In Seuren, Pieter A.M.; and Kempen, Gerard, editors, *Verb Constructions in German and Dutch*

(2003), Benjamins, Amsterdam.

Coherent Constructions in German: Lexicon or Syntax?

Owen Rambow
Columbia University
rambow@cs.columbia.edu

1 Introduction

This paper addresses the issue of embedded *zu*-infinitival clauses in German from the perspective of a formal grammatical framework, tree adjoining grammar (TAG) and related tree rewriting systems.[1] Two radically different analyses have been proposed for this construction. According to the "syntactic" analysis, embedded infinitivals in German are analyzed essentially as in English, namely as clausal complementation. The "incorporation" analysis suggests that there is a process by which two verbs are combined into a single lexical unit (in some sense), which is head of a single syntactic projection. The incorporation analysis has become widely (in fact, nearly universally) accepted in one form or another in both the transformational and non-transformational literature, and has an undeniable intuitive appeal to the native-speaker linguist. However, this paper argues that there is empirical evidence against the incorporation analysis. Furthermore, methodological parsimony requires that the introduction of machinery to handle the merging of argument lists of two verbs (as required under the incorporation analysis) be motivated by the data, and that no alternate account (which does not rely on the additional mechanism) be available. Unfortunately, the status of much of the crucial data is quite murky. As a consequence, theoretically significant choices in the machinery of syntactic theories need to be made on the basis of difficult grammaticality judgments. This paper does not argue for a syntactic solution as such. Instead, it suggests that the stark contrast between the two analyses is in fact an artifact of the grammatical frameworks in which the construction has been analyzed. The paper proposes an analysis in a grammatical formalism in which all phrase structure is built incrementally in a formal derivation, and in which node labels are represented largely as features. In such a system, it is argued, the difference between the syntactic and the incorporation analyses can be interpreted as a difference in the ordering of steps in the derivation.

[1] I would like to thank Tony Kroch and Beatrice Santorini for helpful discussions.

This paper is structured as follows. In Section 2, I introduce TAG and lexicalized tree-rewriting systems. In Section 3, I summarize the relevant data and describe the incorporation analysis. I then discuss why this analysis is problematical for TAG. In Section 4, I present arguments against the clause-union analysis, and propose a solution using a tree-rewriting system in Section 5.

2 Tree Adjoining Grammar

Work in non-transformational frameworks including LFG (Kaplan and Bresnan, 1982), HPSG (Pollard and Sag, 1994), categorial systems such as CCG (Steedman, 1996; Steedman, 1997), and Tree Adjoining Grammar (TAG) (Joshi et al., 1975; Joshi, 1985), has stressed the importance of the lexicon in the development of a theory of syntax, and these approaches share a desire to locate syntactic variation within and between languages in the lexical component of grammar. In TAG-based linguistic theories[2] (Kroch, 1987; Frank, 1992; Frank, 2001), the lexicalist orientation of the linguistic theory is expressed in a *lexicalized* mathematical formalism (a tree-rewriting system): a Tree Adjoining Grammar is a set of elementary trees, each of which represents a single lexical item and its syntactic projection. The lexicon is thus a grammar in the formal sense defined by TAG.[3]

The elementary trees are combined using the operations of substitution and adjunction. In substitution, a frontier node is replaced by a tree; in adjunction, an interior node is replaced by a tree.

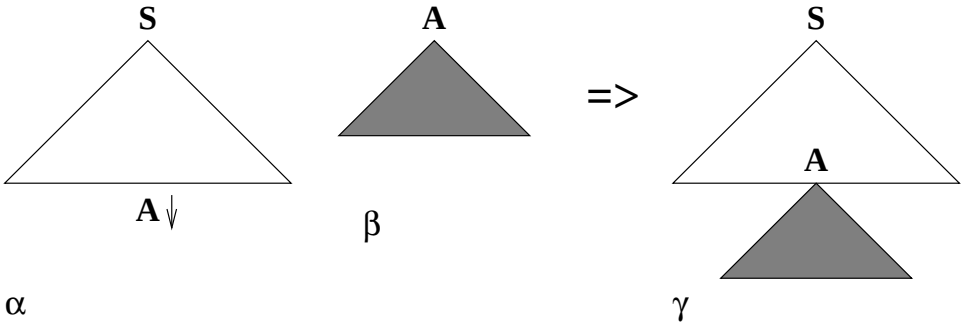

Figure 1: The Substitution Operation

Substitution is illustrated schematically in Figure 1. We can substitute tree β into tree α if there is a nonterminal symbol on the frontier of α which has the same label as the root node of β ('A' in Figure 1). We can then simply append β to α at that node. Nodes at which substitution is possible are called "substitution nodes" and are usually indicated with a downarrow ($\downarrow$). Adjunction is shown in Figure 2. Tree α contains a nonterminal node labeled A (not on its frontier); the root node of tree β (an "auxiliary

[2]Tree adjoining grammar, unlike HPSG or LFG, is simply a mathematical formalism. Hence the need to refer to "TAG-based linguistic theories" rather than simply "TAG".

[3]For general introductions to TAG, see (Joshi, 1987; Joshi, 1994; Abeillé and Rambow, 2000).

tree") is also labeled A, as is exactly one non-terminal node on its frontier (the "foot node"). All other frontier nodes are terminal nodes or substitution nodes. We take tree α and remove the subtree rooted at its node A, insert in its stead tree β, and then add at the footnode of β the subtree of α that we removed earlier. The result is tree γ.

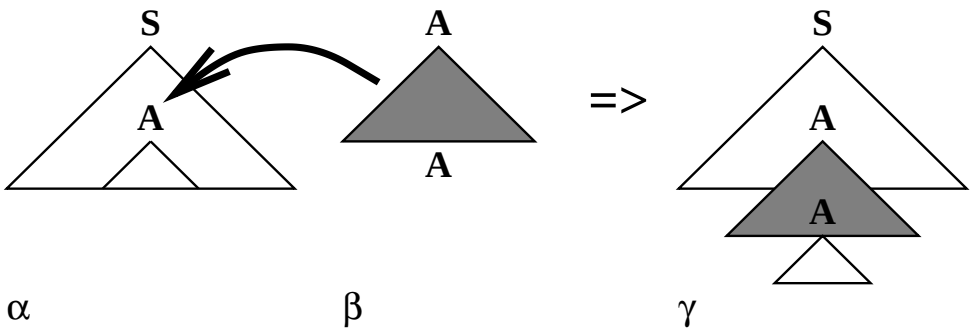

Figure 2: The Adjunction Operation

The interest of the TAG-based approach derives from the fact that the combination of syntactic structures is a formal operation. This has both formal and linguistic advantages. From a formal point of view, the restricted nature of the operations that derive larger structure means that the generative capacity of the formalism is restricted and that efficient (polynomial) parsers can be built. From a linguistic point of view, it means that syntactic theory is restricted in its scope to local domains. More precisely, syntactic theory must specify how to project structure from single lexical items, and how to derive variants (such as trees with *wh*-movement), be it by lexical rules (Abeillé, 1990), metarules (Becker, 2000; Evans et al., 2000), or by a principles-and-parameters type transformational theory based on move-α (Frank, 2001). I will refer to this phase as the lexical derivation (since it affects the syntax of a single lexical projection). Note that in TAG-based linguistic theories, the lexical derivation is not a derivation in the TAG formalism; rather, this phase produces a grammar in the TAG formalism. Structures from this grammar are then combined using the formal operations of TAG. There is no other way of combining two lexical items. For example, there can be no additional restrictions on movement; effects such as long-distance *wh*-movement and island constraints must fall out of the way in which elementary structures are defined in the lexical derivation, and the way in which the formalism combines them in the syntactic derivation.[4] I will refer to this phase as the syntactic derivation (since if combines the syntax of more than one lexeme). The syntactic derivation is restricted only by the definition of the elementary trees and the formal definitions of substitution and adjunction – the linguist using TAG as a metalanguage cannot formulate new or language-specific operations for the syntactic derivation (unlike HPSG or CCG (Steedman, 1991), in which the syntactic derivations are subject to cross-linguistic variation). As a result, if a linguist wants to develop a principles-and-parameters[5] theory, its scope is limited to the lexical derivation. The methodology is illustrated in Figure 3.

[4] See (Kroch, 1987; Kroch, 1989; Frank, 2001) for discussions of *wh*-movement in TAG.

[5] The term refers to a type of linguistic theory, not a particular one.

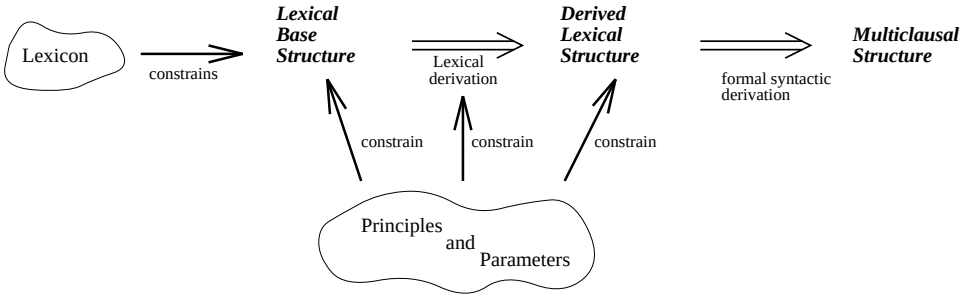

Figure 3: The "basic TAG approach": using TAG as a metalanguage for linguistics

3 Coherence and Clause Union: the Incorporation Approach

3.1 Coherent Constructions

Properties of German embedded infinitivals led Bech (1955) to descriptively identify two classes of constructions, the *coherent* constructions in which two or more verbs are adjacent and in which the verbs' arguments appear to behave as if they were arguments of a single verb, and the *incoherent* constructions, in which the verbs are not (necessarily) adjacent and which displays expected biclausal behavior. Purely descriptively, the following properties have been noted (this list is based on (von Stechow and Sternefeld, 1988), and is not exhaustive):[6]

- Sentences with extraposed clauses and sentences in which only one verb has been fronted are always incoherent constructions:

  (1) a. daß Hans versucht, das Auto zu reparieren
  that Hand tries the car to repair
  that Hans tries to repair the car

  b. Zu reparieren hat Hans das Auto versucht
  to repair has Hans the car tried
  Hans has tried to repair the car

- Fronting of multi-verb sequences without their arguments is possible only in coherent constructions:

  (2) Zu reparieren versucht hat Hans das Auto
  to repair tried has Hand the car
  Hans has tried to repair the car

[6] The terms "matrix" and "embedded" are used purely descriptively to refer to the two verbs involved and to their arguments. The use of this terminology does not imply a bias against the incorporation analysis.

- Long scrambling[7] is possible only in coherent constructions:[8]

  (3) a. daß es Hans zu reparieren versucht
  that it Hans to repair tries
  that Hans tries to repair it

  b. * daß es Hans repariert zu haben bereut
  that it Hans repair to have regrets
  Intended meaning: that Hans regrets having repaired it

  (3b) shows that long scrambling is restricted to certain verbs. I will call verbs that allow for coherent constructions with their complements *coherent* verbs, and those that do not, *incoherent* verbs. (3b) also shows that the juxtaposition of center-embedded infinitivals is not a sufficient condition for coherence. In fact, most authors assume that in the center-embedded construction with *zu* infinitives, coherence is optional even if the matrix verb is a coherent verb.

- Coherent verbs allow for a transposition of the matrix and embedded verbs (the "Third Construction"):

  (4) daß Hans das Auto versucht zu reparieren

- In incoherent constructions, negation in the embedded clause cannot have matrix interpretation, while in coherent constructions, it must.

  (5) a. Weil Hans versucht, das Auto nicht zu reparieren
  because Hans tries the car not to repair
  Only reading: Because Hans is trying not to repair the car

  b. Weil Hans das Auto nicht zu reparieren versucht
  Reading 1: Because Hans is trying not to repair the car
  Reading 2: Because Hans is not trying to repair the car

  Note that (5b) has two readings since the string is ambiguous between a coherent and an incoherent construction.

- In coherent constructions, embedded reflexives can take their antecedent among the matrix arguments:

[7] The term "long scrambling" is used to refer descriptively to cases in which an argument or adjunct of a verb appears to the left of an argument of the matrix verb, or to the left of the matrix verb if it precedes the embedded verb. One often finds the claim that German has no long scrambling. This claim is not an empirical claim, but a theoretical claim, since it is based on the clause union analysis: if coherent constructions (the only context in which long scrambling is licensed) are in fact monoclausal constructions, then of course there is no theoretical distinction between local and long scrambling. That scrambling is syntactically different from *wh*-movement in that *wh*-movement can escape from finite clauses while scrambling never can (a statement occasionally made to show that long scrambling does not exist) is both uncontroversial and irrelevant to the issue.

[8] I ignore differences between clitic climbing and scrambling for the sake of the argument in this paper.

(6) Hans$_i$ hat sich$_i$ zu rasieren versucht
Hans has himself to shave tried
Hans has tried to shave

- As pointed out by Höhle (1978), it is possible in coherent constructions to promote the embedded object to matrix subject through passivization. It is not possible to passivize only the matrix clause:

(7) Zu reparieren versucht wurde der Wagen/ *den Wagen schon
to repair tried was the$_{NOM}$ car the$_{ACC}$ already thrice
dreimal

It has been already tried three times to repair that car

Note that many speakers find the version with the nominative quite marginal as well.

Bech (1955) claims that only subject-control verbs without nominal objects can participate in a coherent construction. Haider (1993) shows that this is not true. For example, in (8) (Haider's 56-b), an object control verb is involved in a coherent construction (as evidenced by the long scrambling of the embedded object).[9]

(8) ? daß ihn mir jemand zu konsultieren geraten hat
that him$_{ACC}$ me$_{DAT}$ someone to consult recommended has
that someone has recommended to me to consult him

3.2 Clause Union: A Incorporation Analysis

The descriptive distinction between coherent and incoherent constructions is interpreted by Evers (1975) in the formal[10] context of transformational grammar. He proposes that sentences with recursively embedded clauses in Continental West Germanic languages differ significantly in their syntactic analysis from their non-center-embedded counterparts in the same languages and in other languages such as English and French. In constructions in in which verbs are adjacent to each other, he proposes a two-pronged process for coherent matrix verbs, which he calls "clause union":

1. The embedded verb moves up to its governing verb and forms a single morphological unit through incorporation ("verb cluster formation"). The argument lists of the two verbs are merged.

2. The process of verb raising dissolves the clause boundary of the embedded clause ("clause pruning").

[9] The somewhat reduced acceptability can be attributed to the multiple scrambling, and the sentence is clearly better than would be expected if *raten* were coherent.

[10] I leave aside the issue to what extent the transformational theory of the 70s was 'formal" – the term is used here to contrast Evers's work with the purely descriptive work of Bech.

This approach has been followed by many other researchers in other frameworks as well, though of course analyses differ greatly as they are expressed in the different frameworks. For example, Haider (1993), working in the Chomskyan transformational framework of the day, rejects the notion that a morphologically complex verb can be formed in the syntax, and instead proposes that the verbs provide a complex basis for single projection. I will refer to all analyses in which the two verbs are explicitly given properties of a single verb as the *incorporation analysis*. I will use the terms "incorporation analysis" and "clause union" interchangeably.

3.3 Tree Adjoining Grammar and Lexicalism

It should by now be clear that the operation of verb cluster formation is potentially troubling for TAG-based approaches: the strict separation between a grammar (= lexicon) on the one hand and formal rules for deriving larger multi-lexemic structures from the lexicon on the other hand would be blurred by a pure incorporation analysis of the coherent construction. More precisely, a incorporation analysis leads to two problems, one linguistic and one formal:

- Linguistically, the scope of syntactic theory is extended beyond the definition of projections from lexical items and of lexical rules to be performed on these projections, since now we need a new mechanism to combine structures for two initially independent lexical items. The principal linguistic advantage of TAG, that the scope of syntactic theory is restricted and that derivations involving multiple projections are handled by the formal machinery of TAG, is lost.

- Formally, a mechanism to include in the lexicon complex verbal forms of potentially unbounded size means that the lexicon (as a formal entity) is no longer finite. This means that we no longer have a TAG (since a TAG requires a finite grammar), and we lose the formal advantages (restricted generative capacity and efficient parsing).

Neither problem is necessarily fatal for TAG. Linguistically, it is probably uncontroversial that we want a productive lexical mechanism, for example to combine a preposition and a verb to create particle verbs (*weggucken, wegfahren, wegwischen* 'to look/drive/wipe away'), even if certain combinations are restricted and certain combinations give rise to idiosyncratic meanings. This mechanism could be extended in order to handle the manipulation of argument lists needed for the cases at hand. On the formal front, the grammar used for parsing could be "tailored" to the input, since the grammar is lexicalized. This means that only those recursive lexical rules need be applied which are justified by the input string, resulting in a finite *ad hoc*-grammar. Nonetheless, the clause union analysis is, basically, at odds with the spirit of TAG. It is therefore not surprising that Kroch and Santorini (1991), in providing a TAG analysis for German and Dutch embedded clauses, suggest that these constructions should not, in fact, be analyzed as forming a clause union.

4 Arguments Against Clause Union

While TAG-based linguistic frameworks may have a particularly acute interest in avoiding a incorporation analysis of the coherent construction, the incorporation analysis needs to be justified in *any* framework: given the existence of the incoherent constructions, it is clear that we need to have a *syntactic* mechanism for subordination of infinitival clauses in German in any case. Therefore, methodological parsimony requires that the introduction of additional machinery to handle the coherent constructions by lexical means be justified, and in particular, it needs to be firmly established that the syntactic machinery is not empirically adequate. It may be objected that there is ample independent evidence in German for a lexical process of incorporation, e.g. the aforementioned particle verbs. However, unlike embedded infinitivals, these processes are not recursive, and furthermore, they do not require the merging of argument lists from two originally independent lexical items. It may also be objected that the causative, which in many languages is a morphological affix, provides independent cross-linguistic evidence for the need for lexical manipulation of argument lists. However, the morphological causative, like the passive, is a manipulation of the argument structure linked to the morphology of a single lexeme, not to a lexical derivation involving *two* (open-class) lexemes. Thus, the apparatus of argument list merger is not independently motivated in syntax, and is unique to the proposed clause union analysis for Continental West Germanic.

In this section, I question the validity of some of the arguments for an independent lexical process of clause union formation by sketching an alternative account in a transformational framework, and discuss additional empirical evidence against it.

4.1 Is Clause Union Necessary?

I will now sketch accounts of the data presented in Section 3.1 in an informal transformational framework. These accounts do not rely on the notion of clause union and instead are purely syntactic in that lexical items retain their individual projections. Some of these accounts have been proposed and discussed previously in the literature (as noted) and can be characterized as an "evacuation" account: the syntactic phenomena are derived by long scrambling out of a constituent and an independent operation performed on the remnant constituent.

Suppose that the only difference between coherent and incoherent verbs were the fact that coherent verbs allowed scrambling out of their complement clause, but incoherent verbs did not. Such a proposal is motivated by the difference between bridge and non-bridge verbs in many languages including English and German, which differ with respect to allowing *wh*-extraction from their finite complement clause. This proposal is supported by the observation that the bridge verbs in German appear to be exactly the coherent verbs. In the case of scrambling, the difference between coherent and incoherent verbs could be represented as a difference in the category or feature content of the maximal projection of the selected clause, and

Suppose furthermore a general process of clausal extraposition (independently needed), and suppose that scrambling can also occur from extraposed clauses. Following the proposal of den Besten and Rutten (1989), we can derive the coincidence of the Third

Construction with coherent verbs by suggesting that it is the result of extraposition and subsequent long scrambling of the arguments out of the extraposed clause. Incoherent verbs do not allow long scrambling from the extraposed clause, and hence do not allow the third construction, as required. A similar approach (following Webelhuth and Besten (1987)) can predict that fronted multi-verb sequences are restricted to coherent verbs: we front the matrix VP after scrambling all arguments of both verbs out of it. Since for the embedded arguments, this represents long scrambling, the construction cannot be derived if the matrix verb is incoherent. Finally, if we assume simply that the negation marker *nicht* modifies the verb into whose projection it is adjoined, we can derive the ambiguity in coherent constructions by assuming that for the matrix negation reading, the embedded arguments have been scrambled to the left of the *nicht* adjoined into the matrix clause. Again, an incoherent verb would not allow this long scrambling. (For an alternative account of the negation facts that resorts neither to clause union nor to long scrambling in the overt syntax, see (Kroch and Santorini, 1991).) Finally, the binding facts can be straightforwardly explained if we assume that the embedded subject, whether represented in the phrase structure as PRO or not, is coreferential with the matrix argument and can serve as local antecedent. This analysis is in fact confirmed by the fact that incoherent verbs also allow this kind of binding: *Hans bereut, sich rasiert zu haben*. (We return to long passives in Section 5.4.)

4.2 Is Clause Union Sufficient?

The syntactic analyses proposed above make additional predictions: if the word orders which are taken to be indicative of the coherent construction are derived by long scrambling, what happens if we only scramble some elements and leave others *in situ*? First, consider the Third Construction. Since it is derived by scrambling arguments from an extraposed clause, we expect that we can leave arguments or adjuncts in the extraposed clause, while scrambling others out. This is indeed the case:

(9) a. (Bayer, 1992)

daß er die Schweine vergessen hat dreimal zu füttern
that he the pigs forgot has thrice to feed

that he forgot to feed the pigs three times

b. (Kroch and Santorini, 1991)

daß uns Hans versuchte, seinen Wagen zu zeigen
that us_{DAT} Hans tried his car to show

that Hans tried to show us his car

c. Uwe Johnson (German novelist), in a letter reproduced in his memoires, *Die Katze Erinnerung*:

…daß ich diesen Beruf nach Fähigkeit und Neigung glaube
that I this profession according-to ability and inclination think
am besten ausfüllen zu können
at best exercise to can

…that I think that it is this profession that my abilities and inclinations make me most suited for

d. Ludwig Tieck, *Franz Sternbalds Wanderungen*:

…daß sie sie glaubten mit Augen zu sehen und mit den
…that they$_{NOM}$ them$_{ACC}$ believed with eyes to see and with the
Händen zu erfassen
hands to grasp

…that they believed to see them with their eyes and to grasp them with their hands

These sentences contradict the characterization of coherent constructions. In these examples, embedded arguments or adjuncts appear between matrix arguments or adjuncts and the matrix verb, stranded from the embedded verb. If they have scrambled from the extraposed embedded clause, we must conclude we have a coherent construction. At the same time, the presence of embedded arguments or adjuncts between the two verbs means the examples cannot form a coherent construction.[11]

In the case of verb fronting, the syntactic analysis, but not the incorporation analysis, predicts that we should be able to front only the embedded verb, even if long scrambling takes place:[12]

(10) a. Zu repaRIEren hat das Auto der HANS versprochen
to repair has the car the Hans promised

It is Hans who has promised to rePAIR the car

Again, this sentence is problematic for the incorporation analysis, since the long scrambling forces a coherent analysis, while the fronting of a single verb precludes it.

Similarly, under the syntactic analysis, we expect to be able to front two verbs while leaving behind arguments of both, and while extraposing the embedded verb in the fronted position along with a remaining argument. Again, this is the case:

(11) (Netter, 1991):

Versucht, einen Freund vorzustellen, hat er ihr noch nie
tried a friend to-introduce has he her not-yet never

[11]It could be argued that elements such as *dreimal* are incorporated into the verbal complex. However, there is no independent motivation for such a proposal, and it clearly cannot account for the other examples, in which full phrases intervene between the two verbs, in the case of (9d) even including two conjoined verbs with adjuncts!

[12]These cases are not very good if the fronted verb and the matrix subject are not stressed (which we indicate through capitalization).

He has never tried to introduce a friend to her

The presence of arguments and adjuncts of both verbs in the Mittelfeld forces an analysis as a coherent construction, while the extraposed embedded verb (with argument) precludes it.

In the case of negation, we expect under the syntactic analysis to be able to have *nicht* with scope only over the embedded clause (by being adjoined to it) even when long scrambling takes place out of that clause. This prediction is borne out:

(12) Wieso redet Jutta so behutsam mit Karsten?
Why is Jutta speaking so carefully with Karsten?

? Weil ihn Jutta nicht zu beleidigen versucht
because him_{ACC} Jutta not to insult tries

Because Jutta is trying not to insult him

In the incorporation analysis, the long scrambling forces the coherent construction, which in turns predicts that only the matrix reading of the negation should be available.

Finally, consider a sentence that allows us to test for binding facts without interference from the control reference:[13]

(13) Zu rasieren erlaubt hat der Meister_i dem Lehrling_j nur $\text{ihn}_{i,*j}$/$\text{sich}_{j,*i}$ selber
to shave allowed has the master the apprentice only him/himself self

The master has allowed the apprentice to shave only him/himself

The possible readings are exactly as expected if the construction were incoherent, and are incompatible with a coherent construction under the assumption that the the verbs form a single domain for anaphor binding, as would be natural in the clause union analysis. But under the incorporation analysis, the verb sequence can only be fronted without arguments if the construction is coherent.

In conclusion, we see that there is empirical evidence against the incorporation analysis. At the same time, much of the relevant evidence is somewhat degraded, and in corpora it is rare.

5 Coherence and Tree Rewriting: The Syntactic Approach

5.1 A Formalism for (Relatively) Free Word Order

In (14), the embedded verb, *geben* 'to give', has two overt nominal arguments, one of which has topicalized into sentence-initial position, and the other of which has scrambled beyond the matrix subject.

[13]This sentence should be compared to that in (Haider, 1993, footnote 20). The contrast in judgments suggests that other factors are at work in Haider's example, given the relative clarity of the judgments here.

(14) [Dieses Buch]$_i$ hat [den Kindern]$_j$ niemand [PRO t$_j$ t$_i$ zu geben]
[this book]$_{\text{ACC}}$ has [the children]$_{\text{DAT}}$ [no-one]$_{\text{NOM}}$ to give
versucht
tried

No-one has tried to give this book to the children

If we associate the nominal arguments of each verb with their verb, and the matrix auxiliary with the matrix verb, we get the following pattern:

(15) Dieses Buch hat den Kindern niemand zu geben versucht
n^{emb} v^{matrix} n^{emb} n^{matrix} v^{emb} v^{matrix}

This sentence cannot be derived by any TAG, if we assume that nominal arguments and auxiliaries are substituted/adjoined into the tree of their main verb. This is because upon adjunction, the frontier of one tree is divided into three segments (as can be seen in Figure 2), and the frontier of the other into two, giving us five segments, but in the string above we have six segments! Simple TAG is therefore not sufficiently powerful to handle the word order variation found in German.

If TAG is not adequate for the description of syntax, what options do we have? The relatively free word order found in German suggests that what is needed is not a formalism whose elementary structures are trees, but one in which the elementary structures are underspecified trees, i.e., *descriptions* of trees. Such a formalism is D-Tree Substitution Grammar (DSG) (Rambow et al., 2001). In lexicalized TAG, elementary objects of a grammar are completely specified trees which fix the syntactic context for the lexical anchor. In DSG, the elementary objects are descriptions of trees ("d-trees"), which represent possible syntactic contexts for the anchor. D-trees are formalized in a logic for describing nodes and the relationships that hold between them, namely dominance, immediate dominance, and linear precedence.[14] A relation of underspecified dominance between two nodes is called a *d-edge*. An example is shown in Figure 4.[15] Immediate dominance is represented as always with a solid line, while d-edges are represented with a dashed line. D-edges may be annotated with *path constraints*, which restrict the sets of node labels (or features of node labels) which may occur on the path in the derived tree which corresponds to the d-edge. A set of nodes which are connected (directly or indirectly, though other nodes) by immediate dominance relations are referred to as a *component*. In Figure 4, there are five components.

If a d-tree has a single component, it is a tree. If a d-tree has more than one component, we can associate with it the set of trees which represent minimal readings of the d-tree, i.e., which have been obtained from the d-tree by identifying pairs of nodes of the d-tree in order to eliminate d-edges, but without introducing new nodes. Furthermore, in identifying nodes, we can only identify nodes that are at the top end of a d-edge with root nodes of components. This requirement means that the components of

[14] In DSG, precedence is always fully specified. It is fairly straightforward to extend the formalism to allow for underspecified linear precedence, but we will use the underspecified dominance relation to describe word order phenomena.

[15] We will discuss this example from a linguistic point of view shortly; right now, it serves to illustrate the formalism.

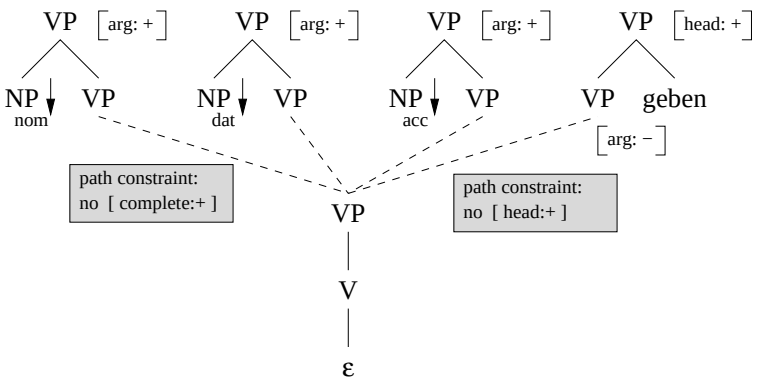

Figure 4: D-tree for German verb *geben* 'to give'

a d-tree will not "disappear" through identification of multiple pairs of nodes – structure, once defined in a d-tree, must be preserved when reading off trees. Figure 5 shows a tree which is a minimal reading of the d-tree in Figure 4. It is clear that we can obtain all six possible orderings of the three arguments of *geben* by choosing different ways of eliminating the d-edges. Note that in TAG we would need six elementary structures in order to capture the relevant syntactic behavior of *geben*, while in DSG we need only one, the d-tree in Figure 4.

in

Figure 5: Tree for German verb *geben* which is a minimal reading of the d-tree in Figure 4

Because of the underspecification, we do not need the operation of adjunction to intermingle the projections of different lexical items, and there is only the operation of substitution. Its definition is basically the same as in TAG: we identify frontier nodes which are labeled with nonterminals as *substitution nodes* (again marked with a downarrow $\downarrow$), and then append another d-tree at such nodes. But since we are appending d-trees, we can choose the root node of any of the d-tree's components as the node which we append to the substitution node. Formally, the operation of substitution consists in forming the union of two tree descriptions and of identifying two nodes, a substitution node and a root node of a component. A derivation consists of a sequence of substitutions. At the end of a derivation, we can obtain trees associated with the derived d-tree in exactly the same manner explained above for elementary d-trees.

We illustrate a derivation in DSG using an example of long scrambling.

(16) a. daß die Kinder dem Lehrer das Buch zu geben
that [the children]$_{NOM}$ [the teacher]$_{DAT}$ [the book]$_{ACC}$ to give
versuchen
try

that the children try to give the teacher the book

b. daß dem Lehrer die Kinder das Buch zu geben versuchen

c. daß dem Lehrer das Buch die Kinder zu geben versuchen

We can represent the matrix verb as shown in Figure 6, and the result of substituting the d-tree for *geben* (at the root of the component containing the verb) into the substitution node for the clausal complement in the d-tree for *versuchen* is shown in Figure 7. From this we can still obtain all six possible word orders, and that this would be impossible using simply LP rules that order sister nodes. (It would also be impossible in LTAG, but see (Joshi et al., 2000) for an alternate discussion of long scrambling in LTAG.) In fact, we also get another two orders, which correspond to cases in which the entire clause has scrambled:

(17) a. daß dem Lehrer das Buch zu geben die Kinder versuchen

b. daß das Buch dem Lehrer zu geben die Kinder versuchen

Note that because of the d-edges, we cannot derive word orders in which embedded arguments follow the embedded verb, which are not grammatical:

(18) a. * daß dem Lehrer zu geben das Buch die Kinder versuchen

b. * daß das Buch zu geben die Kinder dem Lehrer versuchen

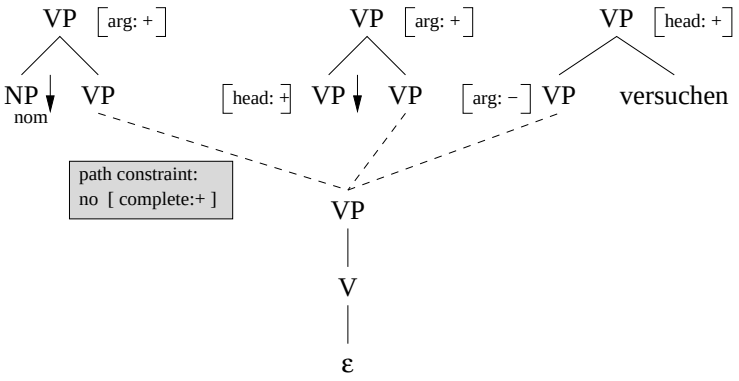

Figure 6: D-tree for German verb *versuchen* ‘to try’

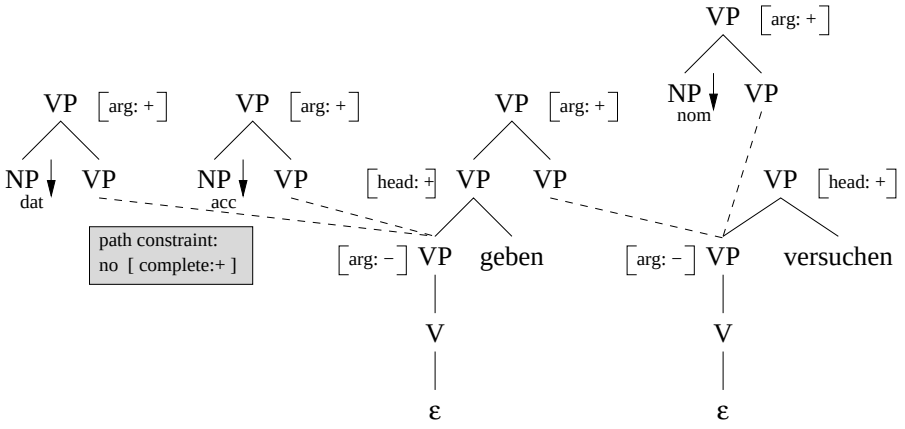

Figure 7: DSG derivation for complex sentence

Formally and computationally, DSG is an appealing formalism for the description of natural languages: if lexicalized,[16] the formalism generates only context-sensitive languages (Rambow, 1994a), and it is polynomially parsable (Rambow et al., 2001).

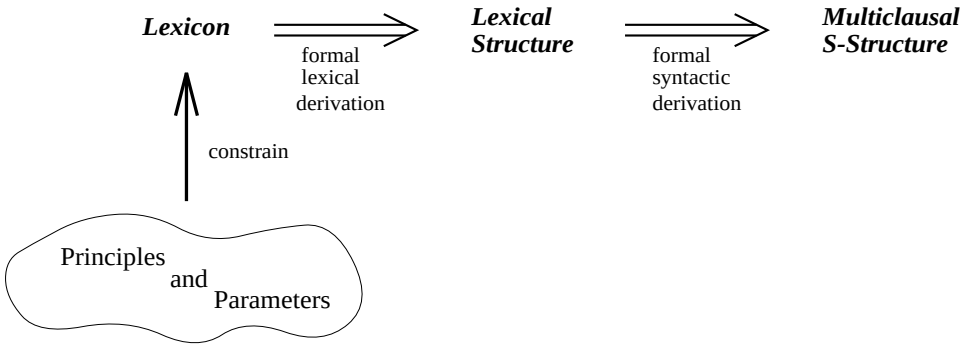

Figure 8: Using V-TAG as a metalanguage for linguistics

5.2 DSG as a Metalanguage for Syntax

DSG, like TAG, is a mathematical formalism and not a linguistic theory. There are many ways in which it can be used to develop a linguistic theory. In particular, it would be possible to use the same approach used for simple TAG, illustrated above in Figure 3: we define basic projections from lexical items along with procedures outside the

[16] Recall that a grammar is *lexicalized* if each tree set is associated with at least one lexical item, as will be the case for linguistically motivated formal grammars.

formalism for deriving related projections (with scrambled arguments, *wh*-movement, and the like). However, the fact that the elementary structures are now tree descriptions suggests a different approach, not previously available. Under the new "unified" approach, all projections from a lexical item are derived *within* the formalism. More precisely, a full derivation proceeds in two steps:

- In the *lexical* derivation, further specifications are added to a d-tree. The lexical derivation may, but need not form a (fully specified) tree. The resulting d-tree corresponds to the (extended) projection from that lexical item.
- In the *syntactic* derivation, the d-trees derived during the lexical derivation are combined using substitution in the usual way to form larger, multi-lexemic structures. At the end of the syntactic derivation, a tree is read off from the derived d-tree.

Thus, there no longer are two *formally* different types of derivations, lexical and syntactic. Instead, both types of derivation are formal operations within DSG, the difference being that the lexical derivation removes underspecification from a single d-tree, while the syntactic derivation combines d-trees and then removes underspecification from the resulting derived tree. This is illustrated schematically in Figure 8.

An important aspect of the linguistic theory for DSG is the use of fewer node labels. This is imposed on us because the lexical entry is underspecified with respect to how the head projects. For example, in German, we use the same lexical entry for verb-final clauses and verb-second (V2) clauses. Normally, the head in verb-final position would be dominated by a VP node, and the head in V2 position by a, say, C′ node. In our approach, both nodes must be labeled VP, since the label is given in the lexical entry common to both constructions. Instead, the difference between the different levels of projection from a lexical item is expressed using a system of binary features, called *categorial features*. The categorial features are independently motivated from semantic or from morphological facts. They constrain the lexical derivation; for example, they make sure that, upon choosing a verbal tree set, we can derive tree sets that correspond to a finite clause with a complementizer, a finite V2 clause, a finite V1 clause (question), or a finite or non-finite verb-final clause. These different lexical derivations are determined by an assignment of features to the lexical head of the tree set and to morphologically independent heads (auxiliaries, complementizers, determiners). There is not necessarily a one-to-one correspondence between features and heads. For a cross-linguistic discussion of how categorial features can derive a range of verb-second behavior, see (Rambow and Santorini, 1995). V2 will not be discussed further in this paper.

Since we are only interested in verb-final clauses in this paper, we omit all features related to V2. We assume that there is a tense feature which is related to case assignment in the usual manner, so that only tensed verbs can assign nominative case, and so that non-finite verbs must have an empty PRO subject. Furthermore, we assume that the component containing the origin of the projection (but not the lexical head) is lacking a feature which all other components have, and which

In addition, we will use the following three features related to the lexical head and its arguments:[17]

- The binary feature head has the value + at all levels of projection that dominate the lexical.
- The binary feature arg has the value + at all levels of projection that dominate at least one syntactic argument of the head. A VP with feature arg:- corresponds, in some sense, to the node label V in traditional systems.
- The binary feature complete has the value + at all levels of projection that dominate *all* of of the syntactic arguments of the lexical head. It designates the completion of the lexical predicate-argument structure in syntax.

This means that at the origin of the projection, all features have the value -. Once a head or an argument introduces a value +, the feature will have that value all the way up the projection.

Consider again the d-tree for *geben* shown in Figure 4. The head component has the feature head:+ at its root, while the components that introduce arguments have arg:+. Furthermore, the d-edge below the nominal components have a path constraint complete:- (this feature may not occur along the path corresponding to the d-edge in the derived tree), since clearly the predicate is not complete at a node if a nominal argument dominates it in the projection. The d-edge below the head has a path constraint of head:-. Note that the path constraints follow naturally from the same principle that introduces features: below the level in the projection at which the feature is set to +, it must be -.

5.3 Deriving Coherent and Incoherent Constructions

In this approach, the difference between coherent verbs such as *versprechen* 'to promise' and incoherent verbs such as *bereuen* 'to regret' is expressed by the feature content on the clausal substitution node in the matrix verb. We will assume that all clausal substitution nodes have feature head:+, meaning that they must dominate the head of the embedded clause. Incoherent verbs furthermore select for a VP with complete:+, i.e., for a syntactically complete predicate (Figure 9). As a result, the entire structure projected from the embedded verb must be dominated by the substitution node, since the path constraints on the dominance links preclude any component from being inserted above the substitution node. Put differently, in the lexical derivation, we can fully specify both the matrix and the embedded d-trees, since no interleaving of components will occur during the syntactic derivation.

For the coherent verbs, we assume that the VP substitution node is marked only for head:+, and that both features arg and complete are undefined. Now we can substitute any component of the *geben* d-tree into the substitution node. If we choose an argument component, we derive a sentence with at least one embedded argument in

[17] We could use names for these features which relate them more clearly to traditional node labels, but these feature names make their use clearer, and it is quite plausible to assume that we have access to a system of features that describe the lexical predicate-argument relations.

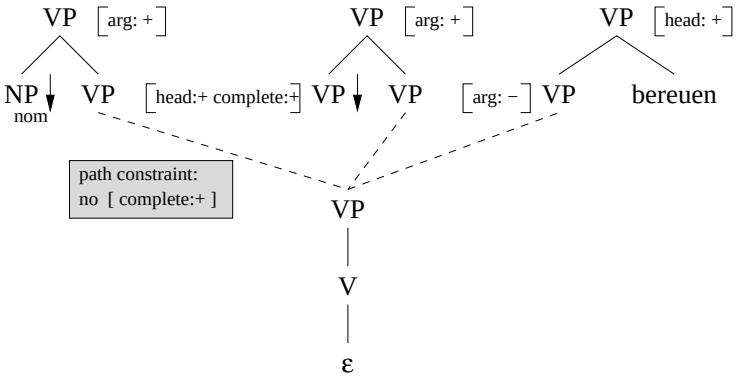

Figure 9: D-tree for incoherent verb *bereuen* 'to regret'

the embedded clause. If we choose the head component, we obtain the derived d-tree shown in Figure 7. Because of the d-edges, we can then read off trees that correspond to the word orders in (16) and (17).[18] If we first eliminate the d-edge which connects the components containing the clausal substitution node, the matrix head *versuchen*, and the matrix and embedded origins of the projections, we obtain the structure in Figure 10. It can be seen immediately that this stage of the derivation represents quite faithfully the incorporation analysis:

- The two verbs have formed a single constituent, whose feature is consistently arg:-, meaning a "lexical" level of projection which does not yet include arguments.
- This verb "cluster" has a list of arguments (i.e., components with argument substitution nodes) which corresponds to the union of the nominal arguments of the two verbs.

Of course, Figure 10 and the associated derivation does not implement the analysis of Evers (1975) faithfully. In particular, the "verb complex" does not in any way appear to have undergone a special morphological process.[19] And, as mentioned, the clauses were never "pruned", since this is a non-transformational system. However, I would like to claim that this derivation captures the underlying intuition of the lexical analysis, which has proved compelling to most native speaker linguists (even if the arguments put forward in favor of it have not necessarily). However, unlike the incorporation analysis, the representation in Figure 10 is just an artifact of our choice of how to read off a tree from the derived tree description: we chose to first eliminate underspecification around

[18]Note that some word orders can be obtained by assuming that an embedded constituent is either in the embedded clause, or in the matrix clause.

[19]The fact that the dominance links of the embedded verb's argument impinge lower in the verb cluster than those of the matrix verb is irrelevant, since the embedded arguments cannot be added except above the matrix verb.

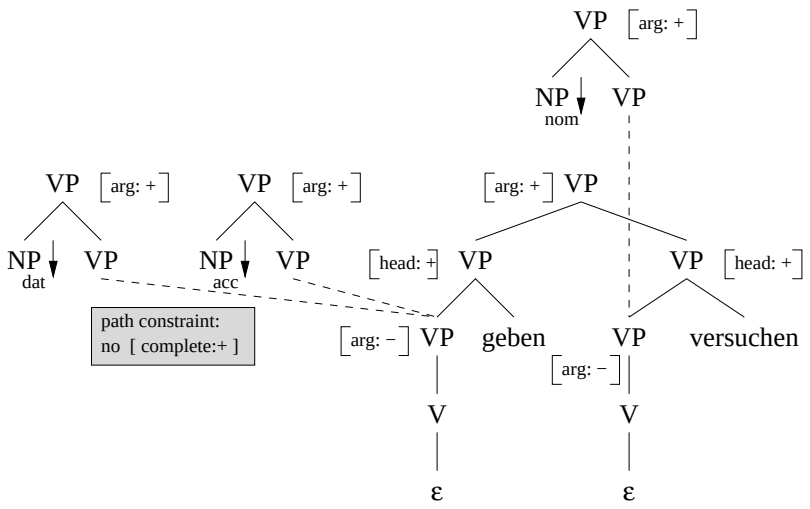

Figure 10: Possible intermediate point in DSG derivation for complex sentence, corresponding to "verb complex"

the two verbs. We could also have first proceeded to eliminate underspecification of the (matrix) arguments, in which case we would not have obtained an intermediate representation which corresponds to a "verb complex". Put differently, we obtain a verb complex in a derivation if the lexical derivation is kept very short and very little fully specified structure is associated with the lexemes prior to the syntactic derivation.

There is another option: we can mark the clausal substitution node with arg:- (with complete undefined). Now, there is only one derivation possible: the only component that can substitute at a node labeled VP head:+ arg:- is the component containing the head. As a result, the lexical derivation of the embedded component can at most eliminate the dominance link between the origin of the projection and the head component. All arguments will have to be inserted in the matrix clause. As a result, we cannot derive any of the word orders in Section 3.1, and in fact, in terms of empirical predictions, this analysis corresponds to the original incorporation analysis.

As we have seen in Section 4.2, there are empirical reasons for choosing the syntactic analysis over the incorporation analysis. However, using DSG, these two analyses are not as starkly different as they are in a transformational framework (where the incorporation analysis requires machinery not needed for similar constructions in other languages) or a pure TAG approach (where the incorporation analysis cannot naturally be represented at all). In fact, the difference is in the value of a single feature, arg, which is either undefined (syntactic analysis) or - (incorporation analysis). This rather slight difference corresponds to the sometimes slight difference in acceptability: usually, the relevant judgments are not clear-cut. By deriving different syntactic behavior from lexically determined features, it becomes more plausible to expect not only lexically idiosyncratic behavior, but also extensive dialectal and even individual variation.

And in either analysis, we can obtain the "verb cluster" representation of Figure 10 as an intermediate step in the reading-off process.

5.4 Accounting for the Data

It can easily be seen that the analyses of the coherent constructions discussed in Section 4.1 can be implemented straightforwardly in this framework if we assume that coherent verbs only specify head:+ on their clausal complements. Furthermore, the cases discussed in Section 4.2 that are problematic for a clause union analysis can also be derived. Under this view, as desired, the difference between coherent and incoherent verbs reduces to the fact that coherent verbs allow scrambling out of their complement.

The long passive facts remain to be explained.[20] I will only sketch a solution here. Long passives pose a problem for a purely syntactic account of coherent constructions in German, since passive is typically assumed to be a lexical transformation, and therefore cannot intervene after the syntactic derivation has started. However, in the framework proposed here, the lexical and the syntactic derivations are formally identical, and differ only in whether trees from one or from more than one set are combined. In (Rambow, 1994b), I suggest that lexical processes such as passive and object-shift can be captured by assigning different features to heads (just as in the case of V2). For example, the double-object construction and the NP-PP construction would both result from assembling the components of phrase structure in the lexical d-tree of a ditransitive verb in different manners, depending on the assignment of features to heads. In the case of the passive, the absorption of the ability to assign (say) accusative case by the verb (represented by the shifting of a feature representing case-assignment or case-agreement) is accompanied by an additional operation: the subject (or agent) is either marked as optional or removed entirely from the lexical tree set. There is no inherent reason why all feature assignments must be made to elementary d-trees. Suppose instead we perform some initial feature assignment to the d-tree, and then start the syntactic derivation to the point shown in Figure 10. We can then finish the feature assignment to the verbal heads in such a way that the embedded verb loses its ability to assign case; at the same time, we mark the matrix subject as being optional. We then obtain the long passive, in a manner quite analogous to the regular passive. (The embedded object would, as required, be assigned nominative case from the matrix verb. The embedded subject could not be marked optional since then there would not be enough case assigners to finish the derivation.) In the case of the incoherent construction, the same procedure could be followed, except that the embedded object could not get case in the matrix clause because the matrix verb has marked its complement with complete:+, so the embedded argument cannot leave the embedded clause.

It may be objected that the idea of passivizing what is in effect a complex clause is unappealing. However, there is cross-linguistic evidence from Italian causatives that such an operation is needed. In Italian causatives, the causative auxiliary and the main verb do not form a morphological word, but the construction allows long passives (Burzio, 1986; Heycock and Santorini, 1988).

[20] The partial VP fronting data also needs to be derived. To do so in DSG, we must assume that arguments can either dominate the origin of the projection, or the lexical head component itself. The details are left to further work.

6 Conclusion

The analysis presented in this paper is based on a formalism whose elementary structures are descriptions of trees. It suggests that the difference between the incorporation analysis and a purely syntactic analysis is really an artifact of the systems that are used to express them. If the mechanisms for projecting from a lexical item (lexical derivation) and for combining one or more such projections (syntactic derivation) are quite different and if this difference is furthermore expressed by differences in node labels (V and VP), then the German data will require a choice. The incorporation solution means that the standard syntactic mechanism for combining two lexical items with separate argument lists cannot be used, and a new lexical mechanism must be invented for German coherent constructions. The incorporation analysis does not appear to be empirically adequate, but the syntactic solution appears not to capture certain intuitions about coherent constructions. However, if we choose to assemble phrase-structure incrementally from underspecified descriptions during both the lexical and the syntactic derivations, we see that both the incorporation analysis and the syntactic analysis differ only in the value of one (or two) features in the projection level selected by the matrix verb. And the intuition of "verb cluster" is an effect of the derivation (in both the "syntactic" and "incorporation" analyses), not expressible explicitly in the competence grammar on their own.

Bibliography

Abeillé, A. (1990). Lexical and syntactic rules in a Tree Adjoining Grammar. In *28th Meeting of the Association for Computational Linguistics (ACL'90)*, Pittsburgh.

Abeillé, A. and Rambow, O. (2000). Tree Adjoining Grammar: An overview. In Abeillé, A. and Rambow, O., editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analyses and Processing*, pages 1–68. CSLI Publications.

Bayer, J. (1992). *Zum* in Bavarian and scrambling. Ms., Universität Aachen.

Bech, G. (1955). *Studien über das deutsche Verbum infinitum*. Det Kongelige Danske videnskabernes selskab. Historisk- Filosofiske Meddelelser, bd. 35, nr.2 (1955) and bd. 36, nr.6 (1957). Munksgaard, Kopenhagen. 2nd unrevised edition published 1983 by Max Niemeyer Verlag, Tübingen (Linguistische Arbeiten 139).

Becker, T. (2000). Patterns in metarules for TAG. In Abeillé, A. and Rambow, O., editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analyses and Processing*, pages 331–342. CSLI Publications.

Burzio, L. (1986). *Italian Syntax: A Government-Binding Approach*. Reidel, Dordrecht.

den Besten, H. and Rutten, J. (1989). On verb raising, extraposition and free word order in dutch. In Jaspers, D., editor, *Sentential complementation and the lexicon*, pages 41–56. Foris, Dordrecht.

Evans, R., Gazdar, G., and Weir, D. (2000). 'lexical rules' are just lexical rules. In Abeillé, A. and Rambow, O., editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analyses and Processing*, pages 71–100. CSLI Publications.

Evers, A. (1975). *The Transformational Cycle in Dutch and German*. PhD thesis, University of Utrecht. Distributed by the Indiana University Linguistics Club.

Frank, R. (1992). *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania.

Frank, R. (2001). *Phrase Structure Composition and Syntactic Dependencies*. MIT Press, Cambridge, Mass.

Haider, H. (1993). *Deutsche Syntax Generativ*. Tübinger Beiträge zur Linguistik 325. Gunter Narr Verlag, Tübingen.

Heycock, C. and Santorini, B. (1988). Remarks on causative and passive. Ms., University of Pennsylvania.

Höhle, T. (1978). *Lexikalische Syntax: Die Aktiv-Passiv-Relation und andere Infinitivkonstruktionen im Deutschen*. Niemeyer, Tübingen.

Joshi, A., Becker, T., and Rambow, O. (2000). A new twist on the competence/performance distinction. In Abeillé, A. and Rambow, O., editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analysis, and Processing*, pages 167–182. CSLI Publications.

Joshi, A. K. (1985). How much context-sensitivity is necessary for characterizing structural descriptions — Tree Adjoining Grammars. In Dowty, D., Karttunen, L., and Zwicky, A., editors, *Natural Language Processing — Theoretical, Computational and Psychological Perspective*, pages 206–250. Cambridge University Press, New York, NY. Originally presented in 1983.

Joshi, A. K. (1987). An introduction to Tree Adjoining Grammars. In Manaster-Ramer, A., editor, *Mathematics of Language*, pages 87–115. John Benjamins, Amsterdam.

Joshi, A. K. (1994). Preface to special issue on Tree-Adjoining Grammars. *Computational Intelligence*, 10(4):vii–xv.

Joshi, A. K., Levy, L., and Takahashi, M. (1975). Tree adjunct grammars. *J. Comput. Syst. Sci.*, 10:136–163.

Kaplan, R. M. and Bresnan, J. W. (1982). Lexical-functional grammar: A formal system for grammatical representation. In Bresnan, J. W., editor, *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Mass.

Kroch, A. (1987). Subjacency in a tree adjoining grammar. In Manaster-Ramer, A., editor, *Mathematics of Language*, pages 143–172. John Benjamins, Amsterdam.

Kroch, A. (1989). Asymmetries in long distance extraction in a Tree Adjoining Grammar. In Baltin, M. and Kroch, A., editors, *Alternative Conceptions of Phrase Structure*, pages 66–98. University of Chicago Press.

Kroch, A. and Santorini, B. (1991). The derived constituent structure of the West Germanic Verb Raising construction. In Freidin, R., editor, *Principles and parameters in comparative grammar*, pages 269–338. MIT Press, Cambridge, Mass.

Netter, K. (1991). Clause union phenomena and complex predicates in German. Technical Report R1.1.B (Part 1), DYANA.

Pollard, C. and Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Rambow, O. (1994a). *Formal and Computational Aspects of Natural Language Syntax*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania, Philadelphia. Available as Technical Report 94-08 from the Institute for Research in Cognitive Science (IRCS) and also at ftp://ftp.cis.upenn.edu/pub/rambow/thesis.ps.Z .

Rambow, O. (1994b). Mobile heads and strict lexicalization. Master's thesis, Department of Linguistics, University of Pennsylvania, Philadelphia.

Rambow, O. and Santorini, B. (1995). Incremental phrase structure generation and a universal theory of V2. In Beckman, J., editor, *Proceedings of NELS 25*, pages 373–387, Amherst, MA. GSLA.

Rambow, O., Vijay-Shanker, K., and Weir, D. (2001). D-Tree Substitution Grammars. *Computational Linguistics*, 27(1).

Steedman, M. (1991). Structure and Intonation. *Language*, 68(2):260–296.

Steedman, M. (1996). *Surface Structure and Interpretation*. MIT Press.

Steedman, M., editor (1997). *The Syntactic Interface*. MIT Press, Cambridge Massachusetts and London, England.

von Stechow, A. and Sternefeld, W. (1988). *Bausteine syntaktischen Wissens*. Westdeutscher Verlag.

Webelhuth, G. and Besten, H. (1987). Remnant topicalization and the constituent structure of VP in the germanic SOV languages. *GLOW Newsletter*, 18:15–16.