A Parallel Chart Parser for German, its Argument Interpretation Strategy, and the Treatment of Infinitives

Christopher Laenzlinger and Martin Simon Ulmann
laenzlinger@latl.unige.ch ulmann@latl.unige.ch

LATL
Department of Linguistics
University of Geneva
CH-1211 Geneva 4[1]

Abstract

In this paper we present a GB-parsing system for German and in particular its strategy for argument interpretation, which copes with the difficulty that word order is relatively free in German (due to scrambling) and also that the arguments can precede their predicate. In this case, the parser makes a provisional interpretation, which is checked when the argument structure of the predicate is available. Moreover, a strategy of argument transfer is used in cases of long-distance scrambling, according to which arguments and adjuncts are attached to the domain of the coherent verb, `ECM` verb, or raising verb, and transferred to the infinitival complement for interpretation.

1 Introduction

Free word order languages raise difficulties for parsing systems based on phrase-structure rule grammars where the constituents are ordered. Indeed, to list all the possible orders leads to an increase of the grammar size and a decrease of performance. There have been several approaches to this problem, notably those based on the `ID/LP` (immediate dominance/linear precedence) grammars (*cf.* Gazdar *et al.* 1985) or functional unification grammars (*cf.* Karttunen & Kay 1985). Within Government and Binding Theory, Kashket (1991) presents a parser for Warlpiri, a non-configurational language, where word order and its variation depends mainly on case marking.

Although German is a partially free word order language, we will make the hypothesis that it has a fixed base word order, which is modified by a set of transformations. In the first part of this paper, we will present our parallel chart parser for German, and its argument interpretation strategy. This strategy is able to handle the difficulties arising from word order variations. In the second part of the paper, we will discuss additional difficulties that occur with the treatment of infinitival constructions and require an extension of the argument interpretation strategy.[2]

[1]The research described in this paper has been supported in part by a grant from the Swiss national science foundation (grant no 11-33731.92).

[2]There are other proposals to deal with infinitival constructions in German: Netter 1986 discusses an LFG approach, and Rambow 1994 uses a variation of TAG.

2 The DIPS Parser

2.1 General Properties of DIPS

DIPS (*Deutsches Interaktives Parsing System* 'German Interactive Parsing System') is a large-scale interactive GB-based[3] parsing system. Its architecture is basically similar to that of IPS (Wehrli 1992) and FIPS (Laenzlinger & Wehrli 1991). The parser produces a set of GB S-structures (trees) from an input sentence. These structures are associated with information concerning traces, argument structure, and case features.

The syntactic structure of constituents corresponds to the GB $\overline{\text{X}}$-schema. We consider German to be an SOV-language, that is, a language in which the object precedes its predicate, when they are in their basic position. Thus, the $\overline{\text{X}}$-schema is parameterized in German as follows: The complement `Compl` precedes the head X^0 for the categories `V, A, I`, whereas it follows the head for the categories `C, D, P, N, Adv, F`.[4] As the specifier is always on the left, the $\overline{\text{X}}$-schema has the structure given in (1).

$$
\begin{array}{lllll}
 & \text{XP} & \rightarrow & \text{Spec } \overline{\text{X}} & \\
(1) & \overline{\text{X}} & \rightarrow & \text{X}^0 \text{ Compl}, & \text{if X}^0=\{\text{C}^0, \text{D}^0, \text{P}^0, \text{N}^0, \text{Adv}^0, \text{F}^0\} \\
 & \overline{\text{X}} & \rightarrow & \text{Compl X}^0, & \text{if X}^0=\{\text{V}^0, \text{A}^0, \text{I}^0\}
\end{array}
$$

On the basis of this schema, the clause structure in German has the general representation given in Figure 1.

This $\overline{\text{X}}$-structure is implemented as a node which basically consists of a head (the lexical item), a bundle of features (such as category, agreement, case, selection, subcategorization), a list of specifiers (`Spec`) and a list of complements (`Compl`). These lists either contain other nodes or are empty.

2.2 Strategy

The parsing strategy is data-triggered (mainly bottom-up), proceeds from left to right, and treats alternatives in parallel by using a chart (*cf.* Kay 1967 and Kaplan 1973). The analysis of a sentence proceeds at two levels: the lexical level and the syntactic level. The lexical analysis looks up the words in the lexicon; each lexical item (word) projects the node corresponding to its category; thus, the lexical features are transferred to the syntactic level in accordance with the Projection Principle (Chomsky 1981). The projected node is then inserted into the chart as an edge. The syntactic analysis builds all possible structures by making use of cross-category projections and attachments, which are further filtered by grammatical constraints; structure building is incremental, as the current constituent is immediately integrated into the existing hypotheses.

Cross-category projection creates a new constituent with the same start and end vertex in the chart as the subconstituent from which it is projected. This kind of projection is limited to some categories and triggered by intrinsic features. For instance, an infinitival verb projects the structure in Figure 1 from `VP` to `CP`.

[3] *Cf.* Chomsky & Lasnik 1992 for a presentation of Government and Binding Theory (GB), and Berwick *et al.* 1991, Wehrli 1988 for a possible implementation of the theory.

[4] We assume Abney's 1987 DP-hypothesis, according to which the head of a noun phrase is the determiner (D^0). The category F stands for a functional head, the complement of which is an AP or a VP.

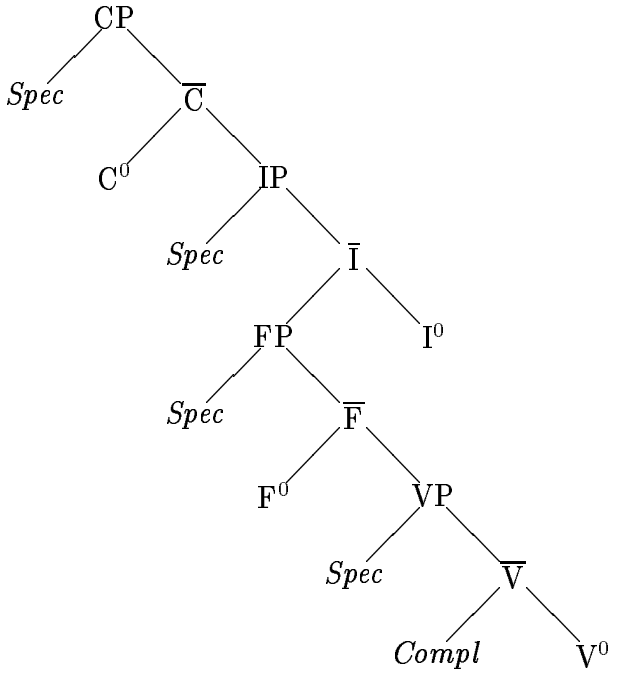

Figure 1: Structure of a German clause

Attachment combines the current constituent with the constituents which immediately precede this current constituent in the chart. The resulting constituents are edges from the start vertex of the left constituent to the end vertex of the current constituent. Attachments can be divided into two different types of combination:

1. A constituent of the left context is attached to the current constituent (left attachment).
2. The current constituent is attached to a constituent of the left context (right attachment).

In order to keep track of where a constituent can be attached in the structure, a list of active nodes specifies the potential attachment sites; this list is systematically updated. It is organized as a stack, with the sites for the left attachments higher than those for the right attachments. For both classes of attachment sites, the highest sites in the tree are put on the stack first, *i.e.* the deeper the site in the tree structure, the higher on the stack. When the two constituents are combined, the sites which are higher on the stack (and thus deeper in the structure) than the current attachment site are removed, because only adjacent constituents can be combinded, *i.e.* no other constituent can intervene. As to right attachment, the sites for right attachment of the two constituents are combined.

The possibilities of attachments are constrained as follows:

1. formal relation, which is restricted to adjacent constituents and ruled by lexical features such as selection and agreement (*e.g.* auxiliary-verb selection, determiner-noun agreement)

2. non-formal relation, which concerns complements and their subcategorizing head.

The second type of relation requires a specific argument interpretation strategy (AIS) to establish the link between the argument and the predicate which subcategorizes it.

2.3 The Argument Interpretation Strategy

The aim of the AIS is to match the arguments with the subcategorization properties (argument structure) of the predicate, and thus to establish an interpretation, which corresponds to the assignment of the thematic roles. The argument structure of a verb (predicate) is provided by the lexicon and specifies the number and type of arguments that the predicate can take; while there can be more than one argument structure for a verb at the lexical level, there is only one argument table for a (verb) node at the syntactic level, which contains the arguments of the clause. This argument table is matched with the corresponding argument structures, which has the effect of filtering the inappropriate argument structures.

The type declaration of this argument table is given in (2). This table is defined as an array of arguments. Each of these arguments consists of a pointer to a node (`ProjectionPtr`), its category, and its grammatical function. The variable *theta* can have three values; if it is 'true', the argument has already received a thematic function, which is assigned to the variable `thetarole`; it is 'false', if the constituent does not get a thematic role (*e.g.* expletive subject); the value 'invalid' stands for a uninterpreted argument (*cf.* section 3.2.3).

(2)

```
ArgumentTable = ARRAY max_nb_of_arg OF Argument;
Argument      = RECORD
                  node : ProjectionPtr;
                  cat  : Category;
                  gf   : GrammaticalFunction;
                  CASE theta OF
                    true    : thetarole : ThematicRole;
                    false   : (* No theta function *)
                    invalid : (* not interpreted yet *)
                  END
                END
```

The AIS has to deal with two types of difficulties: first, the predicate (with its argument structure) is not always available at the time the argument is attached; second, the large number of possible word orders in German makes the argument's grammatical function difficult to determine.

The argument structure is only available if the main verb (predicate) occurs in C^0, that is the second position in the clause (verb second with the main verb), and thus at most one argument precedes the verb. In this case, a final interpretation of the arguments is established immediately (at the moment of attachment); the arguments are inserted into the definitive argument table of the clause and interpreted by being matched with the argument structure of the verb (theta assignment); if more than one interpretation is possible, different hypotheses are considered in parallel. If the verb follows the arguments, they are also inserted into the

argument table, with a provisional interpretation.[5] The matching between the argument table and the argument structure eventually takes place at the time the main verb is attached.

The task of identifying the grammatical function of an argument is complicated by the large number of possible word orders, which results from the interaction of three syntactic processes: verb second, scrambling, and extraposition. The verb second constraint requires that the tensed verb occupies the second position of the main clause; for the first position, however, a large number of constituents (XP) is possible, such as the subject, an object, an adjunct, an empty operator. Scrambling is a process that modifies the order of clause-internal arguments and adjuncts under some constraints (*cf.* Uszkoreit 1987). Extraposition is the occurrence of prepositional or sentential complements or adjuncts after the verb in its base position V^0. Owing to word order variation, the grammatical function of an argument depends not only on its position, but also on case and agreement information and (scrambling) ordering constraints.

The flowchart of the interpretation module is given in Figure 2. The first step is to check whether there are arguments to be interpreted. If so, it is further checked whether the main verb is available, with the argument structures. In case it is not available, the new argument is inserted into the provisional argument table (and its interpretation can be checked only later, when the argument structure is available). If it is available, the new argument is matched with the argument structures; if there is a provisional argument table instead of one argument, the matching is done for each argument in turn. Thus, the list of argument structures is filtered and a list of new argument tables is returned. For each of these argument tables, it is checked whether its arguments obey the ordering constraints. If so, the new structure is completed and for each argument that is not in its base position, a chain is created to link the argument with that position, in which a trace is inserted.

Let us illustrate how the analysis proceeds on the base of the sentence in (3).

(3) Die Kinder haben diesen Bericht gelesen.
'the children have this report read'
the children have read this report.

When the parser reads the verb *haben*, the general clause structure (*cf.* Figure 1) is projected from VP to CP, triggered by the tensed verb, which is placed in C^0 leaving a head trace in V^0 and in I^0. Then, the first constituent *die Kinder* is attached as the specifier of the CP, producing the structure in (4).

(4) $[_{CP}\ [_{DP}$ die Kinder$]\ [_{\overline{C}}$ haben$_i$ $[_{IP}\ [_{FP}\ [_{VP}\ t_i]]\ t_i]]]$

For the verb *haben*, three hypotheses can be considered in parallel. First, *haben* can be a main verb, *e.g. Die Kinder haben viele Spielzeuge* 'the children have many toys'. Second, it can be the root of a verb with particle, *e.g. Die Kinder haben schwarze Hosen an.* 'the children have black trousers on' (The children wear black trousers). Finally, it can be an auxiliary as in example (3).

As the first argument *die Kinder* is morphologically ambiguous between nominative and accusative, it can be interpreted a priori as a subject or as a direct object. Under the

[5] This strategy seems to have psycholinguistic support: German speakers assign an interpretation to arguments even before the predicate is available (*cf.* Bader & Lasser 1993).

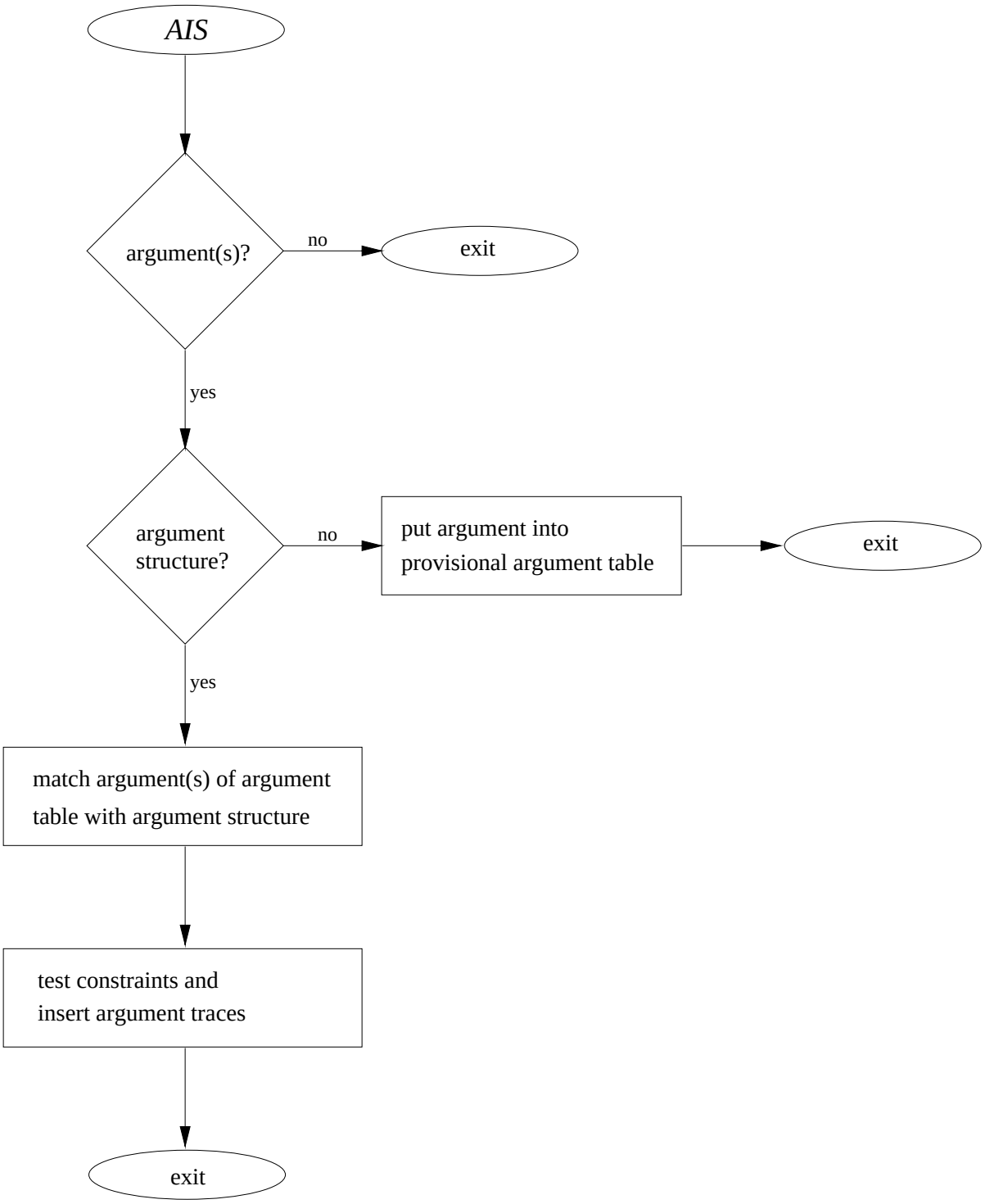

Figure 2: The interpretation module

first hypothesis (*haben* is a main verb), the argument structure of the verb is available and the subject-object ambiguity is solved when the argument *den Bericht*, non-ambiguously accusative, is analyzed as a direct object. This analysis, however, fails when the parser arrives at the participle *gelesen*. For the second and the third hypothesis, *die Kinder* is inserted into the provisional argument table as the subject or the direct object of a forthcoming verb. Afterwards, *diesen Bericht* is inserted into the provisional argument table as direct object. When the past participle is read, the second hypothesis fails since a particle has been expected instead of a participle. The third hypothesis, on the contrary, succeeds; the arguments are matched with the argument structure of *gelesen*, which requires a subject (*die Kinder*) and a direct object (*diesen Bericht*); thereby, the object reading of *die Kinder* is filtered out. A trace is inserted into the specifier of IP for the subject, and another trace into the complement of VP for the direct object. The analysis of (3) is given in (5)

(5) $[_{CP}\ [_{DP}$ die Kinder $]_j [_{\overline{C}}$ haben$_i$ $[_{IP}\ [_{DP}\ t]_j\ [_{DP}$ diesen Bericht $]_k$
$[_{\overline{I}}\ [_{FP}\ [_{VP}\ [_{VP}\ [_{DP}\ t]_k$ gelesen $]\ t_i\ ]]\ t'_i\ ]]]]$

In the following section, we will show how the AIS works in the case of infinitival constructions.

3 The Treatment of Infinitives

3.1 Different Infinitival Structures

German displays two types of infinitives: infinitives introduced by the conjunction *zu*, described in section 3.1.2, and infinitives without *zu*, presented in section 3.1.1.

3.1.1 Infinitives without *zu*

Infinitives without *zu* occur as the complement of modal verbs (*e.g.* *müssen* 'must', *können* 'can', *dürfen* 'may') and exceptional case marking (ECM) verbs (*e.g.* *sehen* 'see', *lassen* 'let/make'). Modals are treated on a par with auxiliaries, *i.e.* they are taken to select an infinitival VP as complement and are not associated with an argument table. In compound tenses, the infinitival form of the modal is usually used instead of its past participle form; in example (6a), the infinitive *wollen* substitutes for the participle *gewollt*. This phenomenon is called *infinitivus pro participio* (IPP) or *Ersatzinfinitiv*. If the verb selecting the IPP is in its base position, the order of the verbs differs from the usual one: auxiliaries that would be at the right of the IPP immediately precede the final predicates, as illustrated in example (6b), where *hätte* precedes *besuchen wollen*.

(6)a. Das Kind hat die alte Frau besuchen wollen.
'the child had the old woman visit want'
The child wanted to visit the old woman.

b. Wenn das Kind die alte Frau hätte besuchen wollen...
'if the child the old woman would-have visit want'
If the child had wanted to visit the old woman...

From a structural point of view, this reordering can be analyzed as verb raising (VR): the verbs which whould precede the uppermost final auxiliary (without VR) are attached to the right of the auxiliary head (right-adjoined position), forming head chains with their base positions, as represented in (7).

(7) $[_{\texttt{VP}}\ [_{\texttt{VP}}\ [_{\texttt{VP}}\ t_i]\, t_j]$ hätte besuchen$_i$ wollen$_j$]

Unlike modals, `ECM` verbs are analyzed as taking an infinitival `CP` as complement and assign (accusative or dative) case to the subject of the infinitival clause. In example (8a), the subject of the embedded infinitival clause *den Gauner* receives accusative case from the verb *sehen*. The phenomena of IPP and verb raising also occur with `ECM` verbs, as example (8a) shows.

(8)a. Nachdem die Polizei den Gauner hatte fliehen sehen...
'after the police the rogue had escape see'
After the police had seen the rogue escape...

b. ...daß ihn$_i$ der Wächter $[_{\texttt{CP}}\ [_{\texttt{DP}}\ t]_i$ den Wagen stehlen] sah.
'that him the guard the car steal saw'
...that the guard saw him steal the car.

Furthermore, the subject of the infinitival clause can be attached to the main clause as a result of scrambling. In example (8b), the subject *ihn* is scrambled out of the embedded clause to a position higher than the subject of the main clause *der Wächter*.

3.1.2 Infinitives with *zu*

The subject of infinitival clauses with *zu* is an empty constituent. In control constructions, the subject is a null pronoun `PRO`, which can be coreferential with (controlled by) the subject (example (9a)) or the object (example (9b)) of the upper clause according to the lexical property of the main verb. In raising constructions, the subject of the infinitive is a trace which forms a chain with the subject of the higher clause (example (9c)).

(9)a. Er$_i$ behauptete, $[_{\texttt{CP}}$ PRO$_i$ sie gesehen zu haben].
'he claimed PRO her seen to have'
he claimed to have seen her.

b. Sie hat ihm$_i$ t$_j$ erlaubt, $[_{\texttt{CP}}$ PRO$_i$ das Buch anzusehen]$_j$.
'she has allowed him, PRO the book to-look-at'
she allowed him to look at the book.

c. Sie$_i$ schien $[_{\texttt{CP}}\ t_i$ ihn gesehen zu haben].
'she seemed him seen to have'
She seemed to have seen him.

The infinitival clause can be extraposed in control constructions, as it is the case in (9b), but not in raising construction.

Among subject-control verbs, there is a class of verbs, called 'coherent verbs', which form a clause union with their infinitival complement (by restructuration). As a consequence, arguments and adjuncts attached to the upper clause can be interpreted with respect to the infinitival clause.

(10) Gestern hat sie$_i$ der Professor versucht [$_{\texttt{CP}}$ t$_i$ zu küssen].
'yesterday has her the professor tried to kiss'
Yesterday the professor tried to kiss her.

In the example (10), the pronoun *sie* is the direct object of the infinitival clause, although it is attached to the main clause.

3.2 Treatment of Infinitival Particularities

3.2.1 Verb Raising

The main problem with VR is that the verbs occur on the right of the uppermost final auxiliary, while their maximal VP constituents remains on the left and contain a head trace. The solution to this problem consists in attaching the structure that contains the verb to the left and in extracting all the heads, which are adjoined to the right of the upper verb. Take for instance the VP in (11a) and the complex VP in (11b); the latter is attached as the complement of the former. As the given word order *hätte besuchen wollen* must not be changed, the constituent that is attached must not contain any lexical material. The heads *besuchen* and *wollen* must be extracted and attached to the right of *hätte*. To obtain the right order, the constituent (11b) is traversed in postorder, so that the head *besuchen* is adjoined first and then the head *wollen*.

(11)a. [$_{\texttt{VP}}$ hätte]

b. [$_{\texttt{VP}}$ [$_{\texttt{VP}}$ besuchen] wollen]

c. [$_{\texttt{VP}}$ [$_{\texttt{VP}}$ [$_{\texttt{VP}}$ t$_i$] t$_j$] hätte besuchen$_i$ wollen$_j$]

The resulting structure is that in (11c); since objects and adjuncts preceding the predicate may be attached as complements to the main verb, the attachment sites for left attachment must be transferred from (11b) to (11c). This solution also works for verb raising in ECM constructions, although the verbal head of the infinitival clause is deeper in the structure.

3.2.2 Control

An infinitive with *zu* projects an infinitival clause to which an empty subject is added, as illustrated in example (12):

(12) [$_{\texttt{CP}}$ [$_{\texttt{IP}}$ [$_{\texttt{DP}}$ e] [$_{\overline{\texttt{I}}}$ [$_{\texttt{FP}}$ [$_{\texttt{VP}}$ zu küssen]]]]]

If the infinitival clause is a complement of a control verb, the empty subject must be coindexed with the controlling argument. The fact that an infinitival clause can precede or follow its controller, as illustrated in (13), requires that the coindexation only applies when both arguments are available, *i.e.* when the second one — the infinitival CP in (13a) and the indirect object in (13b) — is inserted into the definitive argument table. As the lexical information of the verb specifies which argument is the controller, it can be checked if both arguments are already available.

(13)a. Die Mutter erlaubte ihrer Tochter$_i$ nicht, [$_{CP}$ PRO$_i$ ins Theater zu gehen].
'The mother allowed her daughter not to the theatre to go'
The mother did not allow her daughter to go to the theatre.

b. [$_{CP}$ PRO$_i$ Ins Theater zu gehen] erlaubte die Mutter ihrer Tochter$_i$ nicht.
'to the theatre to go allowed the mother her daughter not'
The mother did not allow her daughter to go to the theatre.

3.2.3 Argument restructuration

Since restructuration allows arguments and adjuncts to be attached to the clause containing a coherent verb, while being interpreted with respect to the infinitival clause, the AIS needs to be extended.[6]

Figure 3 is an extended version of the lower part of Figure 2. The first modification concerns the matching procedure: An argument that may be interpreted with respect to the infinitival complement is left in the argument table of the coherent verb — for this reading, no matching takes place and this argument is marked as 'uninterpreted' (The value 'invalid' is assigned to the theta variable of the argument in the argument table, *cf.* the type declaration in (2)). The second modification is an extension inserted after the procedure that makes the argument traces. At this stage, it is checked whether there are arguments marked as 'uninterpreted' and whether the infinitival complement is available. If both of these conditions are fulfilled, the uninterpreted arguments are transferred from the argument table of the main verb to a provisional argument table, which is matched with the predicate of the infinitival complement.

Consider example (10), repeated here in (14a): the pronoun *sie* and the nominative DP *der Professor* are attached to the main clause. When the parser reads *versucht*, it interprets the DP non-ambiguously as subject. As far as the pronoun is concerned, two analyses are taken into account (in parallel). On the one hand, it can be the direct object of the main verb, as shown in (14b); this analysis fails later on, when the infinitival complement cannot be attached.

(14)a. Gestern hat sie$_i$ der Professor versucht [$_{CP}$ t$_i$ zu küssen].
'yesterday has her the professor tried to kiss'
Yesterday the professor tried to kiss her.

[6] Argument scrambling in coherent constructions cannot be treated in the same way as long distance *wh*-movement because (i) the argument is always on the left of its trace in *wh*-constructions, but it can also be on the right in coherent constructions; (ii) more than one argument can be scrambled under restructuration, whereas only one *wh*-argument can move, and (iii) only *wh*-movement applies cyclically.

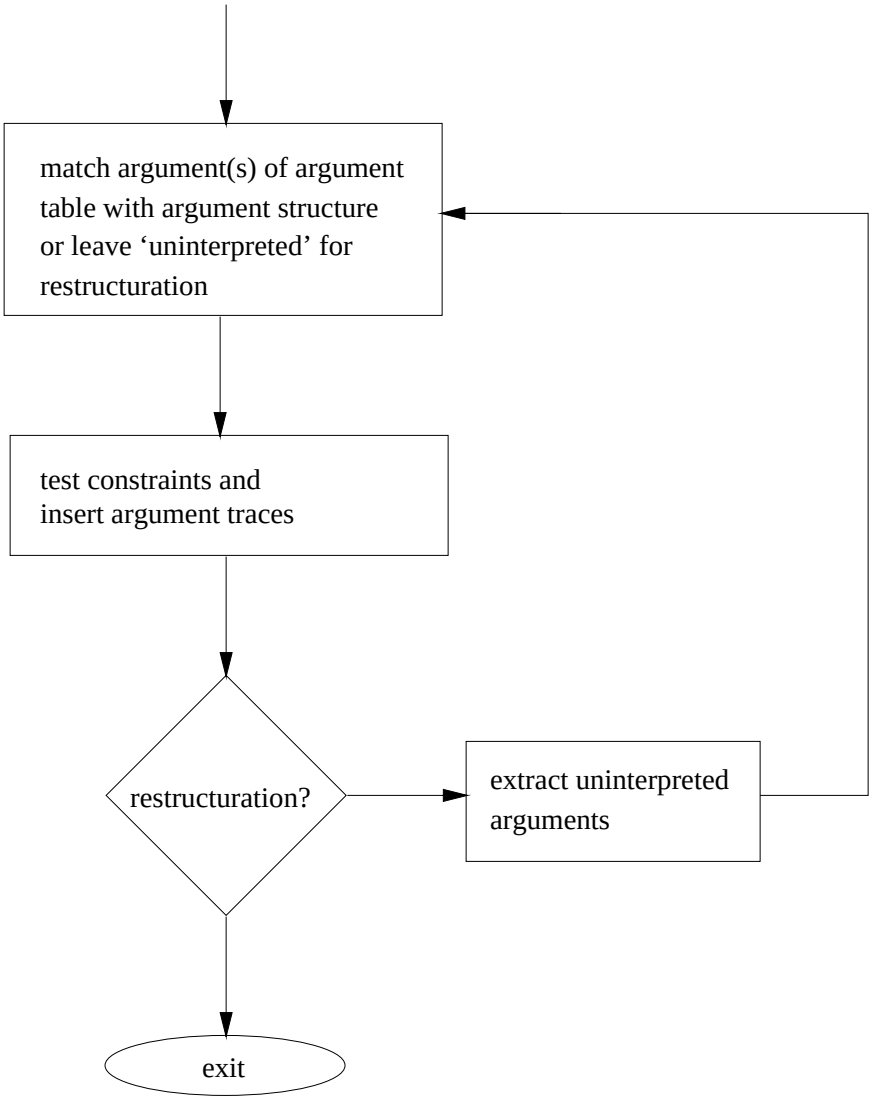

Figure 3: The extension for restructuration of the interpretation module

b. Gestern hat sie der Professor versucht.
'yesterday has them the professor tried'
Yesterday, the professor has tried them (*e.g.* the biscuits).

On the other hand, the pronoun *sie* is regarded as the argument of a following infinitival clause, *i.e.* it is assigned the status 'uninterpreted'. When the infinitival complement *zu küssen* is attached, this uninterpreted argument is treated as a new argument of this infinitive and interpreted as its direct object. The structure provided by the parser is given in (15).

(15) [$_{\mathrm{CP}}$ [$_{\mathrm{AdvP}}$ Gestern] [$_{\overline{\mathrm{C}}}$ hat$_i$ [$_{\mathrm{IP}}$ [$_{\mathrm{DP}}$ sie]$_j$ [$_{\mathrm{DP}}$ der [$_{\mathrm{NP}}$ Professor]]$_k$ [$_{\overline{\mathrm{I}}}$ [$_{\mathrm{FP}}$ [$_{\mathrm{VP}}$ [$_{\mathrm{VP}}$ [$_{\mathrm{CP}}$ t]$_l$ versucht] t_i] [$_{\mathrm{CP}}$ [$_{\mathrm{IP}}$ [$_{\mathrm{DP}}$ PRO]$_k$ [$_{\overline{\mathrm{I}}}$ [$_{\mathrm{FP}}$ [$_{\mathrm{VP}}$ [$_{\mathrm{DP}}$ t]$_j$ zu küssen]]]]]$_l$] t'_i]]]]

Example (16a), discussed by Rambow (1994: 17-23), is a sentence containing multiple coherent verbs, which illustrates the recursive application of argument transfer. The structure of the sentence is given in (16b).

(16)a. ?...weil das Fahrrad niemand zu reparieren verspricht zu versuchen.

'because the bicycle (acc) no-one (nom) to repair promises to try'
...because no one promisis to try to repair the bicycle

b. [$_{\texttt{CP1}}$ weil [$_{\texttt{DP}}$ das Fahrrad]$_i$ niemand [$_{\texttt{CP3}}$ t_i zu reparieren]$_j$ t_k verspricht [$_{\texttt{CP2}}$ t_j zu versuchen]$_k$]

The direct object *das Fahrrad* of the most deeply embedded infinitive *zu reparieren* is attached to the main clause CP1 (long-distance scrambling). In addition, the CP3 is scrambled out of CP2, which is extraposed after the finite main verb.

The parser proceeds as follows: The three arguments preceding the main verb *verspricht* are attached and inserted into the provisional argument table. When the parser reads *verspricht*, the matching procedure applies. *Das Fahrrad* can be nominative or accusative. Therefore, three readings are temporarily possible: subject, direct object and uninterpreted (direct object of a following infinitival complement). Since *niemand* is non-ambiguously interpreted as the subject of *versucht*, the subject reading for *das Fahrrad* fails. On the one hand, the CP3 *zu reparieren* can be interpreted as sentential complement of the main verb *versucht*, which produces an interpretation of *das Fahrrad* as the long-distance scrambled argument of *zu reparieren*, resulting in the grammatical sentence given in (17).

(17) ...weil [$_{\texttt{DP}}$ das Fahrrad]$_i$ niemand [$_{\texttt{CP}}$ t$_i$ zu reparieren] verspricht.
'because the bicycle (acc) no-one (nom) to repair promises'
...because no one promisis to repair the bicycle

This interpretation will fail, since the CP2 *zu versuchen* cannot be attached. On the other hand, CP3 can be left uninterpreted; when the CP2 *zu versuchen* is attached, and interpreted as the sentential complement of *verspricht*, the two uninterpreted arguments *das Fahrrad* and *zu reparieren* are transferred to the CP2 for interpretation. The CP3 *zu reparieren* is interpreted as sentential object of *versuchen*, while *das Fahrrad* is regarded as uninterpreted again, and thus is transferred to the CP3, where it is interpreted as the direct object of *reparieren*.

The same strategy of argument transfer holds for ECM constructions in which a subject is scrambled to the upper clause (*cf.* example (8b)). Almost in the same way, this strategy applies to the arguments of the infinitival clause in raising constructions. While the complements of the infinitival clause are treated in the same way as in the restructuring case, the subject is inserted into the argument table of the raising verb with the grammatical function 'subject', but without thematic role (theta is 'false', *cf.* the type declarations in (2)); thus, it is inserted into the argument table a second time, marked as 'uninterpreted'; therefore, it can be treated like other restructuring arguments, *i.e.* it is transferred to the infinitival clause and interpreted as the logical subject of the embedded clause.

(18) ...daß ihn die Frau zu schlagen scheint.
'that him the woman to beat seems'
that the woman seems to beat him.

Consider example (18). The direct object *ihn* of the infinitival verb *schlagen*, being in a scrambled position, is inserted into the argument table of the raising verb and marked as 'uninterpreted'. The subject *die Frau* is inserted into this argument table as the surface subject of *scheint* without a thematic role; it is reinserted as an uninterpreted argument of a following

infinitival complement so that it will eventually get a thematic role. When the infinitival clause *zu schlagen* is interpreted as the sentential complement of *scheint*, the uninterpreted arguments *ihn* and *die Frau* are transferred to this clause in order to be interpreted with respect to the verb *schlagen*; *die Frau* is taken as the subject of *schlagen* with the thematic role 'agent' and *ihn* as the direct object of the same verb with the thematic role 'patient'.

4 Conclusion

The task of our DIPS parser consists not only in building one or more trees for an input sentence, but also in determining the grammatical function and the thematic interpretation of arguments. We have discussed the parsing strategy in detail and shown that it is adequate for the treatment not only of finite clauses, but also of non-finite clauses. This strategy relies on the following steps: immediate attachment, provisional and definitive interpretation, testing of constraints, creation of chains, and restructuration. We have developed an argument interpretation strategy to deal with constructions in which the arguments can be treated in nearly the same way, regardless of whether they precede or follow the verb. This strategy has been extended to handle long-distance scrambling, so that arguments are transferred from the clause in which they are attached to an embedded clause in which they receive an interpretation.

The fact that general mechanisms, which may operate in combination, are used instead of construction-specific rules allows the coverage of a large number of constructions, while the complexity of the system is kept relatively low. DIPS is a practical system under development, which uses a large-sized lexicon (over 150'000 entries) and which, at present, covers a large range of grammatical constructions such as simple and complex sentences, finite and non-finite clauses, active and passive voice, *wh*-constructions, topicalization, extraposition, scrambling, long-distance dependencies, and verb raising.

Bibliography

Abney, S. 1987. *The English Noun Phrase in its Sentential Aspect*, doctoral dissertation, MIT, Cambridge, Mass.

Bader, M & I. Lasser. 1993. "German verb-final clauses and sentence processing: evidence for immediate attachment", ms., to appear in C. Clifton, L. Frazier & K. Rayner (eds.), *Perspectives in Sentence Processing*.

Berwick, R., S. Abney & C. Tenny (eds.). 1991. *Principle-Based Parsing: Computation and Psycholinguistics*, Kluwer, Dordrecht.

Chomsky, N. 1981. *Lectures on Government and Binding*, Foris Publications, Dordrecht.

Chomsky, N. & H. Lasnik. 1992. "Principles and Parameters Theory", in J. Jacobs, A. van Stechow, W. Sternefeld & T. Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, Walter de Gruyter, Berlin.

Gazdar, G, E. Klein, G. Pullum & I. Sag, 1985. *Generalized Phrase Structure Grammar*, Harvard University Press, Cambridge, Mass.

Kaplan, R. 1973. "A general syntactic processor", in R. Rustin (ed.), *Natural Language Processing*, Algorithmics Press, New York, 193-241.

Karttunen L. & M. Kay, 1985. "Parsing in a free word order language", in D. Dowty, L. Karttunen et A. Zwicky (eds.) *Natural Language Parsing*, Cambridge University Press, Cambridge, 279-306.

Kashket, M., 1991. "Parsing Warlpiri, a free word order language", in Berwick et al. (eds.), 123-151.

Kay, M. 1967. "Experiments with a powerful parser", in *COLING-67*.

Laenzlinger, C. & E. Wehrli. 1991. "FIPS: Un analyseur interactif pour le français", *TA informations*, 32.2, 35-49.

Netter, K. 1986. "Getting things out of order (An LFG-proposal for the treatment of German word order)", in *COLING-86*, 494-496.

Rambow, O. 1994. "Formal and computational aspects of natural language syntax." IRCS Report 94-08, doctoral dissertation, University of Pennsylvania.

Uszkoreit, H. 1987. *Word Order and Constituent Structure in German*, CSLI, Lecture notes 8, Stanford University.

Wehrli, E. 1988. "Parsing with a GB grammar", in U. Reyle and C. Rohrer (eds.), *Natural Language Parsing and Linguistic Theories*, Reidel, Dordrecht, 177-201.

Wehrli, E. 1992. "The IPS System" in *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, Nantes, 870-874.