax node sharing towards *spine sharing*. A shared spine includes all nodes between the footnode and the root node. A formal definition and a formal examination of its complexity, which should be within mildly contextsensitivity, is in preparation.

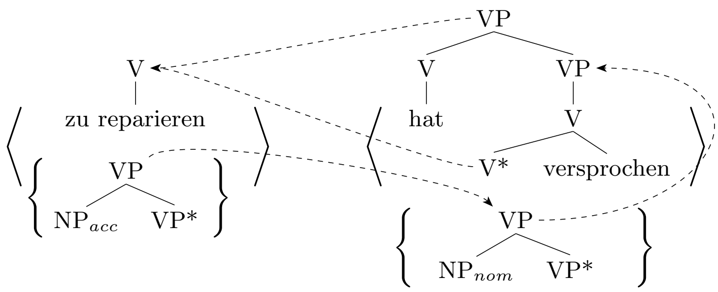

FIGURE 7 Tree tuples for a verb-second coherent construction.

The second challenging group of configurations is made up of fronting phenomena. An example is given in (7):

(7) Zu reparieren hat Peter den Kühlschrank versprochen.
to repair has Peter the fridge promised
'Peter has promised to repair the fridge.'

The governed non-finite verb *zu reparieren* is fronted/topicalized to the left of the finite part *hat* of the main verb. Its object noun *den Kühlschrank*, however, remains in the Mittelfeld. To illustrate the problem that thereby arises, two possible destination trees (δ_1, δ_2) of fronting *hat-versprochen* tuples are depicted in Fig. 8. It easily can be seen that in δ_1 no spine sharing relation can be established between the footnode and the lower VP-node, where the object noun of the fronted *zu*-infinitive is supposed to adjoin. In order to place the lower VP-node onto the shared spine, one could imagine to lift it as done in δ_2. This, however, also forces the complement trees to have a right-branching shape, for both governing and governed verb. Furthermore it looks somewhat non-standard and ad hoc.

Since tree δ_1 is more desirable than δ_2, a further extension of sharing

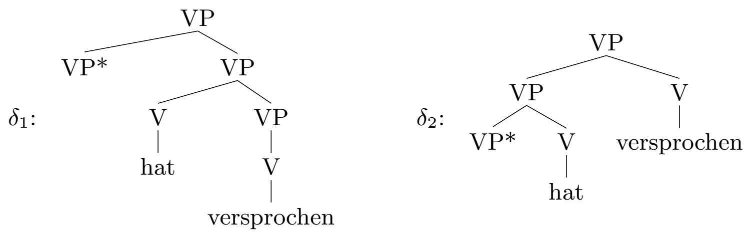

FIGURE 8 Alternative destination trees for the finite verb *hat-versprochen*, which license a fronted *zu*-infinitive.

could be implemented for its rescue. The next natural level of sharing is *tree sharing*. As the term suggests, not only the footnode spine, but the whole auxiliary tree is shared. However, the complexity of this sharing variant is unclear, and we are pessimistic that it reveals the desired computational properties. As no trivial solution is in sight, a somehow relieving fact is that also V-TAG faces severe problems with this kind of constructions (Rambow, 2003).

1.8 Conclusions

The presented MCTAG analysis of coherent constructions and scrambling in German introduces the notion of tree tuples, referred to as TT-MCTAG. As shown, tree tuples qualify for a concise and yet abstracted representation of subcategorization frames in German, which along with node sharing enable a straightforward approach to flexible word order phenomena in terms of TAG: substitution acts as a barrier for permutation, while adjunction widens the domain of locality. Also the difference between coherent constructions and incoherent constructions can be accounted for via the substitution/adjunction dualism. Concurrently, traces are not compulsory as in original SN-MCTAG. Thus, TT-MCTAG incorporates the virtues of SN-MCTAG and V-TAG, while avoiding their drawbacks.

References

Bech, Gunnar. 1955. *Studien über das deutsche Verbum infinitum*. 2nd unrevised edition published 1983 by Max Niemeyer Verlag, Tübingen.

Becker, Tilman, Aravind K. Joshi, and Owen Rambow. 1991. Long-distance scrambling and tree adjoining grammars. In *Proceedings of the 5th Conference of the European Chapter of the Association for Computational Linguistics*.

den Besten, Hans and Jean Rutten. 1989. On verb raising, extraposition and free word order in dutch. In D. Jaspers, W. Klooster, Y. Putseys, and P. Seuren, eds., *Sentential Complementation and the Lexicon*. Foris, Dordrecht.

Frey, Werner. 1993. *Syntaktische Bedingungen für die semantische Interpretation. Über Bindung, implizite Argumente und Skopus*. studia grammatica XXXV. Akademie Verlag, Berlin.

Hinrichs, Erhard and Tsuneko Nakazawa. 1994. Linearizing AUXs in German verbal complexes. In J. Nerbonne, K. Netter, and C. Pollard, eds., *German in Head-Driven Phrase Structure Grammar*, pages 11–37. CSLI,Stanford.

Joshi, Aravind K. 1987. An introduction to tree adjoining grammars. In A. Manaster-Ramer, ed., *Mathematics of Language*, pages 87–114. John Benjamins, Amsterdam.

Joshi, Aravind K., Leon S. Levy, and Masako Takahashi. 1975. Tree adjunct grammars. *Journal of Computer and System Science* 10:136–163.

Kallmeyer, Laura. 2005. Tree-local multicomponent tree adjoining grammars with shared nodes. *Computational Linguistics* 31:2:187–225.

Kiss, Tibor. 1995. *Infinite Komplementation - Neue Studien zum deutschen Verbum Infinitum*. Max Niemeyer Verlag, Tübingen.

Rambow, Owen. 1994. *Formal and Computational Aspects of Natural Language Syntax*. Ph.D. thesis, University of Pennsylvania, Philadelphia. IRCS Report 94-08.

Rambow, Owen. 2003. Coherent constructions in german: Lexicon or syntax? In P. Seuren and G. Kempen, eds., *Verb Constructions in German and Dutch*. Benjamins, Amsterdam.

Rambow, Owen and Giorgio Satta. 1992. Formal properties of non-locality. In *Proceedings of 1st International Workshop on Tree Adjoining Grammars*.

Ross, John Robert. 1967. *Constraints on Variables in Syntax*. Ph.D. thesis, MIT. Reproduced by the Indiana University Linguistics Club.

Søgaard, Anders, Timm Lichte, and Wolfgang Maier. 2007. The complexity of linguistically motivated extensions of tree-adjoining grammar. In *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing (RANLP 2007)*. To appear.

Steedman, Mark. 1984. A categorial theory of intersecting dependencies in Dutch infinitival complements. In W. de Geest and Y. Putseys, eds., *Sentential Complementation*, pages 215–226. Foris, Dordrecht.

Weir, David J. 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms*. Ph.D. thesis, University of Pennsylvania.