Surfaces and Depths in Text Understanding: The Case of Newspaper Commentary

Manfred Stede
University of Potsdam
Dept. of Linguistics
Applied Computational Linguistics
D-14415 Potsdam
Germany
stede@ling.uni-potsdam.de

Abstract

Using a specific example of a newspaper commentary, the paper explores the relationship between 'surface-oriented' and 'deep' analysis for purposes such as text summarization. The discussion is followed by a description of our ongoing work on automatic commentary understanding and the current state of the implementation.

1 Introduction

Generally speaking, language understanding for some cognitive agent means reconstructing the presumed speaker’s goals in communicating with her/him/it. An application-specific automatic system might very well hardwire some or most of the aspects of this reconstruction process, but things get more interesting when the complexity is acknowledged and paid attention to. When moving from individual utterances to understanding connected discourse, an additional problem arises: that of partitioning the material into segments (usually at various levels) and that of inferring the connections between text segments (or between their underlying illocutions).

In recent years, some surface-based approaches to “rhetorical parsing” have been proposed, which try to recover a text’s discourse structure, following the general layout of Rhetorical Structure Theory (Mann, Thompson, 1988). Starting from this idea, in this paper, we imagine to push the goal of rhetorical parsing a bit further. The idea is that of a system that can take a newspaper commentary and understand it to the effect that it can, amongst other things, produce the “most concise summary” of it:

- the topic of the commentary
- the position the author is taking toward it

This goal does not seem reachable with methods of shallow analysis alone. But why exactly is it not, and what methods are needed in addition? In the following, we work through a sample commentary and analyse the steps and the knowledge necessary to arrive at the desired result, i.e., a concise summary. Thereafter, we sketch the state of our implementation work, which follows the goal of fusing surface-based methods with knowledge-based analysis.

2 Sample commentary

Figure 1 shows a sample newspaper commentary, taken from the German regional daily “Märkische Allgemeine Zeitung” in October 2002, along with an English translation. To ease reference, numbers have been inserted in front of the sentences. Let us first move through the text and make some clarifications so that the reader can get the picture of what is going on. Dagmar Ziegler is the treasury secretary of the German state of Brandenburg. A plan for early retirement of teachers had been drafted collectively by her and the education secretary, whose name is Reiche. Sentence 5 points out that the plan had intended education to be exempt from the cutbacks happening all over the various ministries — Reiche’s colleagues in 6 are thus the other secretaries. While the middle part of the text provides some motivation for the withdrawal, 9-14 state that the plan nonetheless should be implemented, for the reasons given in 10-12. Our intended “most concise summary” then would be:

- Topic: Treasury secretary delays decision on teacher staff plan
- Author’s opinion: Government has to decide quickly and give priority to education, thus implement the plan

Notice that a statistical summarization technique (i.e., a sentence extraction approach) is very unlikely to yield

(1) Dagmar Ziegler is up to her neck in debt. (2) Due to the dramatic fiscal situation in Brandenburg she now surprisingly withdrew legislation drafted more than a year ago, and suggested to decide on it not before 2003. (3) Unexpectedly, because the ministries of treasury and education both had prepared the teacher plan together. (4) This withdrawal by the treasury secretary is understandable, though. (5) It is difficult to motivate these days why one ministry should be exempt from cutbacks — at the expense of the others. (6) Reiche’s colleagues will make sure that the concept is waterproof. (7) Indeed there are several open issues. (8) For one thing, it is not clear who is to receive settlements or what should happen in case not enough teachers accept the offer of early retirement. (9) Nonetheless there is no alternative to Reiche’s plan. (10) The state in future has not enough work for its many teachers. (11) And time is short. (12) The significant drop in number of pupils will begin in the fall of 2003. (13) The government has to make a decision, and do it quickly. (14) Either save money at any cost - or give priority to education.

Figure 1: Sample text with translation

It includes the final sentence (most probably because it is the final sentence), but in the context of the other two extracted sentences it does not convey the author’s position — nor the precise problem under discussion.

3 Rhetorical Structure
Since RST (Mann, Thompson 1988) has been so influential in discourse-oriented computational linguistics, we start our analysis with a “man-made” RST analysis, which was produced collectively by two RST-experienced students. See Figure 2.1 (The English reader can relatively easy map the German segments to their translations in Fig. 1 with the help of the sentence numbers added to the text in the tree).

Some considerations motivating this analysis (in terms of segment numbers, not sentence numbers): 1 is seen as the general Background for the satellite of the overall Concession, which discusses the problem arising from the debt situation. Arguably, it might as well be treated as Background to the entire text. The Evaluation between 2-6 and 7-12 is a relation often found in opinion texts; an alternative to be considered here is Antithesis — in this case, however, 7-12 would have to be the nucleus, which seems to be problematic in light of the situation that 3-4 is the main portion that is being related to the material in 13-16.

8-12 explains and elaborates the author’s opinion that the withdrawal is understandable (7). The distinctions between the relations Explanation, Elaboration, and Evidence were mostly based on surface cues, such as tatsächlich (‘indeed’) signalling Evidence. The Elabora-

\footnote{Visualization by the RST Tool (O’Donnell, 1997). Notation follows Mann and Thompson (1988): vertical bars and incoming arrows denote nuclear segments, outgoing arrows denote satellites. Numbers at leaves are sentence numbers; segment numbers are given at internal nodes.}


4 The role of RST trees in text understanding

The prospects for applying RST trees to natural language text understanding were noted in Section 1, and should be implemented education since exercises in Section 1 indicate that the RST tree as a whole should be implemented in 1970's. In exercises, we use a natural language text understanding framework that incorporates the RST tree as a whole. The framework is implemented in 1970's, and is designed for exercises in Section 1, where we use the RST tree as a whole to implement exercises. In exercises, we use the RST tree as a whole to implement exercises. In exercises, we use the RST tree as a whole to implement exercises.

\[ \text{Figure 2: RST tree for sample text} \]
would arrive at such a tree — more specifically, at a formal representation of it.

What kind of information is necessary beyond assigning relations, spans and nuclei? In our representation of the summary tree, we have implicitly assumed that reference resolution has been worked out - in particular that the legislation can be identified in the satellite of the Explanation, and also in its nucleus, where it figures implicitly as the object to be decided upon. Further, an RST tree does not explicitly represent the topic of the discourse, as we had asked for in the beginning. In our present example, things happen to work out quite well, but in general, an explicit topic identification step will be needed. And finally, the rhetorical tree does not have information on illocution types (1-place rhetorical relations, so to speak) that distinguish reported facts (e.g., segments 3 and 4) from author’s opinion (e.g., segment 7). We will return to these issues in Section 6, but first consider the chances for building up rhetorical trees automatically.

5 Prospects for Rhetorical Parsing

Major proponents of rhetorical parsing have been (Sumita et al., 1992), (Corston-Oliver, 1998), (Marcu, 1997), and (Schilder, 2002). All these approaches emphasise their membership in the “shallow analysis” family; they are based solely on surface cues, none tries to work with semantic / domain / world knowledge. (Corston-Oliver and Schilder use some genre-specific heuristics for preferential parsing, though.) In general, our sample text belongs to a rather “friendly” genre for rhetorical parsing, as commentaries are relatively rich in connectives, which are the most important source of information for making decisions — but not the only one: Corston-Oliver, for example, points out that certain linguistic features such as modality can sometimes help disambiguating connectives. Let us now hypothesize what an “ideal” surface-oriented rhetorical parser, equipped with a good lexicon of connectives, part-of-speech tagger and some rough rules of phrase composition, could do with our example text.

5.1 Segmentation

As we are imagining an “ideal” shallow analyser, it might very well produce the segmentation that is underlying the human analysis in Figure 2. The obvious first step is to establish a segment boundary at every full stop that terminates a sentence (no ambiguities in our text). Within sentences, there are six additional segment boundaries, which can be identified by considering connectives and part-of-speech tags of surrounding words, i.e. by a variant of “chunk parsing”: Auf Grund (‘due to’) has to be followed by an NP and establishes a segment up to the finite verb (hat). The und (‘and’) can be identified to conjoin complete verb phrases and thus should trigger a boundary. In the following sentence, weil (‘because’) has to be followed by a full clause, forming a segment. The next intra-sentential break is between segments 11 and 12; the oder (‘or’) can be identified like the und above. In segment 17-18, und zwar (‘and in particular’) is a strict boundary marker, as is the entweder – oder (‘either – or’) construction in 19-20.

5.2 Relations, scopes, nuclei

The lexical boundary markers just mentioned also indicate (classes of) rhetorical relationships. Auf Grund — when used in its idiomatic reading — signals some kind of Cause with the satellite following in an NP. Because the und in 3-4 co-occurs with the temporal expressions jetzt (‘now’) and erst 2003 (‘not before 2003’), it can be taken as a signal of Sequence here, with the boundaries clearly identifiable, so that the RST subtree 2-4 can be derived fully. Furthermore, 5 takes up a single adverbial überraschend from 3, and in conjunction with the weil-clause in 6, the Elaboration can be inferred. weil (‘because’) itself signals some Cause, but the necessity decision (which in the “real” tree in Fig. 2 leads to choosing Result) is difficult here; since 5 merely repeats a lexeme from 3, we might assign necessity status to 6 on the “surface” grounds that it is longer and provides new material. We thus have derived a rhetorical structure for the entire span 2-6. In 7, aber (‘but’) should be expected to signal either Contrast or Concession; how far the left-most span reaches can not be determined, though. Both 8 and 9 provide no reliable surface clues. In 10, tatsächlich (‘indeed’) can be taken as an adverbial indicating Evidence; again the scope towards the left is not clear. So .. etwa (‘thus .. for instance’) in 11 marks an Elaboration, and the oder in 12 a Disjunction between the two clauses. Span 10-12 therefore receives an analysis. In 13, dennoch (‘nonetheless’) is a clear Concession signal, but its scope cannot be reliably determined. Finally, the only two remaining decisions to be made from surface observations are the Elaboration 17-18 (und zwar, ‘and in particular’) and the Disjunction 19-20. Then, making use of RST’s “empty” relation Join, we can bind together the assembled pieces and are left with the tree shown in Fig. 4.
Figure 4: Result of “surface parsing” of sample text

Under the assumption that our discussion reasonably reflects the state of the art in surface-oriented analysis, we can apply surface-based heuristic methods to finding larger segments, i.e., to split the text into its major parts, which has sometimes been called “text tiling”. For instance, a segment boundary might be hypothesized if a certain nucleus stretches to the right is difficult to see, though it is somewhat likely (though wrong in this case!) that the segment hosting the connective, in our case “and”, is a major segment. Such considerations can help to identify smaller parts of a larger segment (such as the evaluation') at the beginning of 17, which has no antecedent hint and hence is a prominent nucleus for our analysis. Hence, the nucleus at the beginning of 17, which we hypothesize as joint, is similar to the nucleus of Contrast/Concession at 13 — it is a major segment in its own right. The nuclei at the end of 17 and the beginning of 18, which we hypothesize as Concession, are similar with the Concession hypothesized at 13 — they are major nuclei with the Concession hypothesis at 13.

Returning to the issue of segmentation, we can also try to apply surface-based methods, we now have to compare its result to our overviews of earlier sections. As for the analysis just proposed, we used lexical knowledge about the satellite span nothing can be said here. For instance, a segment boundary might be hypothesized if a certain nucleus stretches to the right is difficult to see, though it is somewhat likely (though wrong in this case!) that the segment hosting the connective, in our case “and”, is a major segment. Such considerations can help to identify smaller parts of a larger segment (such as the evaluation') at the beginning of 17, which has no antecedent hint and hence is a prominent nucleus for our analysis. Hence, the nucleus at the beginning of 17, which we hypothesize as joint, is similar to the nucleus of Contrast/Concession at 13 — it is a major segment in its own right. The nuclei at the end of 17 and the beginning of 18, which we hypothesize as Concession, are similar with the Concession hypothesized at 13 — they are major nuclei with the Concession hypothesis at 13.

Returning to the issue of segmentation, we can also try to apply surface-based methods, we now have to compare its result to our overviews of earlier sections. As for the analysis just proposed, we used lexical knowledge about the satellite span nothing can be said here. For instance, a segment boundary might be hypothesized if a certain nucleus stretches to the right is difficult to see, though it is somewhat likely (though wrong in this case!) that the segment hosting the connective, in our case “and”, is a major segment. Such considerations can help to identify smaller parts of a larger segment (such as the evaluation') at the beginning of 17, which has no antecedent hint and hence is a prominent nucleus for our analysis. Hence, the nucleus at the beginning of 17, which we hypothesize as joint, is similar to the nucleus of Contrast/Concession at 13 — it is a major segment in its own right. The nuclei at the end of 17 and the beginning of 18, which we hypothesize as Concession, are similar with the Concession hypothesized at 13 — they are major nuclei with the Concession hypothesis at 13.

Returning to the issue of segmentation, we can also try to apply surface-based methods, we now have to compare its result to our overviews of earlier sections. As for the analysis just proposed, we used lexical knowledge about the satellite span nothing can be said here. For instance, a segment boundary might be hypothesized if a certain nucleus stretches to the right is difficult to see, though it is somewhat likely (though wrong in this case!) that the segment hosting the connective, in our case “and”, is a major segment. Such considerations can help to identify smaller parts of a larger segment (such as the evaluation') at the beginning of 17, which has no antecedent hint and hence is a prominent nucleus for our analysis. Hence, the nucleus at the beginning of 17, which we hypothesize as joint, is similar to the nucleus of Contrast/Concession at 13 — it is a major segment in its own right. The nuclei at the end of 17 and the beginning of 18, which we hypothesize as Concession, are similar with the Concession hypothesized at 13 — they are major nuclei with the Concession hypothesis at 13.

Returning to the issue of segmentation, we can also try to apply surface-based methods, we now have to compare its result to our overviews of earlier sections. As for the analysis just proposed, we used lexical knowledge about the satellite span nothing can be said here. For instance, a segment boundary might be hypothesized if a certain nucleus stretches to the right is difficult to see, though it is somewhat likely (though wrong in this case!) that the segment hosting the connective, in our case “and”, is a major segment. Such considerations can help to identify smaller parts of a larger segment (such as the evaluation') at the beginning of 17, which has no antecedent hint and hence is a prominent nucleus for our analysis. Hence, the nucleus at the beginning of 17, which we hypothesize as joint, is similar to the nucleus of Contrast/Concession at 13 — it is a major segment in its own right. The nuclei at the end of 17 and the beginning of 18, which we hypothesize as Concession, are similar with the Concession hypothesized at 13 — they are major nuclei with the Concession hypothesis at 13.

Returning to the issue of segmentation, we can also try to apply surface-based methods, we now have to compare its result to our overviews of earlier sections. As for the analysis just proposed, we used lexical knowledge about the satellite span nothing can be said here. For instance, a segment boundary might be hypothesized if a certain nucleus stretches to the right is difficult to see, though it is somewhat likely (though wrong in this case!) that the segment hosting the connective, in our case “and”, is a major segment. Such considerations can help to identify smaller parts of a larger segment (such as the evaluation') at the beginning of 17, which has no antecedent hint and hence is a prominent nucleus for our analysis. Hence, the nucleus at the beginning of 17, which we hypothesize as joint, is similar to the nucleus of Contrast/Concession at 13 — it is a major segment in its own right. The nuclei at the end of 17 and the beginning of 18, which we hypothesize as Concession, are similar with the Concession hypothesized at 13 — they are major nuclei with the Concession hypothesis at 13.
sions. First, rhetorical parsing should allow for under-
specified representations as — intermediate or final —
outcome; see (Hanneforth et al., submitted). Second, text understanding aiming at quality needs to go further
than surface-oriented rhetorical parsing. With the help
of additional domain/world-knowledge sources, attempts
should be made to fill gaps in the analysis. It is then
an implementation decision whether to fuse these addi-
tional processes into the rhetorical parser, or to use a
pipeline approach where the parser produces an un-
specified rhetorical tree that can afterwards be further en-
riched. Third, probabilistic or statistical knowledge can
also serve to fill gaps, but the information drawn from
such sources should be marked with its status being inse-
cure. As opposed to decisions based on lexical/linguistic
knowledge (in 5.2), the tentative decisions from 5.3 may
be overwritten by later knowledge-based processes.

6 Knowledge-Based Understanding

“Understanding a text” for some cognitive agent means to
fuse prior knowledge with information encountered in the
text. This process has ramifications for both sides: What
I know or believe influences what exactly it is that I “take
away” from a text, and my knowledge and beliefs will
usually to a certain extent be affected by what I read. Nat-
urally, the process varies from agent to agent: They will
understand different portions of a text in different ways
and to different degrees. Thus, when we endeavour to
device and implement models of text understanding, the
target should not be to arrive at “the one and only” result,
but rather to account for the mechanics of this variability:
the mechanism of understanding should be the same, but
the result depend on the type and amount of prior knowl-
edge that the agent carries. In the end, a representation
of text meaning should therefore be designed to allow for
this flexibility.

6.1 KB Design

In line with many approaches to using knowledge for
language processing, we adopt the framework of termi-
nological logic as the vehicle for representing both the
background knowledge necessary to bootstrap any under-
standing process, and the content of the text. Thus the ba-
sic idea is to encode prior, general knowledge in the TBox
(concepts) and the information from the text in the ABox
(instances). For our example, the subworld of govern-
ment, ministries and legislation has to be modelled in the
TBox, so that entities referred to in the text can instantiate
the appropriate concepts. We thus map the rhetorical tree
built up by shallow analysis to an ABox in the LOOM
language (MacGregor, Bates, 1987); for a sketch of re-
presenting rhetorical structure in LOOM, see (Stede, 1999,
ch. 10).

6.2 “Ideal” text understanding

Each leaf of the tree is now subject to detailed semantic
analysis and mapped to an enriched predicate/argument
structure that instantiates the relevant portions of the
TBox (quite similar to the ‘Text Meaning Representation’
of (Mahesh, Nirenburg, 1996)). “Enriched” indicates that
beyond the plain proposition, we need information such
as modality but also the type of illocution; e.g., does the
utterance represent a factual statement, the author’s opin-
ion, or a proposal? This is necessary for analyzing the
structure of an argument (but, of course, often it is very
difficult to determine).

One central task in text understanding is reference reso-
lution. Surface-based methods can perform initial
work here, but without some background knowledge, the
task can generally not be completed. In our sample
text, understanding the argument depends on recogniz-
ing that Kabinettsvorlage in (2), Lehrerpersonalkonzept
in (3), Konzept in (6), and Reiches Personalkonzept in (9)
all refer to the same entity; that Ziegler in (1) and Fi-
nanzminister in (4) are co-referent; that Finanz- und
Bildungspapier in (3), Reiches Ministerkollegen in (6),
and die Regierung in (13) refer to portions of or the com-
plete Brandenburg government, respectively. Once again,
hints can be derived from the surface words (e.g., by com-
pund analysis of Lehrerpersonalkonzept), but only back-
ground knowledge (an ontology) about the composition
of governments and their tasks enables the final decisions.

Knowledge-based inferences are necessary to infer
rhetorical relations such as Explanation or Evaluation.
Consider for example segment 15-16, where the relation-
ship between ‘time is short’ (a subjective, evaluative
statement) and ‘begin already in the fall of 2003’ (a state-
ment of a fact), once recognized, prompts us to assign
Explanation. Similarly, the Elaboration between this seg-
ment and the preceeding 14 can be based on the fact that
14 makes a statement about the ‘future situation’ in Bran-
denburg, which is made more specific by time being short
and the fall of 2003. More complex inferences are nec-
essary to attach 14-16 then to 13 (and similarly in the
segment 7-12).

6.3 “Realistic” text understanding

Even if it were possible to hand-code the knowledge base
such that for our present sample text the complete repre-
sentation can be constructed — for the general text analy-
sis situation, achieving a performance anywhere near the
“complete and correct solution” is beyond reach. As in-
dicated at the beginning of the section, though, this is not
necessarily bad news, as a notion of partial understand-
ing, or “mixed-depth encoding” as suggested by Hirst
and Ryan (1992), should be the rule rather than the ex-
ception. Under ideal circumstances, a clause at a leaf of the
rhetorical tree might be fully analyzed, with all refer-
ences resolved and no gaps remaining. In the worst case, however, understanding might fail entirely. Then, following Hirst and Ryan, the text portion itself should simply be part of the representation. In most cases, the representation will be somewhere in-between: some aspects fully analyzed, but others not or incompletely understood. For example, a sentence adverbial might be unknown and thus the modality of the sentence not be determined. The ABox then should reflect this partiality accordingly, and allow for appropriate inferences on the different levels of representation.

The notion of mixed depth is relevant not only for the tree’s leaves: Sometimes, it might not be possible to derive a unique rhetorical relation between two segments, in which case a set of candidates can be given, or none at all, or just an assignment of nucleus and satellite segments, if there are cues allowing to infer this. In (Reitter and Stede, 2003) we suggest an XML-based format for representing such underspecified rhetorical structures.

Projecting this onto the terminological logic scheme, and adding the treatment of leaves, we need to provide the TBox not only with concepts representing entities of “the world” but also with those representing linguistic objects, such as clause or noun group, and for the case of unanalyzed material, string. To briefly elaborate the noun group example, consider Reiches Ministerkollegen (‘Reiche’s colleagues’) in sentence 6. Shallow analysis will identify Reiche as some proper name and thus the two words as a noun group. An ABox instance of this type is created, and it depends on the knowledge held by the TBox whether additional types can be inferred. Reiche has not been mentioned before in the text, because from the perspective auf the author the name is prominent enough to be identified promptly by the (local) readers. If the system’s TBox contains a person of that name in the domain of the Brandenburg government, the link can be made; otherwise, Reiche will be some un-identified object about which the ABox collects some information from the text.

Representations containing material with different degrees of analysis become useful when accompanied by processes that are able to work with them (‘mixed-depth processing’). For summarization, this means that the task becomes one of fusing extraction (of unanalyzed portions that have been identified as important nuclei) with generation (from the representations of analyzed portions). Of course, this can lead to errors such as dangling anaphors in the extracted portions, but that is the price we pay for robustness — robustness in this refined sense of “analyse as deeply as you can” instead of the more common “extract something rather than fail.”

7 Implementation Strategy

Finally, here is a brief sketch of the implementation work that is under way in the Computational Linguistics group at Potsdam University. Newspaper commentaries are the genre of choice for most of our current work. We have assembled a corpus of some 150 commentaries from “Märkische Allgemeine Zeitung”, annotated with rhetorical relations, using the RST Tool by O’Donnell (1997). It uses an XML format that we convert to our format of underspecified rhetorical structure (‘URML’ Reitter & Stede 2003).

This data, along with suitable retrieval tools, informs our implementation work on automatic commentary understanding and generation. Focusing here on understanding, our first prototype (Hanneforth et al., submitted) uses a pipeline of modules performing

1. tokenization
2. sentence splitting and segmentation into clauses
3. part-of-speech tagging
4. chunk parsing
5. rhetorical parsing
6. knowledge-based processing

The tagger we are using is the Tree Tagger by Schmid (1994); the chunk parser is CASS (Abney 1996). The remaining modules, as well as the grammars for the chunk parser, have been developed by our group (including student projects).\(^2\) The rhetorical parser is a chart parser and uses a discourse grammar leading to a parse forest, and is supported by a lexicon of discourse markers (connectives). We have started work on reference resolution (in conjunction with named-entity recognition). Addition of the knowledge-based component, as sketched in the previous section, has just begun. The main challenge is to allow for the various kinds of underspecification within the LOOM formalism and to design appropriate inference rules.

As implementation shell, we are using GATE (http://www.gate.ac.uk), which proved to be a very useful environment for this kind of incremental system construction.

8 Conclusions

Knowledge-based text understanding and surface-based analysis have in the past largely been perceived as very different enterprises that do not even share the same

---

\(^2\)In addition to this “traditional” pipeline approach, Reitter (2003) performed experiments with machine learning techniques based on our MAZ corpus as training data.
goals. The paper argued that a synthesis can be useful, in particular: that knowledge-based understanding can benefit from stages of surface-based pre-processing. Given that

- pre-coded knowledge will almost certainly have gaps when it comes to understanding a “new” text, and
- surface-based methods yield “some” analysis for any text, however sparse, irrelevant or even wrong that analysis may be,

a better notion of robustness is needed that explains how language understanding can be “as good (deep) as possible or as necessary”. The proposal is to first employ “defensive” surface-based methods to provide a first, underspecified representation of text structure that has gaps but is relatively trustworthy. Then, this representation may be enriched with the help of statistical, probabilistic, heuristic information that is added to the representation (and marked as being less trustworthy). Finally, a “deep” analysis can map everything into a TBox/ABox scheme, possibly again filling some gaps in the text representation (Abox) on the basis of prior knowledge already encoded in the TBox. The deep analysis should not be an all-or-nothing step but perform as good as possible — if something cannot be understood entirely, then be content with a partial representation or, in the worst case, with a portion of the surface string.

Acknowledgements

Thanks to: Thomas Hanneforth and all the students of our Systemkonstruktion seminar for the implementation of the rhetorical parser prototype; anonymous reviewers for helpful comments on the paper; Markische Allgemeine Zeitung for providing us with plenty of commentaries.

References


