

Syntax Coordination: Interaction of Discourse and Extrapositions*

Susanne Kronenberg, Franz Kummert
Technical Faculty, University of Bielefeld
email: susanne@techfak.uni-bielefeld.de

ABSTRACT

The model presented here for parsing spoken German offers a way to proceed incrementally discontinuous constructions. Typical constructions in task oriented dialogs are *Extrapositions to the right*. These are defined as syntactical constructions where a constituent is extraposed in the *Nachfeld*. Basing on the assumption that the extraposed constituent is not part of the source sentence the model works by coordinating the syntactical information given by the source sentence and the extraposed constituent to complete the extraposed constituent to a whole sentence. Therefore, the standard LR(1)-parser is extended by two additional actions. The parsing strategy works by deriving the first part of the construction until the extraposed constituent is reached. The both new actions enable the parser to proceed further by taking the syntactical information of the source sentence to complete the extraposed constituent to a sentence of its own. The repeated use of these actions guarantees that every intended reading will be performed.

1 INTRODUCTION

The paper will present a model for parsing incrementally discontinuous constructions of spoken German. Typical constructions in task oriented dialogs are *Extrapositions to the right*, where constituents, like VP-complements, VP-adjuncts or NP-modifiers are extraposed to the right over the sentence parenthesis. The problem by parsing this kind of constructions are the crossed dependencies which can appear between the extraposed constituent and not-extraposed constituents. The model presented here proceeds on the assumption that the extraposed constituent is not part of the source sentence but that it is a sentential structure of its own, following an idea from Shieber, Pereira and

Dalrymple on the analysis of ellipsis [DSP91]. Out of this fact the extraposed constituent is not modelled by the sentence grammar and is therefore part of the discourse analysis. For deriving a complete sentence out of it the parallelism between the source sentence and the extraposed constituent is used. For instance in the sentence '*Du schraubst den Würfel an mit der Schraube.*' (You fixed the cube with the screw.) the extraposed prepositional phrase '*mit der Schraube*' is considered to be a sentential like construction. Using the source sentence '*Du schraubst den Würfel an.*' the extraposed prepositional phrase will be completed to a sentence by inserting the prepositional phrase in the source sentence structure.

The model bases on a LR(1)-parser and an unification grammar. The regularities where an extraposition can appear in an utterance are extracted by an analysis of a corpus of task oriented dialogs [SER94], which is acquired in the special research area 'Situerte Künstliche Kommunikatoren'. On this foundation the LR(1)-parser is extended by a two additional actions called *insert* and *exchange*. This means that the already parsed sentence structure is taken and the extraposed constituent is inserted in this structure by the grammar rules to derive the intended sentence. The kind of action which will be used depends on the kind of extraposition construction (see section 2).

2 EXTRAPOSITION TO THE RIGHT

In written German the *Nachfeld* can only contain an extraposed subordinate clause, but never nominal arguments or adjuncts [RJ94]. In spoken German it is well accepted that verb arguments or adjuncts are extraposed out of the *Vor-* or *Mittelfeld* into the *Nachfeld*. This word-order variation is called *Extraposition to the right*. Therefore, a parser for spoken German has to be expanded to model not only typical word-order variations given in written German but also to

*This work has been supported by the German Research Foundation (DFG) within SFB 360.

derive the set of sentences which is defined through *Extrapolations to the right*. The following examples are out of a corpus of task oriented dialogs [SER94]. In these dialogs *Extrapolations to the right*¹ appear much more frequently than extrapolations to the left.

Example 1:

Du hast doch eben einen Würfel angeschraubt *mit der Schraube*.

You have as you know just a cube tighten with a screw.

You just have as you know tighten a cube with a screw.

Example 2:

Und die machst du jetzt mal zusammen *aufeinander so*.

And these do you now once together one on top of the other so.

And now you fix these on top of each other.

Example 3:

Das sieht aus *wie ein Flugzeug*.

This looks like an airplane.

Example 4:

Aus diesen wunderschönen Bauteilen soll zum Schluß erstmal ein Flugzeug dabei rauskommen, *so ein Propellerflugzeug*.

Out of this lovely components shall finally once an airplane thereby come out, such a propeller airplane.

Out of this lovely components shall finally come out an airplane, such a propeller airplane.

As seen in the examples above two different kinds of *Extrapolations to the right* exist: *Ausklammerungen* (exclamations) and *Nachträge* (supplements)² [SK98].

- An *Ausklammerung* construction is defined as a syntactical construction where a constituent appears in the *Nachfeld*. The grammatical function of this constituent is not already realized by any other constituent in the *Vor-* or *Mittelfeld* (see examples 1-3).
- A *Nachtrag* construction is defined as a syntactical construction where a constituent appears in the *Nachfeld*. But in contrast to *Ausklammerung* constructions a constituent exists in the *Vor-* or *Mittelfeld* with the same grammatical function like the one of the extrapolated constituent. This means that this constituent in the *Vor-* or *Mittelfeld* can be substituted by the extrapolated constituent in the *Nachfeld* (see example 4).

These different syntactical constructions cause different strategies for deriving these two kinds of extrapolation constructions.

3 THE PARSER

3.1 The Underlying Model

Like seen in examples 1-4 extrapolated constituents occur in the *Nachfeld*. The problem by incrementally parsing these kind of constructions is that the source constituent is already parsed long before the extrapolated constituent will be reached. Therefore, the standard LR(1)-parser has to be changed so that it is still possible to process the input string word by word for producing partial results as soon as possible. But additional to this the new readings of the already parsed sentences which will be induced by the extrapolated constituent have to be performed. This will be done by assuming that the extrapolated constituent is not part of the source sentence but that it is a sentential structure of its own, following an idea from Shieber, Pereira and Dalrymple on the analysis of ellipsis [DPS91]. Therefore, the extrapolated constituents will not be modelled by the sentence grammar but will be considered to be part of the discourse analysis. For deriving a complete sentence out of it the parallelism between the source sentence and the extrapolated constituent is used. This will be done by using the syntactical structure of the already parsed source sentence and the grammatical function defined through the extrapolated constituent. By the coordination of these both information the parser is enabled, to extend the extrapolated constituent to a complete sentence. For doing so, the parser has to be extended by two additional actions. Which of these will be used depends on the kind of extrapolated construction defined by the grammatical function of the extrapolated constituent and the syntactical information given by the source sentence (see chapter 3.2 and 3.3).

3.2 Ausklammerung

The problem by parsing *Ausklammerung* constructions is that crossed dependencies can occur between the extrapolated constituent and the source constituent in the *Vor-* or *Mittelfeld*. The following sentence:

Du schraubst den Würfel zusammen *mit der Schraube*.

You fix the cube to the screw.

has three different readings (see Figure 1).

¹The extrapolated elements are printed cursive.

²The german expressions will be used because this is a german word-order phenomenon

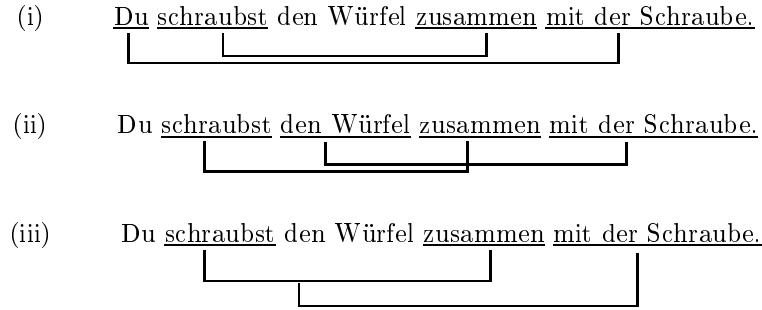


Figure 1: Crossed- and Nested-Dependencies in *Ausklammerung* constructions

The problem by deriving incrementally all possible readings is that the source constituents in (1.i) and (1.ii) are separated from the extraposed constituent by the verb phrase or the separable prefix. By deriving these readings on means of a standard LR(1)-parser the source constituents are already reduced with the verb to a verb phrase. No derivation of the intended reading is possible because it is not any longer possible to combine the source constituent with the extraposed constituent on foundation of the grammar rules.

Therefore, the LR(1)-parser will be extended by a third action called *insert*. The first part of the construction will still be parsed only by use of the standard *reduce* and *shift* action until the extraposed constituent is reached. As soon as the extraposed constituent is moved on the stack and is reduced as far as possible the next look ahead symbol which will be seen in the input string marks the point where the *insert* action has to be used. The *insert* action is associated with a grammar rule, like a reduce action. This tells the parser where the extraposed constituent has to be inserted. In the example above two different *insert* actions are possible. The one which is associated with the grammar rule to reduce the verb phrase and the extraposed prepositional phrase to a complex verb phrase to derive the reading of example (1.iii). The other which is associated with the grammar rule to reduce a noun phrase and a prepositional phrase to a complex noun phrase. This will be done twice in the sentence under consideration to derive the meaning of example (1.i) and (1.ii). All three uses of the *insert* action are performed by the parser so that every reading will be derived [Tom86]. The parsing results for the example sentence will be therefore:

1. *Du mit der Schraube schraubst den Würfel zusammen.*
You with the screw fix the cube.

2. *Du schraubst den Würfel mit der Schraube zusammen.*
You fix the cube with the screw.
3. *Du schraubst mit der Schraube den Würfel zusammen.*
You fix with the screw the cube.

3.3 Nachtrag

Extraposed constituents in *Nachtrag* constructions are those with the same grammatical function of one of the constituents already realized in the *Vor- or Mittelfeld*. As mentioned above the extraposed construction is considered to be a sentential like construction of its own which can be completed to a whole sentence by help of the syntactical information given by the source sentence. The source constituent can therefore be substituted by the extraposed constituent. Considering the sentence:

Das sieht wie ein Flugzeug aus *wie ein Propellerflugzeug.*
This looks like an airplane like a propeller airplane.

the definition of a *Nachtrag* construction causes two different sentences:

1. *Das sieht wie ein Flugzeug aus.*
This looks like an airplane.
2. *Das sieht wie ein Propellerflugzeug aus.*
This looks like a propeller airplane.

To derive both readings the LR(1)-parser is extended by a fourth action called *exchange*. This works by first deriving the source sentence until the extraposed constituent is reached. As soon as the extraposed constituent is shifted on the stack the next look ahead symbol tells the parser to use the *exchange* action. An *exchange* action is associated with a grammar

symbol which indicates the grammatical function of the constituent in the source sentence which has to be exchanged by the extraposed constituent. If more than one constituent in the source sentence has the same grammatical function associated with the *exchange* action the action will be repeated as long as every reading is produced. For the example above both intended readings (see (1) and (2)) are produced.

4 RESULTS

The extended LR(1)-parser presented here is tested on 25 sample sentences [SK98]. Every sentence contains one of the *Extraposition to the right* constructions introduced in section 2. These sentences are out of a corpus of task oriented dialogs [SER94]. Some sentences are simplified by leaving out reparations because these are not in focus of this model. This means that a sentence like "*Dann können sich doch äh diese Löcher überschneiden von der von dem, was unten liegt.*" (Then äh these holes could overlap from the from the one which lies below.) will be simplified to "*Dann können sich diese Löcher überscheiden von dem, was unten liegt.*" (Then these holes could overlap from the one which lies below). The length of the sentences is 10 words on the average. All sentences will be parsed correctly by tipped input. The parsing time is 23 msec on the average using a DIGITAL AlphaStation 500/400. For the moment every possible reading as presented above will be produced by the parser although not every reading is really intended. Additional corpus analysis shall give some criterions to grade the readings in a way that only the one preferred by the given context will be produced. It is intended to use not only speech context but also visual context for deciding which reading is induced in the given situation. Sentences without *Extrapositions to the right* are also correctly derived by the parser. So no over-generalizations will occur by introducing these both new actions into a standard LR(1)-parser.

5 CONCLUSION AND OUTLOOK

Extrapositions to the right are typical syntactical constructions in spoken German. These constructions induce long-distance-dependencies between the extraposed constituent and the source constituents in the *Vor-* or *Mittelfeld*. The model presented here offers a way to parse incrementally all possible readings which are induced by these long-distance-dependencies. This is done by assuming that the extraposed constituent is not a part of the source sentence structure but

that it is a sentential structure of its own. The completion of the extraposed constituent to a whole sentence is done by the discourse analysis. The syntactically information given by the source sentence and extraposed constituent is coordinated to receive the intended sentence. For deriving these kind of constructions the standard LR(1)-parser has to be extended by two additional actions. The actions enable the derivation of all possible readings by the repeated use of the same action or by competitive entries in the action table. The requirement of an incremental parsing strategy still holds by first producing the syntactically structure of the source sentence and additional to this the sentence structure of the extraposed construction. Therefore, this parsing strategy offers a way to process discontinuous extraposition constructions incrementally without any extensions of the underlying grammar.

Extraposition to the right constructions are often produced coordinate by both recipients of the dialog. Therefore, the end of a turn symbolises not any longer the end of these constructions. Further corpus analysis shall give information about how many *Extrapositions to the right* are accepted in one utterance to define the end of these syntactical entities. Context information have to be included to decide which of the possible readings of the extraposition construction is intended by the given situation.

References

- [DSP91] Mary Dalrymple, Stuart Shieber, and Fernando Pereira. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14:399–452, 1991.
- [RJ94] Owen Rambow and Aravind K. Joshi. A processing model for free word order language. In C. Clifton, Jr., L. Frazier, and K. Rayner, editors, *Perspective on Sentence Processing*. Lawrence Erlbaum Associates, 1994.
- [SER94] G. Sagerer, H.-J. Eikmeyer, and G. Rickheit. "Wir bauen jetzt also ein Flugzeug ...": Konstruieren im Dialog - Arbeitsmaterialien. Technical report, Universität Bielefeld, 1994.
- [SK98] K. Skuplik and W. Kindt. Ausklammerungskonstruktionen: Definition und empirische Untersuchung. Technical report, Universität Bielefeld, 1998.
- [Tom86] M. Tomita. *Efficient Parsing for Natural Language*. Kluwer Academic Publishers, Hingham, Ma., 1986.