An Implementation of the Argument Dependency Model

Wolfgang Menzel and Klaus Dalinghaus
Universität Hamburg, Department Informatik

Abstract

Human sentence processing proceeds in a left-to right incremental manner and is able to assign partial structural interpretations even if all the lexical items involved are not yet known. To mimic such a behaviour in a computational system is particularly difficult for head final languages where the rich information about the valency of the verb comes last. In such cases the Argument Dependency Model facilitates early hypotheses about the thematic status of a constituent by establishing direct argument-to-argument dependencies independent of the verb. An implementation of the model within the framework of Weighted Constraint Dependency Grammar is presented and evalated in a pseudo-incremental processing mode. Thanks to its non-monotonic nature the system can replicate the predictions of the Argument Dependency Model about the dynamic nature of revision processes during sentence comprehension with a very high degree of reliability.

Menschliche Sprachverarbeitung erfolgt inkrementell von links nach rechts und ist in der Lage, partielle Strukturhypothesen aufzustellen, auch wenn die beteiligten lexikalischen Elemente erst teilweise bekannt sind. Die maschinelle Nachbildung eines solchen Verhaltens ist besonders problematisch für kopffinale Sprachen, in denen die reichhaltige Valenzinformation des Verbs erst sehr spät verfügbar ist. In solchen Fällen gestattet das Argument Dependency Model frühzeitige Hypothesen über die thematischen Rollen einer Konstituente, indem es unabhängig vom Verb, direkte Abhängigkeitsbeziehungen zwischen seinen Argumenten aufbaut. Der Bericht stellt eine Implementation des Modells im Rahmen der Weighted Constraint Dependency Grammar vor und evaluiert diese in einem pseudo-inkrementellen Verarbeitungsmodus. Dank seiner nichtmonotonen Verarbeitungsweise kann das System die Vorhersagen des Argument Dependency Model über die dynamischen Revisionsprozesse beim Sprachverstehen mit einem sehr hohen Grad an Genauigkeit replizieren.

1 Introduction

Analogies between human cognitive faculties and the information processing capabilities of man-made artifacts has been a source of mutual inspiration, starting with the very early attempts to devise formalisms and machinery for formal information handling. Even if not explicitly expressed, both sides expect benefits from such a comparison. Although there is a large gap between the underlying processing principles, the hope persists that insights into functional aspects of the biological model might help to derive guidelines for system improvement. On the other hand, attempts to explain human behavior, eventually require to formulate them in terms of formal models, which then can be checked against real data by means of the currently available technology.

One of the most obvious characteristics of human language understanding is its incremental nature. Humans are processing sentences in a left to right manner making choices as early as possible always risking the necessity of later revisions which might become necessary to accommodate the subsequent input information. Indeed the vast majority of psycholinguistic insights into the mechanisms of the human language faculty relate to the time course of language processing. In such a situation any attempt to compare the performance of the human model with that of artificial systems beyond a simple introspective account, requires algorithmic solutions with comparable processing properties. Unfortunately, the lack of a truly incremental processing mode is one of the most serious drawbacks of current natural language processing technology.

The motivation to investigate prerequisites and possibilities for incremental processing of natural language utterances, however, reaches far beyond a purely cognitive perspective. Left-to-right processing is also an essential feature for all kinds of online processing tasks, which first of all are relevant in speech processing scenarios, like dialogs or dictation. Here, the speaking time becomes a precious resource, which needs to be used in order to maintain fluency and provide for a natural man-machine interaction.

A large number of issues arises if an algorithmic scheme for incremental sentence processing needs to be devised:

1. Which kind of mechanisms are necessary to gradually extend a parsing problem by successively incoming word forms?
2. Are the available heuristics reliable enough to take early decisions and, if necessary, to initiate revisions of intermediate results?
3. Which kind of heuristics can be applied to reuse as much as possible of the already available information in subsequent processing steps in order to facilitate efficient processing?
4. Can models of grammar, which have been developed for the non-incremental case be used for or ported to an incremental processing scheme?

This report investigates some of these problems in the context of a broad coverage parsing model implemented within the framework of Weighted Constraint Dependency Grammar (Schröder, 2002; Foth, 2006).

2 Incremental Sentence Processing

Incremental processing of an input stream is an inherently recursive process: After a certain portion of the input has been analysed, it is extended by additional elements (an increment) and subjected to the very same processing component. This process becomes non-monotonic as soon as the necessity arises to revise (parts of) the already computed output information according to the extended input. Two different approaches can be distinguished in such a case: a cautious one, which maintains all the different output alternatives until enough information is available to take a safe decision, and an eager one, which adopts the most plausible interpretation as soon as possible and does so even at the risk of later revisions becoming necessary. Obviously, human sentence processing takes the latter approach. This does not come as a surprise, because eager processing is connected with a number of cognitive advantages:

1. Taking early decisions provides for a better, hence, more economical allocation of cognitive resources.

2. Having a single preferred reading available early enough allows the hearer to derive strong expectations about upcoming observations. Maintaining a variety of pending hypotheses will inevitably result in very many different expectations, which not only are expensive to be matched to the actual input, but due to their diversity have little predictive power.

3. Having a unique interpretation available rather early brings the hearer into a favourable position to quickly react to the incoming information, an important prerequisite for efficient communication strategies.

Even among the eager approaches different temporal processing schemes can be distingushed. Usually, a piece of the structural description is built as soon as all its constituting parts are available. Unfortunately such a schema necessarily involves some kind of delay. While in case of a phrase structure grammar all daughter nodes of a rule to be applied need to be available, for a dependency model at least the two word forms to be connected by a dependency relation must be already known. Even if incomplete partial structures are hypothesized on a sub-rule level, e.g. by means of chart parsing techniques, a final decision on their appropriateness can only be taken after the complete constituent has been established.

This situation is particularly problematic in head final languages, since crucial information about valences and valence requirements, usually contributed by the head (e.g. the verb), comes last. Therefore, reliable hypotheses about the distribution of functional roles can only be established at a very late point in time. This again is in stark contrast to human sentence processing where hypotheses e.g. about the thematic status of a constituent are made very early and independent of the verb and its lexical information and are revised later on if verb specific information like government or agreement requirements have to be accomodated.

Recently, large coverage models of natural language syntax became available, which are able to process sentences with almost no failure and a fairly high accuracy. One of them, MaltParser Nivre et al. (2006), even adopts a strictly incremental processing scheme. Here, a stack serves as an intermediate storage device able to hold unattached word forms until an appropriate attachment point becomes available. For each incoming word form a support vector machine operating on features of the parse history deterministically decides which one of the following operations is applied:

- shift the incoming word form on the stack, i.e. wait
- attach the word form on the stack to the incoming one and reduce the stack
- attach the incoming word form to the one on the stack and shift it on the stack
- reduce the stack

Although the decision is fully deterministic, i.e. no revisions are ever considered, the parser achieves state-of-the-art performance for a number of languages. Still, it suffers from the delay which is incurred by shifting input elements onto the stack temporarily. Changing this behaviour to an earlier commitment would require an additional source of information about role attachments independent of the verb and a non-deterministic mechanism for revising decisions if necessary.

A framework which actually does this is the Argument Dependency Model (ADM), a dual pathway approach of sentence comprehension. It is based on psychological evidence for both, the syntactic structure of a sentence and its thematic interpretation being created by a human in an incremental way. In order to accomplish this, it establishes a hierarchy of direct complement-to-complement relationships which are fully independent of the valence requirements of individual verbs. This approach therefore not only makes verb-independent information available for disambiguation, but also facilitates an early assignment of a functional role to a certain part of the sentence even if the governing verb is not yet known. The model has been successfully used to explain EEG-data obtained from humans while listening to German subordinate clauses (Bornkessel, 2002). To investigate whether the ADM might also help to facilitate early decisions and revision processes in an computational parser we integrated it into an existing broad coverage WCDG for German and evaluated it on the very same utterances, which also have been used in the psycholinguistic experiments.

3 The Argument Dependency Model

Psycholinguistic evidence gives rise to the expectation that the degree of meaning derived from a given sentence during incremental comprehension is a function of the morphological informativeness of the sentential arguments (Schlesewsky and Bornkessel, 2004). This can be explained by a model of language comprehension which incorporates two different pathways of processing: a syntactic and a thematic one. Such a model is the Argument Dependency Model (ADM) (Bornkessel et al., 2005) which is based on the language comprehension model of Friederici (1999, 2002). Which of the two pathways is chosen depends on the morphological case marking borne by the incoming arguments (Bornkessel, 2002): The thematic pathway is activated by an unambiguously case marked argument whereas the syntactic pathway is chosen in the other case. The choice of a pathway does not mean that the other pathway is cut off but that the chosen pathway leads the comprehension.

Thematic information provides a general conceptual specification of the relations between the arguments of a sentence and between the arguments and the verb. Hence, the thematic pathway generalizes over a number of verbs by being based on so called proto-roles. These proto-roles are introduced as Proto-Agent, Proto-Patient, and Proto-Recipient by Primus (1999) and Dowty (1991). Another kind of classifying these proto-roles as Actor and Undergoer is given by van Valin Jr. and LaPolla (1997). Proto-roles as defined by Primus (1999) are dependent on each other and, thus, can be arranged according to a thematic hierarchy:

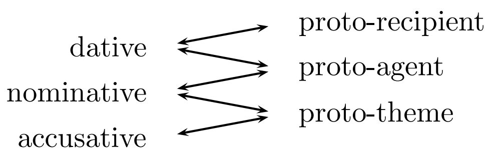
$$\text{Proto-Agent} <_{\Theta} \text{Proto-Recipient} <_{\Theta} \text{Proto-Patient} \tag{1}$$

The ideal Proto-Agent is nominative and in control (animate). It may not depend on any other argument ($-$DEP). In contrast, a Proto-Patient (+DEP) always depends on either a Proto-Agent or a Proto-Recipient. The Proto-Recipient, finally, can be either depending, dependent, or both at the same time ($\pm$DEP).

This thematic hierarchy can be mapped to a complementing case hierarchy:

$$\text{nominative/absolutive} < \text{accusative/ergative} < \text{dative} < \text{other oblique cases} \tag{2}$$

according to the following many-to-many relationship



Nominative is an ideal proto-agent, since it has full control.

Ich zerbrach die Vase.

Dative can be proto-agent, but has no full control

Mir zerbrach die Vase.
**Mir zerbrach die Vase absichtlich.*

whereas accusative can never be proto-agent

**Mich zerbrach die Vase.*

Dependency relationships between the constituents of a sentence are established by means of a number of principles which either operate on the thematic pathway or the syntactic one:

Principle	**Thematic pathway**	**Syntactic pathway**
ECONOMY	The first argument is assigned the status [−dep] if at all possible.	An argument is associated with the grammatical function compatible with the least number of syntactic dependencies.
DEPENDENCY	For any two arguments A and B, either A must hierarchically dominate B in terms of thematic status, or B must dominate A.	For any two arguments A and B, either A must c-command B, or B must c-command A.
DISTINCTNESS	For relations consisting of ≥ 2 arguments, each argument must be maximally distinct from every other argument in terms of thematic status.	For any two arguments A and B, either A must asymmetrically c-command B. or B must asymmetrically c-command A.

These principles can be refined for individual languages, e.g. for German

MAPPING: if no verb-specific information contradicts, nominative (+ANIMATE) receives −DEP
BLOCKING: accusative is not compatible with −DEP

They are complemented by the usual verb specific requirements (VERB-FIT), e.g.

thematic: object-experiencer verbs require that the dative-marked argument must dominate the nominative-marked one
syntactic: subject-predicate agreement must be obeyed

Finally, a general preference for argument role attachment over adjunct attachments is assumed.

Most of these principles are non-monotonic ones. They assign default values (e.g. ECONOMY) or require to determine an optimum between different alternatives (e.g. DISTINCTNESS). It is this non-monotonicity which provides an explanation for the re-interpretation processes which can be observed with human sentence processing.

In the following example the role assignment solely relies on the thematic pathway. After the first constituent (*der Junge*) is encountered, it is assigned the topmost rank in the hierarchy (−DEP). Consequently, the second one receives +DEP according to the distinctness principle. No reanalysis occurs, since the assignment is compatible with the specific requirements of the verb.

..., dass	*der Junge*	*den Großvater*	*besucht.*	
	−DEP			ECONOMY
		+DEP		DISTINCTNESS
				VERB-FIT

In the second example, a reanalysis is necessary on the second constituent, because its nominative case ultimately requires −DEP and therefore the first constituent needs to be reanalysed as +DEP. The object-experiencer verb finally triggers yet another re-interpretation since it forces the dative to dominate the nominative.

..., dass	*dem Jungen*	*der Film*	*gefällt.*	
	−DEP			ECONOMY
		−DEP		MAPPING
	+DEP			DISTINCTNESS
	−DEP			VERB-FIT
		+DEP		DISTINCTNESS

A reanalysis is also necessary with ternary verbs, where the maximally distinct assignment on the dative NP needs to be corrected as soon as the accusative becomes available.

... hat	*der Mann*	*dem Jungen*	*einen Klaps*	*gegeben*	
	−DEP				ECONOMY
		+DEP			DISTINCTNESS
			+DEP		BLOCKING
		±DEP			DISTINCTNESS
					VERB-FIT

If in case of ambiguous case assignments the syntactic pathway is activated, the only reason for a reanalysis can be the agreement requirements of the verb.

... , dass	*Maria*	*Geigerinnen*	*zuhörten*	
	SUBJ			ECONOMY
		OBJA		DISTINCTNESS
	OBJD	SUBJ		VERB-FIT

If however the second NP bears an unambiguous case marking the processing switches from the syntactic pathway to the default case of thematic processing

… hat	*die Geigerin*	*der Direktor*	*gesucht*	
	SUBJ			ECONOMY
		?		MAPPING vs. ECONOMY
		−DEP		MAPPING wins
	+DEP			DISTINCTNESS
				VERB-FIT

4 Incremental Processing with CDG

A preliminary investigation using a Weighted Constraint Dependency Grammar (WCDG) has shown that in principle the model can be used in an incremental processing mode (Foth et al., 2000a). This study, however has been carried out using a fairly restricted model for a specific type of utterances (namely simplified utterances from a appointment negotiation domain). Moreover, similar to MaltParser (Nivre et al., 2006), this system did not implement a truly incremental processing scheme but had to wait instead until a suitable attachment point became available in the right context. A partial parsing scheme was used to deal with the incomplete nature of a sentence during incremental analysis. Relaxing the constraints on the possible top-most nodes of a dependency tree, arbitrary categories can take this role, thus allowing the tree to break into fragments if no better attachment was available. This in some cases led to unintuitive weight assignments in the grammar.

It has been shown that the parser was indeed robust enough to deal with many instances of sentence prefixes, i.e. incomplete utterances. Moreover, its ability to supply information about constraint violations in addition to structural hypotheses has turned out to be an extremely valuable feature in the incremental case: Constraint violations for an incomplete utterance reflect expectations for the not yet observed input and therefore have the potential to guide further processing.

Trying to adopt this approach in a broad coverage grammar, however, seems to be not as easy. At least two aspects contribute to these difficulties:

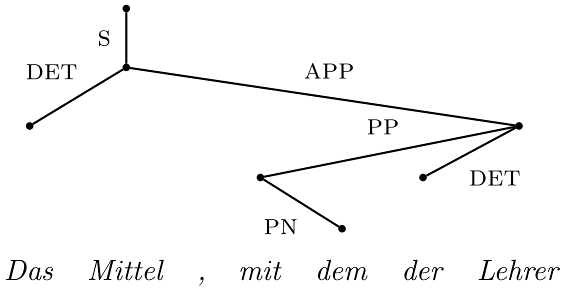
- The parser enforces attachment decisions for all the word forms in the utterance. This might easily lead to meaningless attachments as long as the sentence is not yet complete, because a suitable attachment point is not yet available (see figure 1).

  Therefore a processing mechanism is required which explicitly reflects the (possibly) incomplete nature of the incoming utterance.

- Considering the usually very large variety of possible continuations for a sentence prefix, many constraints have to be relaxed in order to tolerate missing information. Such a neutralization of syntactic constraints might be possible in a restricted domain, but has disastrous effects with a broad coverage grammar, where the restrictive potential of grammar constraints is already fairly low, because a large number of alternatives has to be accommodated. Somehow, this loss of constraining information needs to be compensated for.

In principle, WCDG is an approach which is compatible with the requirements of an eager processing mode:

- Irrespective of the input being partial or complete, WCDG is able to determine the optimal structural interpretation.

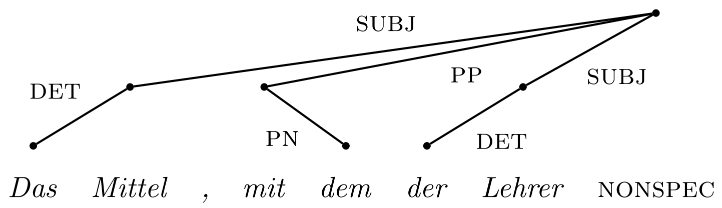


constraint	weight	from	to
Fragment	5.000e-02	Mittel	NIL
Antezedens fehlt	1.100e-01	dem	mit
PP vor Nomen	9.100e-01	mit	Lehrer
mod-Distanz	9.524e-01	Lehrer	Mittel
mod	9.950e-01	mit	Lehrer
mod	9.950e-01	Lehrer	Mittel
neut	9.998e-01	Mittel	NIL
direction	9.999e-01	mit	Lehrer

Figure 1: Structural interpretation of an incomplete sentence without NONSPEC-modeling and the constraints it violates. The additional level for the reference of the relative pronoun has been suppressed.

- Dependency structures in general are robust against fragmentation: Any part of a dependency tree is valid dependency tree again. WCDG inherits this property.
- By evaluating individual dependency relations (or pairs of them) the scope of constraints is fine-grained enough to support local decisions, i.e. the attachment of a right complement can be established without waiting for possible other complements not yet observed.
- Constraint violations incurred by a partial analysis might well be interpreted as expectations about the upcoming right context.

From a procedural point of view, differently sophisticated algorithms for incremental processing can be devised. They distinguish themselves by the amount of information which is passed between two subsequent calls on an incrementally extended input stream. The most simple one is prefix parsing, where for each new increment, the complete input sequence available so far is subjected to the parser again. No information about the results of a preceding call is maintained or passed to subsequent processing steps. Although being algorithmically simple, this approach is not very interesting neither from a computational nor from a cognitive point of view. It renders one of the most attractive features of incremental processing void, namely the possibility to use listening time as processing time, a major prerequisite for the near linear time behaviour of human sentence processing. Prefix parsing instead invests the same amount of time into the processing of the last incremental extension as a non-incremental procedure would do. To improve on this, more information about the hypothesis space of the preceding cycle needs to be made available. A series of increasingly permissive structural heuristics



constraint	weight	from	to
Antezedens fehlt	1.100e-01	dem	mit
Kasus Dispräferenz Dativ	9.400e-01	mit	NONSPEC
NONSPEC-Kante	9.700e-01	Mittel	NONSPEC
NONSPEC-Kante	9.700e-01	mit	NONSPEC
NONSPEC-Kante	9.700e-01	Lehrer	NONSPEC
Komplementdistanz	9.940e-01	Mittel	NONSPEC
Komplementdistanz	9.960e-01	mit	NONSPEC

Figure 2: Structural interpretation of an incomplete sentence with NONSPEC-modeling and the constraints it violates

(Foth et al. (2000a)) could then be applied to keep some of the alternative interpretations as possible revision candidates while others are frozen to the currently optimal decision:

1. Keep only the dependency edges of the optimal parse tree.
2. Additionally keep all dependency edges which differ from the optimal parse with respect to their lexical readings.
3. Additionally keep all dependency edges which differ from the optimal parse in the subordination of the last word form.
4. Additionally keep all dependency edges which differ from the optimal parse in the subordination of all word forms on the path from the last word form to the root node of the tree.
5. Additionally keep all dependency edges which differ from the optimal parse in the subordination of all word forms on the path from the last word form to the root node of all partial trees of the preceding solution.

Foth et al. (2000a) compared these different heuristics and found a tradeoff between speedup and quality. Note, however, that the comparison in Foth et al. (2000a) has been made using a small scale grammar and a combinatorial search procedure, which builds dependency trees by successively adding more edges. Such an approach turned out to be infeasible if applied to large scale grammars as used in the experiments reported here. For them a repair approach has been more successful, which sucessively transforms parts of a structure in order to remove the most severe constraint violations (Foth et al. (2000b)). At the same time, the transformational approach also lends itself for a completely different kind of informational coupling between subsequent incremental cycles: Instead of (or in addition to) *restricting* the remaining search space the currently available optimal structure for a sentence prefix can be used to *initialize* or even *guide* the transformation procedure in the subsequent step.

Using the optimal structure of the current step as a starting point for the next one was the approach taken in the experiments reported here. It provides a fairly weak coupling between subsequent incremental cycles and therefore no impressive results in terms of its temporal characteristics could be expected. It is meant, however, as a first attempt to study possibilities of incremental processing with a large scale WCDG under the conditions of a truly eager processing regime.

In order to facilitate eager processing, the basic solution procedure of WCDG had to be extended to be able to also consider the not yet observed right context of the currently available sentence prefix. If not, the parser is forced to subordinate everything under an existing word form, or to break the tree into several fragmements. This was the approach followed in Foth et al. (2000a). The other extreme would consist in providing a set of virtual nodes which could be filled later on with the lexical information of the incoming word forms.

Two problems, however, have to be solved for such a model to become feasible:

1. How many of these additional nodes are actually necessary?
2. Are there sufficient conditions under which two of them can be identified as referring to the same input word form?

Because there are no straightforward heuristics to answer these questions, we adopted a compromise: A single node (called NONSPEC) is used to approximately model the unseen right context and all word forms requiring a modifiee to the right are attached to it. Its approximative nature results from the fact that it merely serves as a projection plane for arbitrary right-bound attachment expectations and it remains unclear whether this node actually represents one or several different word forms. Moreover, is has no lexical properties and none can be projected onto it. Thus, its only purpose consists in avoiding meaningless attachments and a further propagation of their consequences across the already seen sentence prefix (Figure 2).

The NONSPEC node competes with other nodes in the parsing problem in the ususal manner: Attachments are established, scored and possibly modified, if they turn out to be the source of a major constraint violation. From the perspective of an external observer only the consequences of this competition become visible as a kind of reanalysis effect, since the optimum might switch to an alternative structure as the scores of the individual dependency edges develop. Internally it is just an accumulation of constraint violations at a certain point in the structure. i.e. a reranking. Of course, the final optimal structure of a sentence should not include a NONSPEC node. This, however, is not directly enforced by the parser.

5 Implementation of the ADP Model in CDG

The following description focusses on the core constraints needed for integrating the Argument Dependency Model into the rule set of a WCDG. Since NONSPEC neither represents the identity of a word form nor does it contain any lexical information, all constraints of the grammar have to be adapted accordingly, in order to avoid unwarranted conclusions or access to non-existing information. Here, it is assumed that the original grammar as used e.g. in (Foth, 2006; Foth and Menzel, 2006) has already been adapted accordingly. This, in particular, requires to model the unknown right context by means of the NONSPEC-node and preventing all right-bound constraints from access to non-existing lexical information. These additional extensions are not described here.

While in Section 3 the Argument Dependency Model has been introduced by means of a feature assignment mechanism, its implementation in WCDG is based on true argument-to-argument relationships which directly represent the corresponding dependencies as postulated by the model. For this purpose, an additional level ARG has been introduced. It is used to build a chain of argument dependencies which strictly obeys the hierachical ordering conditions of the DEPENDENCY principle by placing arguments higher up in the chain if they exhibit fewer dependencies than others. See Figure 3 for an example.

The linear structure of this dependency chain is ensured by two constraints which rule out branchings (no two arguments can share the same attachment point: `X!ARG/\Y!ARG`) and bidirectional attachments (if two edges model an attachment sequence the topmost node must be different from the lowest one: `X!ARG/Y!ARG` and `X^id != Y@id`).

```
// Zwei ARG-Kanten dürfen sich bei einem Wort nicht treffen.
{X!ARG/\Y!ARG} : 'ARG-Kollision' : arg : 0.5 :
  false;

// Zwischen zwei Argumenten kann es nur eine Kante geben.
{X!ARG/Y!ARG} : 'bidirektionale ARG-Kante' : arg : 0.5 :
  X^id != Y@id;
```

Here, the constraint header `{X!ARG/\Y!ARG\}` refers to two edges on the ARG-Level sharing the same attachment which is not the root of the tree. Expressions like `X^id` and `X@id` give access to the features of a word form at the upper and the lower node of an edge, respectively. In this particular case they return their positional indices `id` in the sentence. For more details on the WCDG constraint language refer to (Foth et al., 2003).

The hierarchy of dependencies is established by a constraint which has access to a corresponding WCDG-subsumtion hierarchy

```
// Die ARG-Kante, die unterhalb einer anderen liegt muss auch ein
// Label tragen , dass in der Argument-Hierarchie unter dem der ersten Kante liegt.
{X!ARG/Y!ARG} : 'Hierarchie nicht beachtet' : arg : 0.5 :
  subsumes(ArgLabel, X.label, Y.label)
  & Y.label != X.label;
```

where the subsumtion of argument dependency relationships is defined as follows

```
// Hierarchie für die ARG-Label
ArgLabel ->
  '[-dep]' -> '[+-dep]',
  '[+-dep]' -> '[+dep]',
  '[+dep]' -> '';
```

Only nominals (common nouns, proper noun, numbers, personal, relative, or demonstrative pronouns, etc.) are considered as arguments (in case of a conjunction its leftmost element) and arguments need to be distinguished from non-arguments.

```
// Nominale sind Ausgangspunkt einer ARG-Kante.
{X:ARG\/Y:SYN} : 'ARG-Definition' : arg : 0.0 :
```

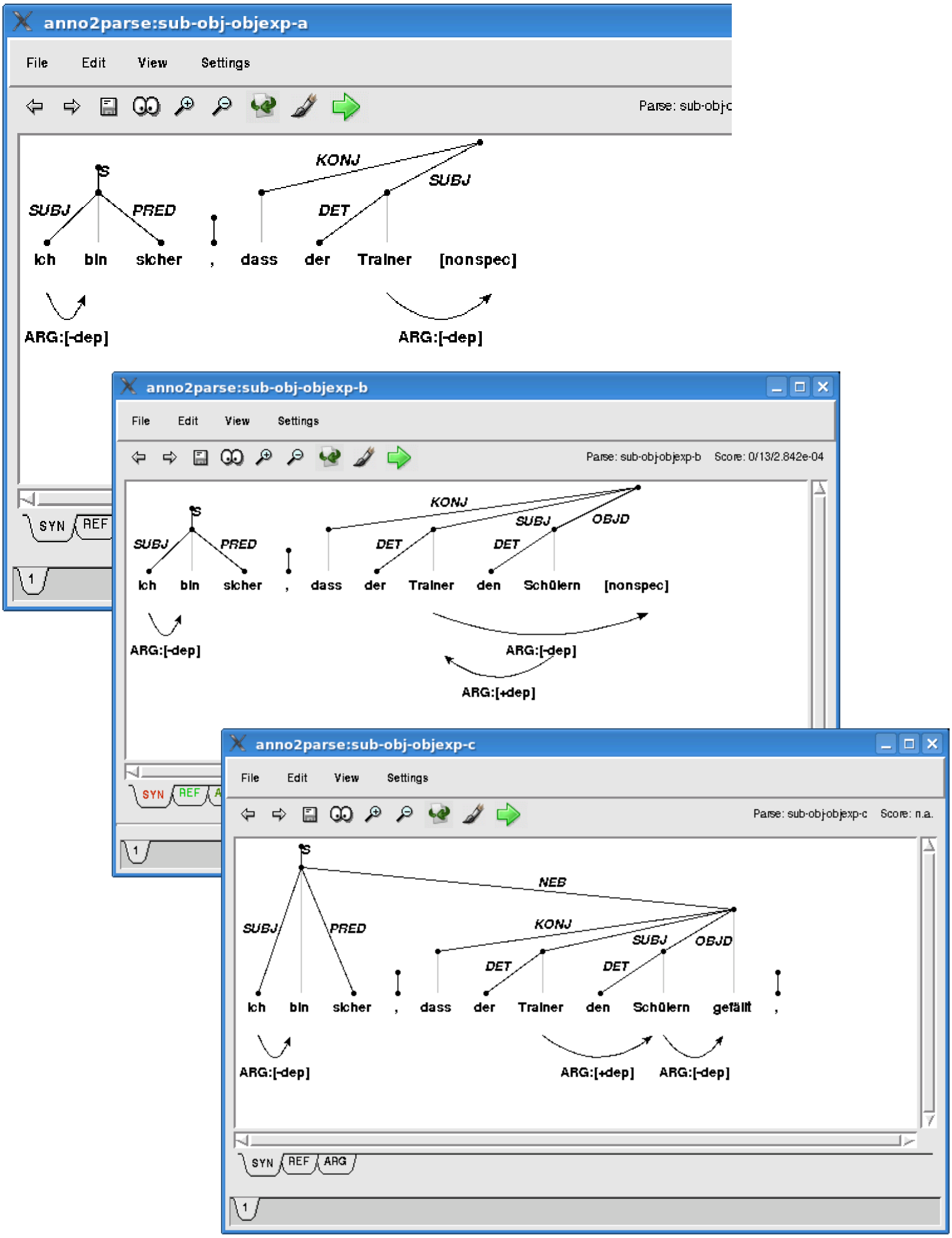


Figure 3: Left-to-right incremental processing of a German subordinate clause. Note the early assignment of a thematic interpretation on the ARG-level which is revised after receiving the verb.

```
  ~ isa(X@,Nominal)
  // CJ werden ausgenommen, da die gesamte Konjunktion das Argument ist.
  | Y.label = CJ
  // APP werden ausgenommen, da sie ein Argument modifizieren.
  | Y.label = APP
  // GMOD werden ausgenommen, da sie ein Argument modifizieren.
  | Y.label = GMOD
  // TODO: PN dürfen nur Argument sein, wenn sie dem Verb zugeordnet werden.
  | Y.label = PN
  <-> root(X^id);
```

All other word forms are attached to the artificial root node NIL with the empty label.

```
// ARG-Kanten nach NIL haben das leere Label
{X|ARG} : 'NIL-Definition (ARG)' : arg : 0.0 :
  X.label = '';

// Alle Kanten mit dem leeren Label gehen nach NIL
{X!ARG} : 'NIL-Label (ARG)' : arg : 0.0 :
  X.label != '';
```

The chain of argument-to-argument dependencies is anchored eventually at the finite verb of the clause. This relationship must carry the label −DEP:

```
// Kante zum finiten Verb heißt [-dep]
{X!ARG\/Y:SYN} : 'Verbindung zum Verb muss [-dep] sein' : arg : 0.5 :
  isa(X^,finit) <->
   X.label = '[-dep]';
```

Consequently, only nominals and the finite verb can serve as a potential attachment point for argument dependencies:

```
// Nur Nominale und finite Verben kommen als Zielpunkt einer ARG-Kante in Frage.
// eine ARG Kante darf nicht bei einem PN ankommen
{X!ARG\Y:SYN} : 'ARG-Kategorie' : arg : 0.0 :
  (isa(X^,Nominal)
  // TODO PN (siehe oben)
  & Y.label != CJ
  & Y.label != APP
  & Y.label != GMOD
  & Y.label != PN) |
  isa(X^,finit);
```

For this and most of the subsequent constraints to be valid one has to ensure also that the arguments under consideration and their governing verb belong to the same clause. At the moment, this is achived by means of a simple check for the presence of a comma. This is a corpus specific solution valid only for the particular test cases (c.f. Section 6.1).

The ECONOMY principle of the syntactic pathway is formulated by means of a constraint which has access to the label of the syntactic assignment (`Y.label = SUBJ`). Together with the above mentioned constraint that requires −DEP to attach to the verb it ensures the subject to be thematically independent of other arguments.

```
{X!ARG\/Y:SYN} : 'Verb fordert SUBJ als [-dep]' : arg : 0.5 :
  spec(X^id) & isa(X^,finit) & ~ exists(X^obj_experiencer) ->
    exists(X@case)
    & compatible(Features, nom, X@case)
    & Y.label = SUBJ;
```

The only exception are object experiencer verbs, which are explicitly marked in the dictionary and assign −DEP to the object with the case as specified there.

```
{X!ARG\/Y:SYN} : 'obj experiencer verb fordert Objekt als [-dep]' : arg : 0.5 :
  spec(X^id) & isa(X^,finit) & exists(X^obj_experiencer) ->
    exists(X@case)
    & compatible(Features, X^obj_experiencer, X@case)
    & edge(Y, Object_Experiencer);
```

Here, the value of the feature `obj_experiencer` at the verb has to be compatible with the case feature of the argument (`compatible(Features, X^obj_experiencer, X@case)`) and the corresponding syntactic edge must be of type `Object_Experiencer`, namely a dative object or an ethical dative. Note that both constraints explicitly require the verb to be already specified (`spec(X^id)`).

As long as the verb is still unknown, the thematic pathway is used to assign −DEP to a nominative by means of a preferential constraint

```
{X!ARG\/Y:SYN} : 'Preferenz [-dep] für nom' : arg : 0.97 :
  edge(Y, nom_Label) ->
    X.label = '[-dep]';
```

Although being rather weak, this preference is still stronger than the preference for a dative case being −DEP

```
{X!ARG\/Y:SYN} : 'Preferenz [-dep] für dat' : arg : 0.99 :
  edge(Y, dat_Label) ->
    X.label = '[-dep]';
```

Accusative case, on the other hand, has a preference for +DEP

```
{X!ARG\/Y:SYN} : 'Preferenz [+dep]' : arg : 0.99 :
  edge(Y, acc_Label) | edge(Y, gen_Label) ->
    X.label = '[+dep]';
```

Finally, DISTINCTNESS is modelled by means of another constraint, which refers to a more technical subsumtion hierarchy (`ArgLabelBig`), where instances of label pairs which satify the principle are directly enumerated

```
{X!ARG/Y!ARG} : 'Rollen sollten maximal auseinander liegen' : arg : 0.99 :
  subsumes(ArgLabelBig, X.label, Y.label)
  & Y.label != X.label;
```

with `ArgLabelBig` being defined as

```
ArgLabelBig ->
  '[-dep]' -> '[+dep]',
  '[+-dep]' -> '[+dep]',
  '[+dep]' -> '';
```

6 Parsing performance

One of the major benefits of the Argument Dependency Model is its ability to very early assign a proto-semantic interpretation to parts of a sentence as it unfolds over time without necessarily having access to the verb-related information which in verb final languages comes late. This advantage, however, requires to take early decisions based on perhaps incomplete information with the necessity to revise them later on if more evidence becomes available. These revision processes have been studied extensively in a number of psycholinguistic experiments (Bornkessel, 2002).

Having a computational version of the model available the natural question arises, whether similar effects to the ones found with human listeners can also be observed in the behaviour of a computational parser. Such an investigation can be carried out with respect to

- the quality of parsing decisions in particular for the intermediate processing steps, and
- resource requirements of the parser at different points in time while processing the sentence incrementally from left to right.

Using the predictions of the model as a gold standard a quality centered evaluation can be conducted in a rather straightforward manner. Finding relationships between reanalysis processes and their resource requirements is more difficult, because it requires to identify suitable formal parameters which could be taken as indicators of parsing effort.

Because of the underlying processing paradigms in the natural and computational system being completely different ones, no easy mapping between processing time requirements can be established. Instead, computational psycholinguistics has used probabilities as a linking hypothesis, based on the common observation that frequent items are processed more rapidly (Keller, 2003) and low probabilities can signal processing difficulties. Nevertheless, for the purpose of this investigation we will simply use CPU time as a measure of resource requirements leaving open the question whether other parameters of the WCDG parser might provide a better foundation to draw conclusions about its cognitive plausibility from.

6.1 Data

To facilitate a direct comparison with the available findings about the psychological adequacy of the Argument Dependency Model we used the same set of sentences, which have been compiled as test stimuli for the experiments with human subjects (Bornkessel, 2002).

The corpus comprises sentences according to a uniform sentence pattern consisting of a verb final target subclause preceded by its matrix clause and followed by a second subclause.

A	S-O	active	ambiguous	sg-pl	*... dass Christian Pastorinnen abrät*
A’	S-O	active	ambiguous	pl-sg	*... dass Pastorinnen Christian abrät*
B	O-S	active	ambiguous	sg-pl	*... dass Christian Pastorinnen abraten*
B’	O-S	active	ambiguous	pl-sg	*... dass Pastorinnen Christian abraten*
C	S-O	object-exp	ambiguous	sg-pl	*... dass Christian Pastorinnen behagt*
C’	S-O	object-exp	ambiguous	pl-sg	*... dass Pastorinnen Christian behagt*
D	O-S	object-exp	ambiguous	sg-pl	*... dass Christian Pastorinnen behagen*
D’	O-S	object-exp	ambiguous	pl-sg	*... dass Pastorinnen Christian behagen*
E	S-O	active	unamb.	sg-pl	*... dass der Betrüger den Winzern abrät*
E’	S-O	active	unamb.	pl-sg	*... dass die Winzer dem Betrüger abraten*
F	O-S	active	unamb.	sg-pl	*... dass dem Betrüger die Winzer abraten*
F’	O-S	active	unamb.	pl-sg	*... dass den Winzern der Betrüger abrät*
G	S-O	object-exp	unamb.	sg-pl	*... dass der Betrüger den Winzern behagt*
G’	S-O	object-exp	unamb.	pl-sg	*... dass die Winzer dem Betrüger behagen*
H	O-S	object-exp	unamb.	sg-pl	*... dass dem Betrüger die Winzer behagen*
H’	O-S	object-exp	unamb.	pl-sg	*... dass den Winzern der Betrüger behagt*

Table 1: Sentence patterns of the test corpus

matrix clause	target subclause			final subclause
Gestern wurde gesagt, dass die	*Winzer*	*dem Betrüger*	*abraten,*	*obwohl das nicht wahr ist.*
	NP1 ↑	NP2 ↑	V ↑	

The relevant points in time, where decisions about the thematic status of constituents have to be taken or possibly revised are marked by arrows.

The target subclause is modified into 16 different test conditions, along the following four parameters (Table 1)

- constituent order: subject before object vs. object before subject
- verb type: active (*abraten*) vs. object-experiencer (*behagen*) [1]
- case marking: ambiguous (Hans, Bäuerinnen, ...) vs. unambiguous (dem Betrüger)
- morphological variation: singular before plural vs. plural before singular (with a possible inflectional adaptation of the verb form to ensure subject-verb agreement).

For each of these 16 conditions 80 sentences have been generated yielding a corpus consisting of 1280 sentences, which can be used to systematically study the different cases of reanalysis processes. Basically, two different kinds have to be considered here: syntactic and thematic revisions, where a syntactic reanalysis can only appear in case the first argument carries an ambiguous case marking (conditions A-D/A’-D’):

[1] Note that only dative objects are considered.

		syntactic reanalysis	
		no	yes
thematic	no	A, B'	C', D
reanalysis	yes	C, D'	A', B

If in any of these conditions a reanalysis becomes necessary it occurs during the integration of the verb information. A more fine grained investigation of the time course of revision processes can be conducted using conditions E-H/E'-H', where an early revision (at the second NP can be distinguished from a late one at the verb.

		early revision	
		no	yes
late	no	E, E'	F, F'
revision	yes	G, G'	H, H'

6.2 Experiments

Four different research questions have guided the experimental work carried out in this study:

- Does the pseudo-incremental parsing scheme, which uses the current optimal structure as a starting point for the processing of the incrementally extended sentence, provide any significant improvement over a simple parsing of incrementally extended sentence prefixes?
- Do the different reanalysis processes have an impact on the resource requirements of the parser?
- Do the parsing decisions on incomplete sentences comply with the predictions of the Argument Dependency Model, i.e. can the reanalysis effects be observed in the intermediate parser output?
- How sensitive is the output quality to a modification of fundamental modelling assumptions.

To study these issues in more detail three different experiments have been carried out

Experiment 1 : Prefix parsing on the complete corpus as baseline

Experiment 2 : Pseudo-incremental parsing on the complete corpus

Experiment 3 : Pseudo-incremental parsing with a modified grammar on a reduced subcorpus of 800 sentences.

Results for quality and time consumption under the different experimental conditions are given in Table 2. All sentences have been presented in three incremental steps: The sentence up to the point after the first NP, the second NP and the verb. The final subclause which was needed to factor out sentence final integration effects in the psycholinguistic experiments has not been considered here.

Note that many entries in Table 2 are duplicates, since their corresponding sentence prefixes coincide. In particular, this holds for the following conditions:

	Experiment 1 prefix parsing						Experiment 2 pseudo-incr. p.			Experiment 3 modified grammar					
	processing time			sentence acc.			processing time			processing time			sentence acc.		
	NP1	NP2	V	NP1	NP2	V	NP1	NP2	V	NP1	NP2	V	NP1	NP2	V
A	0.66	3.80	5.94	100	96	100	0.66	3.21	6.75	0.83	3.57	13.40	100	100	100
A'	0.55	3.73	8.75	100	86	89	0.55	3.20	9.16	0.68	3.07	11.54	100	90	100
B	0.67	3.85	8.51	100	96	100	0.67	3.23	8.08	0.84	3.58	11.93	100	100	100
B'	0.55	3.73	6.81	100	86	89	0.55	3.21	7.34	0.67	3.08	10.47	100	90	90
C	0.67	3.85	7.44	100	96	100	0.67	3.24	8.51	0.84	3.59	14.14	100	100	100
C'	0.55	3.73	6.43	100	86	96	0.55	3.22	6.31	0.68	3.07	8.01	100	90	100
D	0.67	3.85	9.25	100	96	100	0.67	3.25	7.88	0.84	3.59	11.21	100	100	100
D'	0.55	3.73	8.43	100	86	98	0.55	3.23	9.46	0.68	3.07	10.09	100	90	100
E	1.00	7.42	5.63	100	100	100	1.00	6.98	5.73	1.23	7.01	8.62	100	100	100
E'	1.20	8.03	7.91	100	100	100	1.20	6.76	9.41	1.52	8.15	17.21	100	100	100
F	1.24	11.26	10.66	100	100	100	1.24	10.02	6.84	1.53	9.42	13.38	100	100	100
F'	0.87	4.47	5.45	100	24	100	0.87	5.19	5.55	1.10	5.90	6.48	100	100	100
G	0.99	7.35	7.08	100	100	100	0.99	6.97	6.94	1.24	7.01	10.75	100	100	100
G'	1.19	7.96	11.01	100	100	100	1.19	6.72	11.87	1.53	8.21	10.99	100	100	100
H	1.24	11.26	9.06	100	100	100	1.24	9.98	8.17	1.53	9.43	16.36	100	100	100
H'	0.87	4.47	5.01	100	100	100	0.87	5.20	5.62	1.10	5.87	9.59	100	100	100

Table 2: Experimental results with parsing method frobbing/combined. Processing time is measured in seconds per sentence on a 2.8 GHz Pentium D Processor. Quality is presented as sentence accuracy for both levels (syntactic and thematic).

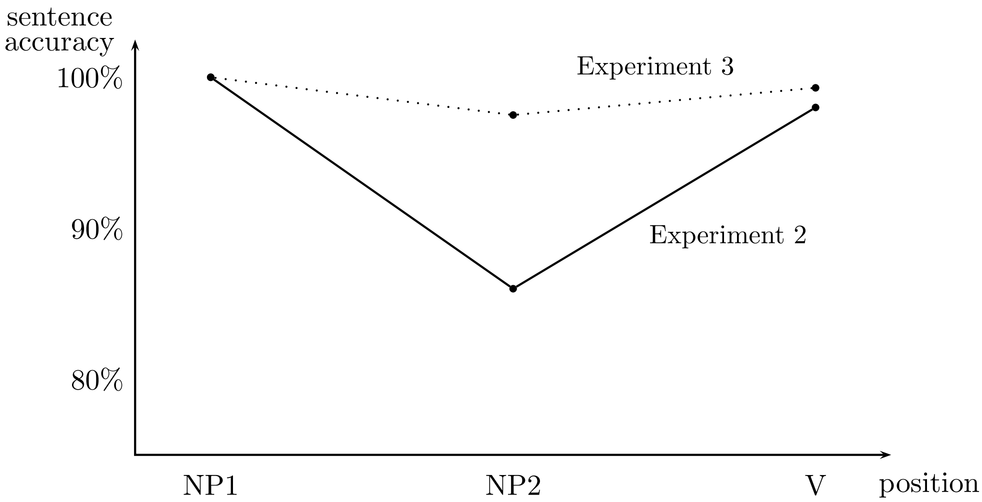


Figure 4: Parsing Quality at the critical points with the original grammar (Experiment 2) and the reduced preference for non-argument attachments

<table>
<tr><td colspan="2">A(NP1) = B(NP1) = C(NP1) = D(NP1)
A(NP2) = B(NP2) = C(NP2) = D(NP2)</td><td colspan="2">A’(NP1) = B’(NP1) = C’(NP1) = D’(NP1)
A’(NP2) = B’(NP2) = C’(NP2) = D’(NP2)</td></tr>
<tr><td>E(NP1) = G(NP1)
E(NP2) = G(NP2)</td><td>E’(NP1) = G’(NP1)
E’(NP2) = G’(NP2)</td><td>F(NP1) = H(NP1)
F(NP2) = H(NP2)</td><td>F’(NP2) = H’(NP2)
F’(NP1) = H’(NP1)</td></tr>
</table>

The quality figures for Experiment 2 are the same as for Experiment 1 and therefore not repeated. Since different hardware has been used for Experiments 1 and 2, the time figures for prefix parsing had to be normalized to the time for the first increment of the pseudo-incremental case.

6.3 Parsing quality

At a first glance, the global quality of parsing decisions seems very high. With a completely correct structure for whole sentences (including the matrix clause and the final subclause) on both levels (syntax and thematic) in 97.3% of the test sentences it exceeds the quality level of current state-of-the-art dependency parsers on general text corpora by far. Given, however, the very regular structure of the test data this result is not really a surprising one. With a few exceptions (e.g. the ambiguity at the verb *gefallen* between its object-experiencer interpretation and the participle form of *fallen*) no serious parsing problems do occur.

A closer inspection reveals that intermediate results at the critical points in the sentence are far less reliable (c.f. Figure 4). The treatment of the second NP seems particularly error prone. Nevertheless, for the majority of test sentences the interpretation-switching behaviour predicted by the Argument Dependency Model can be replicated by the WCDG parser during left-to-right incremental parsing.

The vast majority of error cases (93%) are confusions of an argument with either a genitive modifier (GMOD), an apposition (APP), or an ethical dative (ETH). They obviously have been caused by a mismatch between fundamental assumptions of the Argument Dependency

Model and the original WCDG grammar: While in the Argument Dependency Model argument attachments take priority over all other attachments, WCDG attaches a noun phrase as an argument only, if this is licensed by the subcategorization pattern of the verb. This mismatch could explain the comparatively low accuracy at the second NP, since at that position the verb information is not yet available. As a consequence, the parser runs into systematic problems in cases with a nominative-genitive or genitive-dative syncretism, since here an alternative non-argument interpretation is readily available. Indeed the extremely high error rate at the second NP of condition F'/H' (*... dass den Winzern der Betrüger*) has been caused by such an ambiguity, which can only be resolved after the agreement information and the valence requirements of the verb become available. The problem is particularly severe, since the test data set does not contain any sentences with non-argument NPs, making the parser fail on all the relevant instances.

Therefore, the third experiment has been conducted to estimate the potential for improvement. It shows that already a very simple modification of the grammar (penalizing the non-argument attachment alternatives) almost completely removes the problem. The dotted line in Figure 4 shows the corresponding result for the labels GMOD and ETH. Similar results can also be expected for arguments wrongly attached as appositions to the preceding noun. Note that the measures for the sentence accuracy in Figure 4 cannot be directly compared, since Experiment 3 has been carried out on only a subset of the test corpus.

This finding at least confirms the assumption, that there is potential to further improve the parsing results in accordance with the predictions of the Argument Dependency Model. Considered in isolation, however, the result is not very meaningful, since a modified preference of the grammar might of course negatively affect its performance on a general purpose corpus. Whether this is the case and if so, whether an acceptable balance can be found remains as a goal for further investigations.

In general, it is not clear, how the performance of an incremental model can be evaluated on a standard treebank at all. Thus, transferring the implementation of the Argument Dependency Model used so far to the case of unrestricted text as used in standard evaluations poses yet another challenge, since no gold standard for parsing unrestricted text in an incremental manner is available so far. To also demonstrate the appropriateness of the model derived for the controlled conditions of the stimulus sentences under the open world conditions of general text data, it would be necessary, to make assumptions about the intermediate interpretations derived by humans during online comprehension. The Argument Dependency Model is only of limited use in that situation.

6.4 Resource requirements

6.4.1 The benefit of pseudo-incremental parsing

Usually, incremental sentence processing is expected to facilitate a noticeable reduction in the resource requirements for processing the next increment based on the information already derived from the preceding ones. Unfortunately, the experimental results for prefix parsing and pseudo-incremental parsing show no substantial difference. Although a small gain (10%) in parsing efficiency has been obtained after the second NP this effect vanishes completely when integrating the Verb. The effect is by no means stable and obviously highly sensitive to details of the grammar design. In principle, one would expect the parser to benefit most, if the new increment can be integrated into the currently available structure without any revisions. Actually, this is not the case. Even if neither an early nor a late revision is necessary (conditions A, B', E, and E'), a slightly larger speedup of 12% at the second NP turns into a later 11% slowdown at the verb, while the conditions which do require a reanalysis benefitted instead.

This cognitively highly inadequate behaviour clearly demonstrates that verb related information of the current WCDG grammar is still crucial to effectively guide the transformation process of the parser towards the optimal structural interpretation. To achieve a clear advantage, the parser would certainly also require strong heuristics about which dependency edges can safely be exempted from further repair steps.

6.4.2 Reanalysis effects

Comparing the four different cases of reanalysis (RA) with respect to their specific time requirements during pseudo-incremental parsing a general qualitative tendency can be observed. Let X $<_{CPU}$ Y denote the relationship that condition X is faster than condition Y the following hierarchical ordering has been found

$$\text{no RA} <_{CPU} \text{syntactic RA} <_{CPU} \text{both RAs} <_{CPU} \text{thematic RA}$$

having only a single exception, namely C' $<_{CPU}$ B' holds although C' requires a syntactic reanalysis while B' does not. While the first two of these hierarchical relationsships seem plausible, both from a cognitive and a computational point of view, the last one (both RAs $<_{CPU}$ thematic RA) deserves more attention. One possible reason for this unexpected behaviour might be that the influence of the thematic constraints within the grammar is too weak at the moment and therefore thematic reanalysis is pursued by the parser only with a low priority. To check this hypothesis, systematic experimentation with a range of modelling variants will be necessary.

Another question concerns a possible correlation between the position in the sentence where the evidence for revision becomes available and the computational effort at this position. To answer this question, the conditions with an unambiguous case marking (E to H) have to be compared. Here, at least for the singular-first sentences a highly systematic relationship has been found (c.f. Figure 5).

It shows that the need for reanalysis systematically leads to higher resource requirements and that these requirements are triggered exactly at the point in the sentence where the evidence becomes available. Unfortunately this behaviour does not extend to the conditions E'-H' where the plural noun phrase precedes the singular (c.f. Figure 6).

Here, the case of the first NP is actually no longer unique (nominative or accusative). This does not pose a serious processing problem to the Argument Dependency Model which simply checks whether nominative can be assigned, without caring about alternative interpretations. The same situation, however, confronts the WCDG parser with a greater space of transformation possibilities, which at least have to be checked before they can be rejected due to not leading to a competitive score. Again, it needs to be determined whether this relationship can be reversed by e.g. strenghtening the the influence of the thematic default rule ECONOMY.

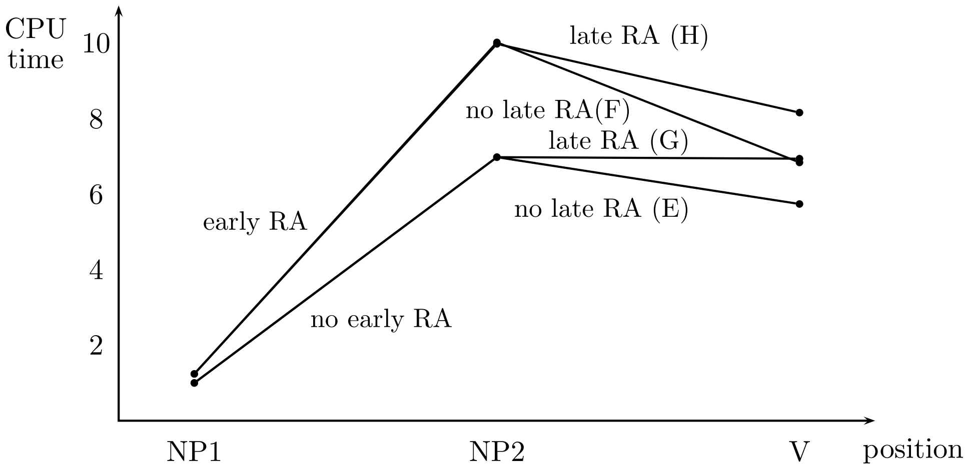


Figure 5: Computational effort at the critical points for sentences with the singular NP preceding the plural

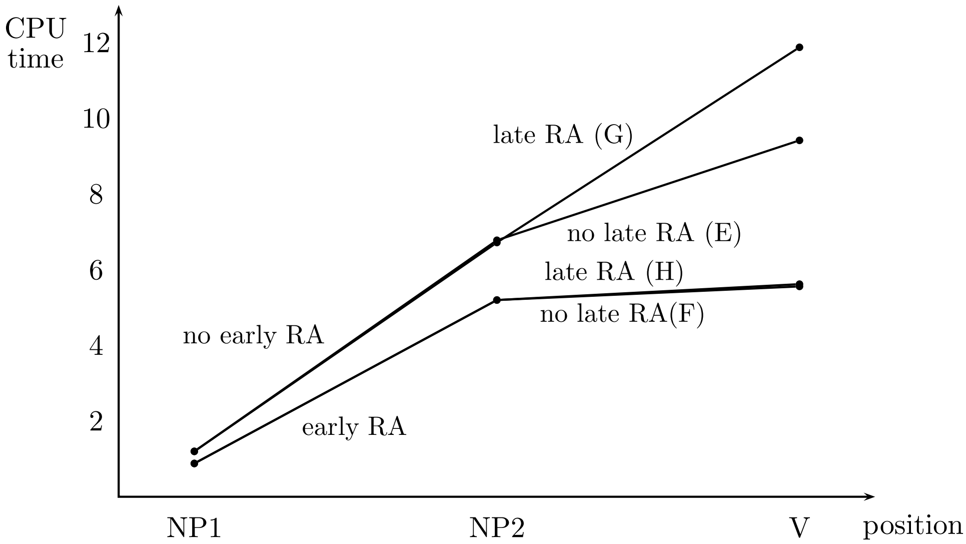


Figure 6: Computational effort at the critical points for sentences with the plural NP preceding the singular

7 Conclusions

A model which captures dependencies between the arguments of a common head has been implemented using a parsing formalism based on weighted constraints. Modeling such dependencies is motivated by the need to compensate the loss of disambiguating information in a partially available sentence, thus making possible well-informed early decisions about the type of attachment even in cases where the attachment point has not yet been observed. It has been shown that weighted constraints are an appropriate means to model grammar in a way that mirrors human preferences about the interpretation of incomplete sentences and that produces as a consequence of their application the non-monotonic syntactic re-interpretations of the sentence as it unfolds. Applying the suitably extended broad-coverage parser to the very same test data, which have been used in psycholinguistic experiments, shows that this pattern of early commitment can be computationally replicated with a very high degree of reliability. Moreover, at least under certain conditions, the reanalysis effort did occur exactly at the position in the sentence where the necessity for revision arises.

Providing each incremental parsing step just with information about the optimal structure determined in the preceding one, did not yield the expected gain in efficiency compared to pure prefix parsing. Further experiments with a richer interface and effective heuristics about those parts of the structure which can and need to be excluded from further modification are certainly necessary.

References

Ina Bornkessel. *The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation*. PhD Dissertation, University of Potsdam, 2002.

Ina Bornkessel, Stefan Zysset, Angela D. Friederici, D. Yves von Cramon, and Matthias Schlesewsky. Who did what to whom? the neural basis of argument hierarchies during language comprehension. *NeuroImage*, 26(1):221–223, 2005.

David Dowty. Thematic proto-roles and argument selection. *Language*, 67(3):547–619, 1991.

Kilian Foth and Wolfgang Menzel. Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In *Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006*, Sydney, 2006.

Kilian Foth, Stefan Hamerich, Ingo Schröder, Michael Schulz, and Tomas By. [X]CDG User Guide, 2003.

Kilian A. Foth. *Hybrid Methods of Natural Language Analysis*. PhD thesis, Universität Hamburg, Fachbereich Informatik, 2006.

Kilian A. Foth, Wolfgang Menzel, Horia F. Pop, and Ingo Schröder. An experiment in incremental parsing using weighted constraints. In *Proc. 18th International Conference on Computational Linguistics, Coling 2000*, pages 1026–1030, Saarbr"ucken, 2000a.

Kilian A. Foth, Wolfgang Menzel, and Ingo Schröder. A transformation-based parsing technique with anytime properties. In *Proc. 4th Int. Workshop on Parsing Technologies*, pages 89–100, Trento, Italy, 2000b.

Angela D. Friederici. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science*, 6(2):78–84, 2002.

Angela D. Friederici. The neurobiology of language comprehension. In Angela D. Friederici, editor, *Language comprehension: A biological perspective*, pages 263–301. Springer-Verlag, Berlin, Heidelberg, New York, 1999.

Frank Keller. A probabilistic parser as a model of global processing difficulty. In *Proc. 25th Annual Conference of the Cognitive Science Society*, pages 646–651, Mahawah, NJ, 2003. Erlbaum.

Joakim Nivre, Johan Hall, Jens Nilsson, Gülsen Eryigit, and Svetoslav Marinov. Labelled pseudo-projective dependency parsing with support vector machines. In *Proceedings of CoNLL 2006*, pages 221–225, 2006.

Beatrice Primus. *Cases and Thematic Roles: Ergative, Accusative and Active.* Max Niemeyer Verlag, Tübingen, 1999.

Matthias Schlesewsky and Ina Bornkessel. On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua*, 114(9–10):1213–1234, 2004.

Ingo Schröder. *Natural Language Parsing with Graded Constraints.* PhD thesis, Universität Hamburg, Fachbereich Informatik, 2002.

Robert D. van Valin Jr. and Randy J. LaPolla. *Syntax: Structure, Meaning, and Function.* Cambridge University Press, Cambridge, 1997.