DTAG?

Kim Gerdes
Lattice, Université Paris 7

1. Introduction

This paper reports on work in progress on the creation of a metagrammar for German verbal constructions. Section 2 circumscribes the field we are working on: We describe known and less known problems and try to delineate the limits of a standard Tree Adjoining Grammar for German; we see which structures we get easily, and which structures we will never get. In Section 3 we put the German data in perspective to the other TAG languages, French and English, and we define and justify our choice to create a limited German TAG, designed for a specific generation task. We then propose in Section 4 some of the possible elementary trees that can live up to our expectations: Compromising on the semantic interpretability of the derivation structure as well as on the principles underlying TAG allows us to get most of the word orders necessary for the generation task. Yet, what we gain in usefulness in generation, we pay in linguistic descriptiveness. Last but not least, in Section 5 we discuss problems and limits of the metagrammar implementation and we give some indication on how the desired elementary tree sketches can be created and maintained with a metagrammar.

We presuppose the comprehension of the terms *metagrammar* and *topological field* of a German sentence, which we cannot define here. For details, see Gaiffe et al. 2002 and Kathol 1995 respectively.

2. Scrambled Minds

The two existing Tree Adjoining Grammars of interesting grammatical coverage have been created for English (XTAG, 1995) and French (FTAG, Abeillé 1991), two languages with quite rigid word order and little case marking. On the other hand it has long been shown that German is beyond the (derivative) generative capacity of TAG: There are no verbal elementary trees carrying the nominal arguments (and thus verifying the predicate-argument cooccurrence constraint) that can be combined to cover some of the (so called) scrambled word order of German (Becker et al. 1991, 92). Let us see where the problem is.

2.1. Argument Scrambling

Sentence (1) is the standard example for scrambling: The two constituents in the Mittelfeld (the positions between finite verb and non-finite verbs) have 'exchanged' their position. Note that this is a very natural order that even is the standard order if the direct and the indirect objects are pronominalized as in (1).

(1) a. Peter hat das Buch meinem Vater zu lesen versprochen.
Peter has the book to my father to read promised.
'Peter has promised to my father to read the book.'
b. Peter hat es ihm zu lesen versprochen
Peter has it him to read promised.
'Peter has promised him to read it.'

By definition, a control verb like *versprochen* 'promised' assigns a theta role to the subject just like the embedded verb. So the (semantic) predicate argument structure is shown in Figure 1a (leaving aside for the moment the tense auxiliary because of its uncertain semantic role). As the derivation structure of TAG is a tree, we have to restrict our analysis to one of the predicate-argument links, the subject role of the infinitive or the subject role of the control verb.

Suppose we wanted to follow the usual XTAG/FTAG analysis and have the control verb *versprochen* 'promised' govern its subject. Suppose further that we handled the auxiliary as an adjunction to the matrix verb. Our goal is then to obtain the derivation structure in Figure 1b. The first elementary tree of *versprochen* 'promised' (Figure 2) can adjoin to the root node of the infinitive and subsequently, its subject substitution node appears at the right place. However, in this case, its dative substitution node remains outside of the infinitive's elementary tree. We do not obtain the word order of sentence (1); the dative would be supposed to appear at the

* I would like to thank the XTAG group at the University of Pennsylvania and an anonymous reviewer for their helpful comments on my work. I am also grateful to Tilman Becker for his help finding the least ugly trees and keeping up my faith in TAG. I assume the customary responsibility for content and shortcomings.

 Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6), pp. 242-251. Universitá di Venezia.

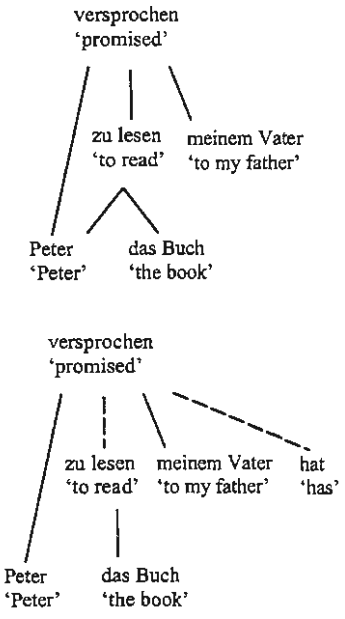

Figure 1: predicate-argument graph and derived tree

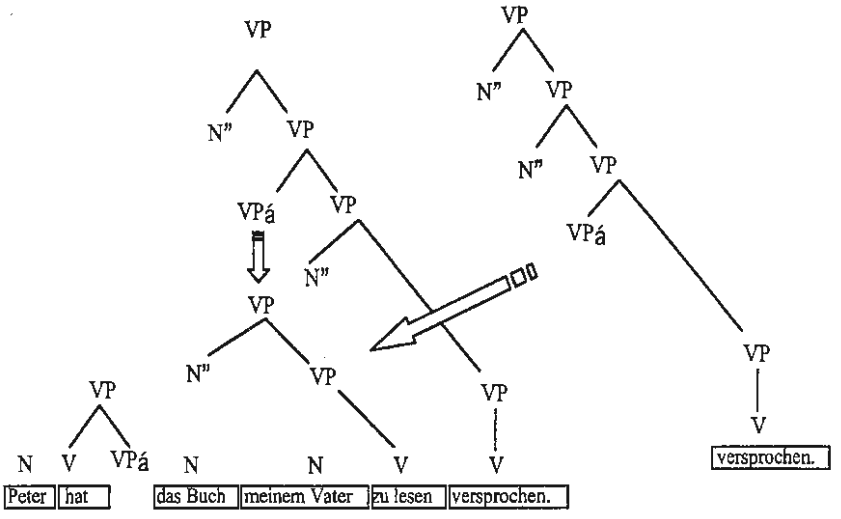

Figure 2: Scrambling is beyond TAG

right of the infinitive giving the so-called intraposition structure of German: *Peter hat das Buch zu lesen meinem Vater versprochen.* (Peter has the book to read to my father promised). The second elementary tree for *versprochen* 'promised' shown in Figure 2, allows on the contrary the dative argument to appear in the right place. Then, however, the subject node appears between the infinitive and its argument and the subject cannot be substituted at its Vorfeld (sentence initial) position.

The reader easily verifies that it actually does not make a difference if we decided to treat *versprochen* 'promised' as a raising verb, i.e. leaving the subject to the embedded verb *zu lesen* 'to read'. In this case we could not find a way to adjoin the auxiliary *hat* 'has' to its head *versprochen* 'promised'.[1] The only thing that would help for this particular case is to have the auxiliary carry the subject. This possibility could be a good choice if we wanted to develop a useful parsing grammar, as the case of a subject not neighboring the V2 position is very rare (mainly for negative pronouns like *niemand* 'nobody' or in spoken language with a strong accentuation). It would however weaken further the semantic interpretability of the derivation tree and does not solve the theoretical problem of the 'free'[2] placement of the arguments of the left and right bracket's verbs.

2.2. Adjunct Scrambling

It is less known that even with all the arguments in standard order, the position of the modifiers can block a correct TAG analysis. The preferred interpretation of sentence (2) has a prepositional phrase modifying the control verb *versprochen* 'promised', although this modifier finds itself between the infinitive *zu lesen* 'to read' and the infinitive's argument *das Buch* 'the book'

(2) Peter hat meinem Vater das Buch ohne Zögern zu lesen versprochen.
Peter has to my father the book without hesitation to read promised.
'Peter did not hesitate to promise to my father to read the book.'

In Fig. 3 we depict the corresponding elementary trees. The elementary tree of the adverbial modifier *ohne Zögern* 'without hesitation', cannot reach any node of its governor's elementary tree. The only possible TAG analysis will give us the semantically 'wrong' derivation tree where the reading is done with hesitation.

1. If the auxiliary does not adjoin to the past participle's tree, the information that a past participle adjoined would have to be passed through the infinitive's tree. This puts lexical information into the features and breaks all usual TAG principles.
2. Of course this order is only free in the sense that it does not affect the predicate argument structure of the sentence. The order depends on the communicative structure of the sentence.

2.3. Relatively Difficult

German relative clauses put up two hurdles for TAGs: The inner structure of the relative clause and its placement in the main clause.

The standard TAG formalism excludes trees with more than one foot node (as this would complicate considerably the derivation structure). Thus, we cannot express the idea that an adjunct turns its governor into an adjunct itself. However, this is the case for relativized adverbial modifiers. Take (3) as an example.

(3) der Balkon auf dem er singt.
the balcony on which he sings
'the balcony he sings on'

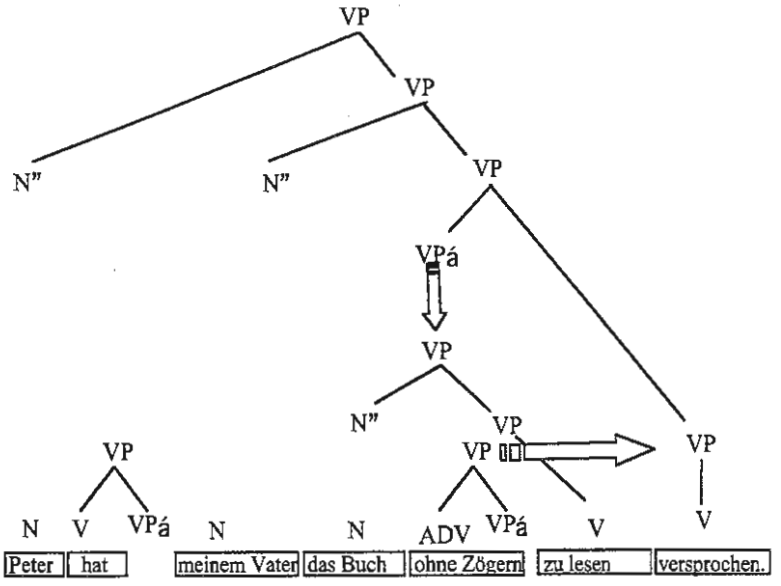

Figure 3: Adjunct scrambling is even worse

The only possibility of analyzing such a phrase in TAG is to combine the elementary tree for *singt* 'sings' and the one for the relativized modifier into one elementary tree of *auf etwas singen* 'sing on something', as if the balcony was an argument of the singing. This clearly violates the principle of non-compositionality.

In general, as soon as we want to move an element out of a substituted position, we are beyond the (derivative) generative capacity of TAG. Take example (4) of a German relative clause.

(4) der Film von dem ich nur die Musik kenne.
the movie of which I only the music know
'the movie I only know the music of.'

The predicate argument structure seems clear: *Musik* 'music' is the object of *kenne* 'know', and *von dem Film* 'of the movie' modifies *Musik*. The only way out for this case is to have separate *kenne* 'know' trees for the relativized modification of its arguments. Again, we do not obey the principle of non-compositionality and obtain severely messed-up derivation trees.

Extraposed relative clauses constitute another difficulty for the TAG formalism: The preferred placement for relative clauses in German is not directly behind the modified noun, but in the Nachfeld of the sentence's main domain. This preference is independent of the noun's position in the Vor- or Mittelfeld[3] and it is particularly strong in two circumstances: First, when the right bracket is occupied only by a short and semantically weak element, like for example a so-called separable verbal prefix. Secondly, when the relative clause is rather long. In the following example (5), the separable prefix *ab* of the verb *abschließen* 'lock' has to take the right bracket of the main domain.[4] The Nachfeld position of the relative clause in (5) is clearly preferable.

(5) a.?? Peter schließt die Tür, die ich gestern bemalt habe, ab.
Peter closes the door, that I yesterday painted have, up.
b. Peter schließt die Tür ab, die ich gestern bemalt habe.
Peter closes the door up, that I yesterday painted have.
'Peter locks the door I painted yesterday.'

3. The agreement in number and gender of the relative pronoun with its head allows the reconstitution of the dependency relation. The placement behind the noun might be preferable in the rare case when neither agreement nor semantic plausibility allows associating the relative clause with its governor. Another rarity consists of two nouns of the same domain being modified by relative clauses. In this case again, one of the relative clauses has to be adjacent to its head.

4. The separable prefixes of some verbs behave syntactically just like bare infinitival arguments of the verb: They usually take the syntactic place the verbal dependant would take, and the more semantic weight they have, the easier is their independent placement into the Vorfeld. For a more detailed discussion see Gerdes & Kahane 2001.

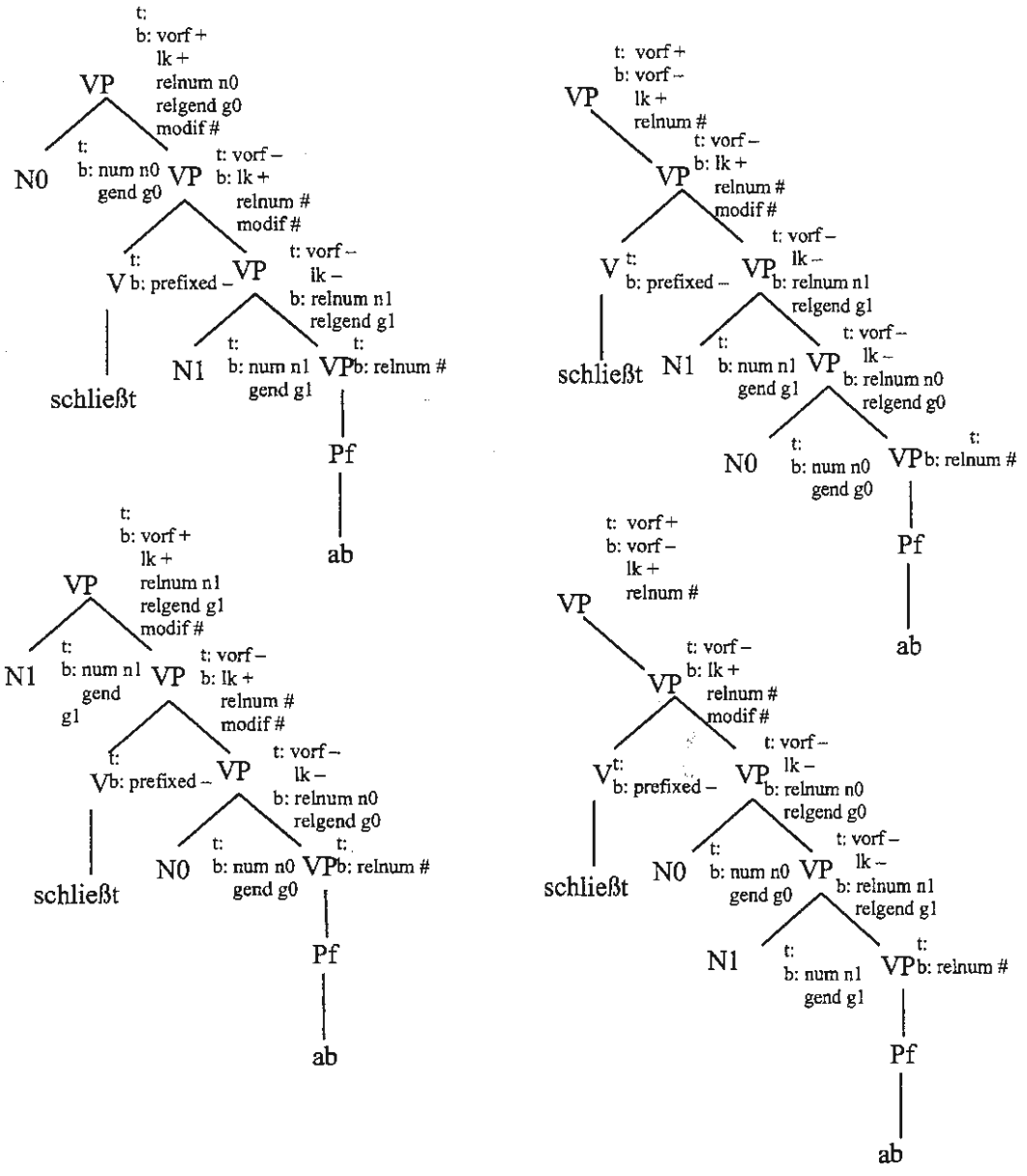

Figure 4: The simplified V2 trees for a transitive verb with separable prefix

This is another instance of an extraction of an element out of a substituted position. For the sake of the semantic consistency principle we have to realize the verbal prefix as a co-anchor of the verbal tree[5] and the verbal branch down to the prefix in the right bracket disallows the relative's adjunction to its head.

3. XTAG – FTAG – DTAG?

The question we have to address now is whether all this is any worse than the anglo-roman situation. English and French don't have scrambling, but it is clear that there are many cases in these languages, too, where the only possible analyses with a regular TAG will give us derivation trees with incorrect dependencies with our TAG principles falling by the wayside (Shieber & Schabes 1994, Rambow 1995, Candito & Kahane 1998). For example, the case of relative clauses is in fact very similar in the three languages in question: The gloss I give for sentence (3) is a correct English sentence (although the stranded preposition is preferable), parallel constructions to (3) and (4) exist in French (e.g. *Le film dont je ne connais que la musique* 'The movie I only know the

5. One sees easily that the situation is in fact identical even if we treated the verbal prefix as an argument or had it adjoin into the main verb's elementary tree.

music of'), and French has extraposed relatives just like English (in particular for parallel constructions to the separable prefix: *She threw the book away that I wanted to read.*)[6]

Several extension of the TAG formalism have been proposed attempting to remedy these shortcomings (VTAG, DTG, TDG, ...). The scrambling case of sentence (1), for example, could still be handled in tree local multi-component TAG, a little add-on to regular TAGs that does not change the weak generative power. However, if we added a 4th verb to the sentence, this will no longer be possible.

3.1. A generation grammar!

Leaving aside the theoretical and linguistic problems of these new formalisms, including multi-component TAG, their biggest flaw is the lack of implemented tools, a fact that creates the desire among researchers in language engineering for a 'real' German TAG. In particular the language generation task can live without the complete set of word orders. One could even go as far as saying that the limited word order possibilities are a feature not a bug: A formalism that allows all the possible orders would force the generation module to choose between the different orders. These choices are dependent mainly on the information structure of the sentence, a problem far beyond today's running generation systems. The snag is of course that the choice implied by the formalism is not necessarily the best choice in a given context. It has been shown though that a sentence in the so-called 'standard word order' can obtain most information structures by prosodic means (Lenerz 1977, Choi 99).[7]

For our German TAG, we could thus be content with the standard word order. However, even for a restricted generation task, it might be preferable in some cases to have access to verbal trees allowing different word orders. The difficulty of this work lies in the fact that one has to find a compromise between restrictions of the formalism (and the metagrammar description) and the concrete necessities for generation. We try to obtain a German TAG linguistically as good as possible and practically as useful as possible for the generation job. On the basis of corpora for quite simple generation tasks in German,[8] we decided to add all verb internal argument permutations to our grammar (in order, for example, to encode a lexically triggered preference of some verbs to place the accusative argument before the dative, or a preference to place temporal or local modifiers into the Vorfeld). We disallow inter-verbal argument exchange with two exceptions too common to leave aside even for a simple generation system: Topicalization into the Vorfeld and extraposition of relative clauses into the Nachfeld. Further we allow placement of embedded non-finite verbs only in the right bracket of its governor's domain and its extraposition into the Nachfeld.[9] We believe not to need any verbal positions like (partial) VP fronting or intraposition. The grammar ends up with many more trees than the standard word order, but we suppose that when these trees are inadequate, the tree description[10] allows sorting them out with no trouble.[11]

Of course such a grammar will be of limited use for parsing because in a corpus, we will encounter the other word orders as well. It seems that the phenomena that cause trouble to TAG are more frequent in German than in Anglo-Roman, where the 'pure' TAG grammars can obtain reasonable parsing results.[12]

4. The compromising trees

The trees presented here are product of a compromise between a minimal violation of the TAG principles (this comes down to a maximal semantic interpretability of the derivation tree), a maximal coverage of the grammar, a maximal usefulness for simple language generation systems, and a maximal simplicity in the meta-

---

6. Another awkward example on relatives is the fact that in the phrase *apples that Mary thinks John likes*, the tree of thinks has to adjoin into the likes tree, whereas in *the woman who thinks John likes apples*, the tree for thinks has an argument position (a substitution node) taken by the like tree. This XTAG analysis is triggered by the restrictions of the formalism and, to my knowledge, not by linguistic intuition.
7. This is in fact one of the definitions of 'standard word order'. The 'standard' order of a verb's nominal arguments can vary lexically (Müller 1999).
8. We actually looked at the requirements for the German generation systems in the MultiMétéo project (Coch 1998) and in the SmartKom project (http://smartkom.dfki.de).
9. Some matrix verbs like *scheinen* 'seems' prefer to construct incoherently, i.e. their verbal argument should go into the Nachfeld of the main domain.
10. Tree descriptions are finer grained than the notion of family and this should allow a smoother interface with the lexicon, the neglected but essential component of any grammar. Future work will show whether this approach resists when leaving the toy state.
11. This will have to be verified when actually using the grammar. The metagrammar setup easily allows generating limited subgrammars if desired.
12. It will be difficult to actually prove this point: One would have to compare an English and a German grammar with equivalent lexicons, equivalent grammatical coverage, and equivalent 'tricking' around the limits and principles of the formalism.

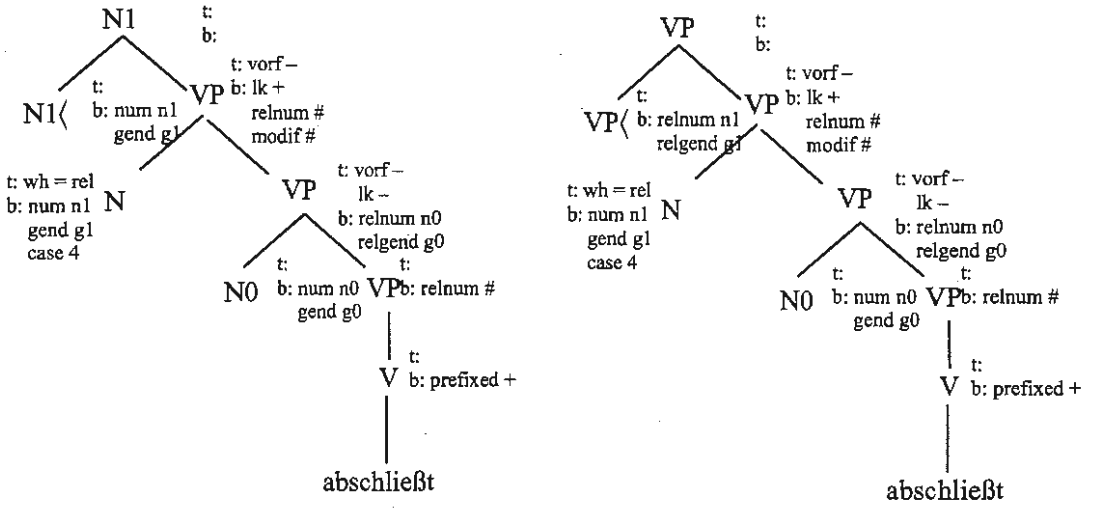

Figure 5: a relative elementary tree for adjacent modification (left), the other for long-distance modification (right)

grammar description. We do not expect syntactic expressiveness from the derived tree though.[13] The trees in Figure 4 illustrate some intermediate results on the way to this compromise.

We do not distinguish VP from S nodes. This simplifies the tree description in the metagrammar (see below) and additionally, it makes it easier to describe relative extraposition from the Vor- to the Nachfeld. Since the long-distance modification by the extraposed relative clause is beyond the generative capacity of TAGs fulfilling the strong cooccurrence constraint, we relax this constraint and push the lexical information of the noun's number and gender into the VP spine of the verbal elementary tree. Now, an adjoining relative phrase can check its agreement on the VP node. Of course, the corresponding derivation tree bares a relative clause modifying a verb and not a noun, and the generation module has to take this into account.[14] See Figure 5 as an example of an extraposed relative phrase and its corresponding elementary tree: The number and gender information of the argument is passed up into the *relnum* and *relgen* features of the verbal spine, where it can be read by the tree of the relative clause. The relative clause then passes it back down to its substitution node for the relative pronoun. Case is assigned directly by the verb.

So the treatment of extraposed relative phrases is one reason for the right branching structure we use. The other reason is that in a flat structure, the TAG formalism does not allow adjunction between sister nodes. As modifiers can appear at any position in the Mittelfeld, we need the VP nodes as landing points for the right adjoining modifiers we use.[15] A third reason is that the right branching tree is easily described in the metagra m-mar, as we will show below, as every nominal argument introduces its verbal spine element into the tree sketch, independently on the final order or realization. We do not stipulate however, that the VP-headed subtrees correspond to linguistic (functional, prosodic, semantic...) objects of their own; their justification is internal to the formalism.

The distribution of the German verb distinguishes different positions in what is usually called 'the topological model' (Drach 37, Bech 55): The finite verb can take the 2nd position of the sentence (V2, for main clauses[16]) or the final position (Vfin, for sentential complements and relative clauses). V2 means that one constituent of any nature has to be placed before the verb. This easily stated constraint is again beyond TAG's expressiveness; the translation into TAG becomes a highly non-trivial exercise of feature manipulation. The basic idea is the

13. Many TAG grammar writers seem to put the derived structure in second place. The end of establishing a correspondence between the word string and the semantically interpretable derivation tree justifies all syntactic means.

14. This boils down to some preprocessing of the derivation tree before the actual TAG generation comes in. The original derivation tree for an extraposed relative originally contains a verb whose nominal argument is again modified by a verb (the verb that opens the relative clause). In order to establish agreement, the preprocessing then not only has to move up the node of the relativized verb from nominal to verbal modification, it also has to compute the address of the node in the verbal matrix tree that will receive the relative adjunction. If this preprocessing comes out to be too costly, one could also put unique features in the tree to make sure that the tree of the relative clause only adjoins to the corresponding node in the verbal spine. We consider further that this preprocessor takes care of the uniqueness of the relative extraposition. An alternative choice would be to put a feature into the verbal spine that rules out double extraposition of relatives.

15. We have to avoid this modifying adjunction from the left into the root VP and into the VP of the verb in second position (V2).This is done with the non-unifying feature value # of the feature *modif.*

16. There are some cases of embedded V2, not taken care of in this grammar.

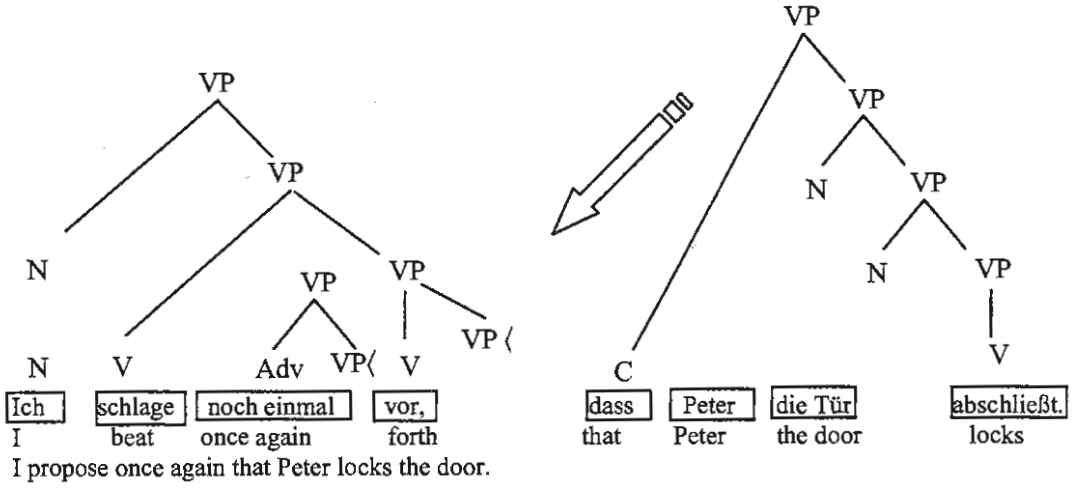

Figure 6: An analysis with sentential complement

following: The Vorfeld can be filled by any complement of the verb, argument or modifier. In both cases, the concerned verb has a VP node in its elementary tree. If this node dominates the Vorfeld-argument it has to prevent the adjunction of modifiers, since there is only one constituent before the finite verb (see the left trees of Figure 4). The second possibility is that the VP remains a unary node, which then forces the unique adjunction of modifier (see the right trees of Figure 4).[17]

Relative clauses and sentential complements behave similarly: The complementizer or the relative pronoun takes the first place, topologically often called complementizer field. This field behaves similarly to the left verbal bracket, and there have been attempts to consider the complementizer and the finite verb in V2 to be appearing in the same position (see for example Kathol 1995, for a contrary view in the light of universal grammar see Rambow & Santorini 1995). For TAG, the semantically empty complementizers should appear either with the matrix verb or with the embedded verb. We consider the complementizers field as unification of Vorfeld and left bracket, as in inherits properties of both fields (Gerdes & Kahane 2001). In the metagrammar, this is easily translated as the union of the two quasi nodes Vorfeld and left bracket. The 'universal' interpretation of a grammar could then consist of saying that in languages where the complementizer and the verb in V2 position do not have a complementary distribution, this node union did simply not take place (see Rambow & Santorini 1995 for the example of Yiddish).

As it is shown in Figure 6, we consider for the moment the sentential complement to substitute into the matrix verb. This allows handling all finite verb final constructions to be described quite uniformly in the metagrammar, but has the important drawback not to be able to treat wh-extraction into the Vorfeld. Contrarily to infinitival construction where every argument can join the Vorfeld, only wh-elements can be extracted out of sentential complements (see sentences (6)). However, this extraction is not a frequent phenomenon, and many native speakers of German consider it as bad style, preferring parenthetical constructions (which are not part of my Grammar for the moment either).

(6) a. ? Was schlägst du vor, dass Peter abschließt?
What beat you forth, that Peter locks?
b. Was, schlägst du vor, schließt Peter ab?
What, beat you forth, closes Peter up?
What do you propose that Peter locks?

Non-finite verbs come in many flavors: Keeping up the XTAG/FTAG division of subject-carrying control verbs on one hand and subjectless auxiliaries and raising verbs on the other, we are obliged to distinguish infinitival trees carrying a nominative argument or not. The infinitive can further care for the Vorfeld position or leave this field for other verbs. Caring for the Vorfeld position means to either place an argument there or forcing a Vor-

17. The expletive *es* stands out in this description because it is neither a modifier nor an argument and it can only appear in the Vorfeld. Its function is to occupy the Vorfeld if the speaker wants to avoid emphasize (topicalization or focalization, see Choi 99) of any semantically full element. The resulting sentence is purely rhematic. Formally the tree sketch corresponding to the expletive *es* is just the same as an adverbial one with the restriction that it can only adjoin into the Vorfeld. However, giving the *es* an individual entry seems at odds with the semantic consistency principle, as the function of this word is precisely to take a topological position without adding semantic information.

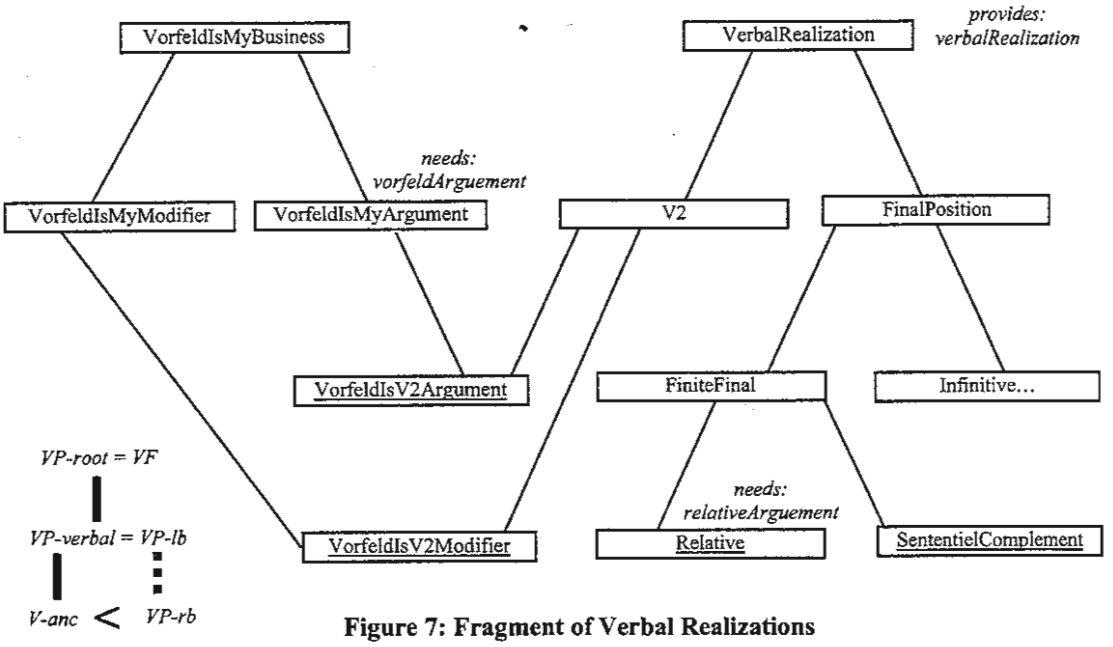

Figure 7: Fragment of Verbal Realizations

feld adjunction to its highest node. We leave aside all other crossing of arguments with other verbs. Even now, without cases of ellipsis we obtain 6 different trees for a simple transitive bare infinitive (or past participle or zu-infinitive that behave identically for the matter[18]).

5. Meta!

The freer word order of German arguments corresponds well to the idea of underspecified tree descriptions in the metagrammar: If we don't indicate the linear ordering between two elements, all the possible orders will be realized in the resulting tree sketches. On the other hand, the metagrammar becomes the actual linguistic description, and the tree sets obtain the status of intermediate products in the parsing process. For example, having two elementary trees for exactly the same word forms and the same sub categorization frame as in Figure 4, makes these cases indistinguishable on the tree level from actually different syntactic instances.[19]

5.1. Implementations

Our first attempt was to model the German verbal grammar around Candito 99's French metagrammar in order to be able to use her compiler. The main difficulty, except for the fact that this tool is no longer maintained, lies in the rigid definitions of the three dimensions she uses: initial subcategorization, redistribution, and realization of arguments. In order to capture the difference between V2 and Vfin positions, which is a general choice for all verbs (independently of their arguments), we have to introduce a realization module not only for arguments, but also for the head, the verb itself.[20] Although it was possible to hack Candito's tool to accept the verbal realization, the tree description for the V2 structure revealed a serious problem in this implementation:

At most one of the arguments of the verb has to be fronted, if none of its argument goes into the Vorfeld, it has to be a modifier, and the highest VP node has to force exactly one adjunction from the left. So we need leaf nodes whose function remains underspecified until the final class crossing and that eventually disappear if e.g. they can unify with their mother node. In Candito 99's implementation, those leaf nodes default to substitution nodes, which prevents them from unifying with their governor. In a word: It was not possible to express this linguistically appealing description of how the V2 elementary tree sketches came into being.

18. The infinitive extraposition to the Nachfeld, only possible for zu-infinitives, does not appear in the infinitival tree itself. It is the matrix verb subcategorizing for a zu-infinitive that has one more class crossing in (which can mean many more elementary trees depending on the arguments of the matrix verb).

19. This situation differs from French and English where, leaving aside some exceptions, different trees correspond to different syntactic constructions.

20. The possibilities described in this dimension correspond to the GB notion of head movement. We consider instead that the head is positioned out of an unordered dependency tree directly into its final position. Nevertheless, our V2-elementary trees without infinitival complement have an empty node at the verb-final position, which can be interpreted as the trace of the verb moved to the V2 position, or topologically (and preferably) as the right bracket that remained empty.

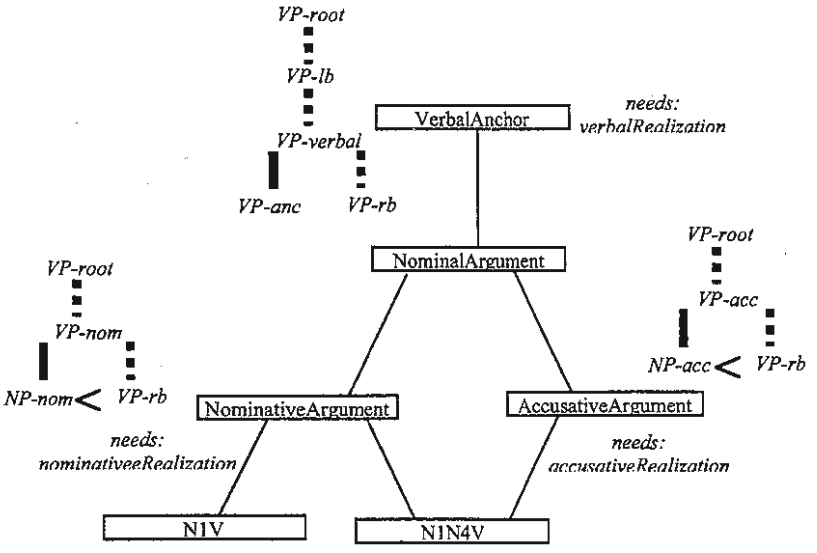

Figure 8: Fragment of the initial subcategorization

Help came from a reimplementation of the metagrammar compiler by B. Gaiffe (Gaiffe et al. 2002). This compiler allows without difficulty to obtain the desired minimal tree for the description[21]. The second a dvantage of the new compiler is that it is based on a more general approach than Candito's three dimensions: Each final class has a set of polar features, called needs and provisions, and all classes that can 'do' something for each other, i.e. mutually neutralize a polar feature, are crossed. Only when all needs and provisions are neutralized the minimal tree is calculated. This resolves the unclear status of the third dimension in the original approach that consisted of a number of quite independent hierarchies for each argument to realize. Moreover, for German at least two of the 'classical' redistribution for French, the passive and the causative in German, don't have a specific behavior different from verbal subcategorization. There is no syntactic reason to consider the passive auxiliary and the past participle as a unit, although they are, certainly, on a semantic level. If we see the metagrammar only as a generator of all necessary tree sketches, we have got already the necessary tree sketch among the trees for predicative adjectives.[22] A third advantage is the notion of tree description that replaces the notion of family: A feature structure, as fine grained as needed, replaces the name of the family. For details see (Gaiffe et al. 2002).

We conclude this section with a few extracts out of the complete German verbal metagrammar. We illustrate how the information on elementary tree sketches is distributed across the hierarchy.

In Figure 7 we depict a fragment of the verbal realization class hierarchy, indicating some partial tree description we use. Final classes are underlined. Each final class inherits the 'verbalRealization' provision and will thus be crossed with classes that need this feature. We just observe the combined tree description that the class VorfeldIsV2Modifier has inherited from all its superclasses: The quasi-node VP-root combines with the Vorfeld and it governs directly the combination of the VP-verbal node (the quasi node governing directly the verbal anchor) and the VP-lb (left bracket) node. Thus the desired V2 position is created. Moreover, we make sure that the right bracket (VP-rb) follows the left bracket. This partial tree description will cross with other classes that need or provide any feature for it. This crossing is completely incremental; no constraint can be erased.

Another subgrammar of the (not necessarily connected) metagrammar is the verbs' initial subcategorization frame (corresponding to Candito's first dimension). We observe in Figure 8 which constraints each class adds to the hierarchy: The verbal anchor class introduces 5 quasi nodes and their specific constraints. Each final verbal class inherits from this class and inherits thus these quasi nodes and their constraints. The final class N1N4V[23] for simple transitive nouns, for example, inherits the verbal anchor's tree description, the nominal argument's tree description, and the accusative argument's tree description. This means that the class's tree description has already the quasi nodes for a verbal anchor, a nominative argument, and an accusative argument, each with their

21. The calculation of the minimal representative of the tree description differs slightly from the original one by Rogers & Vijay-Shanker 1994.

22. The (rare) passives of intransitives may be a better reason to introduce the redistribution step, because these forms need a tree sketch without subject. In any case, if, for a given application, the notion of family containing all the passive trees as it is used in XTAG/FTAG proves to be useful, this 'redistribution step' can easily be added to the grammar.

23 These final class names recall the family names in the classical elementary tree arrangement. To avoid the discussion on whether the term subject corresponds to nominative (and indirect object to dative, etc) we name the arguments directly by their case (1 = nominative, 2 = genitive, 3 = dative, 4 = accusative).

corresponding VP-spine node. Their mutual order however will only be defined when the class's three needs will be satisfied, i.e. when the verb and the two arguments will get realized.

This should suffice to make clear the underlying principle of class inheritance and class crossing. The metagrammar is actually being worked on with the goal to include as many grammatical phenomena as possible. For real use, we will then choose the necessary grammatical phenomena and only generate the required trees.

6. Conclusion

We have seen the possibilities and limits of describing German with Tree Adjoining Grammars. As the formalism is very restricted (and for that computationally attractive), the design of the elementary trees is a constant trade-off between usefulness and descriptiveness. If however, we want a grammar for a well-restricted area, we can nearly always find a way of covering the desired string with a more or less good-looking tree.

One experience that should be shared with TAG writers of other new languages: It seems very difficult to start with the metagrammar as a descriptive tool, and see afterwards what trees fall out. It is much easier to decide first on the concrete tree sketches and then to distribute the information contained in the tree into a hierarchy, in a way to maximize sharing of information. The resulting metagrammar can then be seen as an interesting linguistic description, although its TAG-independent usefulness remains to be shown.

It was thought that the metagrammar would liberate the grammar writer from the actual messy mass of elementary trees, that one could easily extend a metagrammar by just adding some classes that will automatically be crossed with the existing classes. In reality, the grammar writer has to check all new resulting tree sketches, and most of the time lots of features have to be altered all around the existing metagrammar to fit in smoothly the new classes. The metagrammar should maybe better be thought of as a specific elementary tree editor. It is true however that it is easy to create limited subgrammars by simply commenting out the undesired classes in a more complex grammar.

In this paper we cannot go into greater detail on the proposed structure of a German metagrammar. We could only outline some of the problems we encountered on the way to a German TAG, problems which seem to us to be instructive for the general understanding of TAGs and of the idea of a metagrammar.

References

Abeillé Anne, 1991. *Une LTAG pour le français*. Ph.D. thesis. Univ. Paris 7.

Bech G. 1955. *Studien über das deutsche Verbum infinitum*, 2nd edition 1983, Linguistische Arbeiten, Nr. 139, Niemeyer, Tübingen.

Becker, T., A.K. Joshi, O. Rambow 1991. "Long distance scrambling and tree adjoining grammars" in *Proceedings of ACL-Europe*.

Becker, T., M. Niv, O. Rambow 1992. *The Derivational Generative Power of Formal Systems or Scrambling is Beyond LCFRS*. IRCS Report 92-38, IRCS, University of Pennsylvania.

Candito M.-H., Kahane S. 1998. 'Can the TAG derivation tree represent a semantic graph? An answer in the light of Meaning-Text Theory.' *Proc. Fourth Workshop on Tree-Adjoining Grammars and Related Frameworks (TAG+4)*.

Candito, M.H., 1999. *Structuration d'une grammaire LTAG : Application au français et à l'italien*. Ph.D. thesis, University of Paris 7.

Choi, H-W. 1999.*Optimizing Structure in Context – Scrambling and Information Structure*, CSLI, Stanford.

Coch, J., 1998. "MultiMeteo: Interactive weather report generation" in: *Proceedings of the 9th International Natural Language Generation Workshop (INLG'98)* Niagara-on-the-Lake, Canada.

Drach, E. 1937. *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt.

Gaiffe B., B Crabbé, A. Roussanaly 2002. "A New Metagrammar compiler" in *Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6)*. Universitá di Venezia.

Gerdes, K., S. Kahane 2001. "Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy", in *Proceedings ACL 2001*, Toulouse.

Kathol, A. 1995. *Linearization-based German Syntax*, PhD thesis, Ohio State University.

Lenerz J. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*, Tübingen, Narr.

Müller St. 1999, *Deutsche Syntax deklarativ: Head-Driven Phrase Structure Grammar für das Deutsche*, Linguistische Arbeiten 394; Niemeyer: Tübingen.

Rambow, O. 1994. *Formal and Computational Aspects of Natural Language Syntax*, Institute For Research in Cognitive Science, PhD thesis, University of Pennsylvania, Philadelphia.

Rambow, O., K. Vijay-Shanker, D. Weir 1995. "D-Tree Grammars", in *Proceedings of ACL'95*.

Rambow, O, B. Santorini 1995 "Incremental Phrase Structure Generation and a Universal Theory of V2" in *Proceedings of NELS 25*, Amherst, MA.

Rogers, J., Vijay-Shanker, K., "Obtaining Trees from Their Descriptions : An Application to Tree-Adjoining Grammars", *Computational Intelligence*, 10, 4, 1994.

Schabes Y., S. Shieber 1994. "An alternative conception of tree-adjoining derivation". in *Computational Linguistics*, 20(1).

The XTAG Research group 1995. *A lexicalized Tree Adjoining Grammar for English*. IRCS-95-03, Institute of Research in Cognitive Science, Philadelphia.