Querying and visualizing coreference annotation in multi-layer corpora

Christian Chiarcos, Julia Ritz, and Manfred Stede

Collaborative Research Center (SFB) 632 "Information Structure"
Department of Linguistics, Universität Potsdam
Karl-Liebknecht-Str. 24-25, D-14476 Golm (Potsdam), Germany
{chiarcos|jritz|stede}@uni-potsdam.de

Abstract. This paper describes the application of the linguistic information system ANNIS to the querying and evaluation of corpora with coreference and multiple other layers of hierarchical and relational annotation.
ANNIS provides a web-based graphical user interface that allows linguists to run queries against corpora in a data base and to display their results in a convenient way, without the necessity to become experts in any specific programming language. It is based on a generic data model that is capable to represent any kind of linguistic annotation. To our best knowledge, ANNIS is the *first and only* corpus query system that allows to formulate queries across an unrestricted combination of relational and hierarchical annotations.
This is illustrated here for a multi-layer corpus of German newspaper commentaries, the Potsdam Commentary Corpus [27, PCC], with corpus queries that address three independent layers of annotation, constituent syntax, coreference, and annotations for discourse structure.

Keywords: coreference, discourse structure, syntax, corpus querying, Potsdam Commentary Corpus, PAULA, ANNIS

1 Background and Motivation

In the last years, the number of corpora with coreference annotation beside other layers of annotations has significantly increased. At the same time, however, the joint evaluation of these annotations is usually performed by means of specialized scripts programmed by individual researchers for basically their own use. Aside from the system described in this paper, there does not seem to exist a tool to query and explore the full band-width of annotations found in such multi-layer corpora that is targeted towards end users, i.e., linguists, and in particular those working from a theoretical rather than a computational perspective: Existing approaches are either targeted towards expert programmers rather than linguists [4, 14, 2] – whereas we provide a linguistic information system tailored to end users –, or they are restricted with respect to the types and/or number of annotation layers to be combined [21, 26, 23] – whereas we support the free combination of any types of linguistic annotation.

We present ANNIS, an information system for richly and heterogeneously annotated corpora developed at the Collaborative Research Center (SFB) 632 "Information Structure" (University of Potsdam, Humboldt University Berlin). ANNIS is specifically designed to visualize, to retrieve from, and to mine corpora annotated at multiple layers. As an illustrative example for complex multi-layer queries, we address a long-standing assumption in cognitive-functional literature, i.e., the effect of discourse-structural subordination on anaphoric accessibility [12] relative to antecedent-based factors such as grammatical roles [13].

Below, corresponding queries are illustrated for the ProCon10 corpus, a selection of 10 German newspaper commentaries drawn from the Potsdam Commentary Corpus [27, PCC].

2 ANNIS: An Information System for Multi-Layer Corpora

ANNIS is an *ann*otation *i*nformation *s*ystem[1] specifically developed for the processing of multi-layer annotated corpora at the Collaborative Research Center (SFB) 632 "Information Structure". ANNIS provides a web browser-based search and visualization environment designed to access richly annotated corpora with heterogeneous annotation schemes [6, 30]. In its current implementation, ANNIS2, the backend of the system is based on a relational database (PostGreSQL).

ANNIS provides a unified access to various annotations for syntax, semantics, morphology, prosody, discourse structure and coreference. The relevant annotation layers of the ProCon10 corpus, as further described below, include `tiger` (constituent syntax à la TIGER [3]), `mmax` (coreference annotated with MMAX2 [21]), and `rst` (discourse structure annotations created with the RSTTool [22]).

In order to integrate these different annotations, the various source formats are converted to a generic XML format, PAULA [6]. PAULA developed out of early drafts of the ISO TC37/SC4 Linguistic Annotation Framework (LAF) [15], and like GrAF [16], the upcoming standard format developed for the LAF, PAULA is based on directed acyclic graphs. It is generally accepted that all linguistic annotations can be modelled as labels attached to a directed acyclic graph [1, 16, 14], hence our claim that ANNIS (whose data base scheme is just another linearization of the object model underlying PAULA) is capable to represent any kind of linguistic annotation.

Converting annotations to PAULA requires a reanalysis of these source formats according to the data structures of the PAULA Object Model, i.e., a labeled directed acyclic (hyper)graph. Primary data structures are *nodes* and *edges* for which various subtypes are distinguished: A node is either a *token* (a character span in the primary data), it is a *markable* (a span of tokens), or it is a *struct* (parent of other nodes). Edges are defined by the pair of nodes they connect:

[1] Originally, ANNIS was an acronym for " *ann*otation of *i*nformation *s*tructure", but in recent years, ANNIS was increasingly applied beyond the scope of the Collaborative Research Center (SFB) 632 "Information Structure".

Between a struct and its children exist *dominance relation*s, any other relation is classified as a *pointing relation*. All types of nodes and edges can be labeled with *features*, i.e., attribute-value pairs that express the actual annotations. In order to group together nodes, edges and labels of one single layer, they can be assigned a *namespace*, here `tiger`, `mmax` and `rst`.

After the conversion to PAULA, annotations that refer to the same piece of primary data can be merged into PAULA projects with multi-layer annotations [10], and then fed into the ANNIS database.

3 The Potsdam Commentary Corpus in ANNIS

The Potsdam Commentary Corpus [27, PCC] is a collection of 220 German newspaper commentaries (2.900 sentences, 44.000 tokens) taken from the online issues of the Märkische Allgemeine Zeitung (MAZ subcorpus) and Tagesspiegel (ProCon subcorpus).[2] Among the corpora available for German, it is outstanding with respect to the broad range of different types of linguistic information conveyed in its annotations. Among others, these include morphosyntax and syntax, coreference and information structure, as well as discourse structure, connectives and illocutions. Aside from ANNIS, we are not aware of any linguistic information system available that is able to provide access to these different layers of annotation simultaneously. In this paper, we show how ANNIS can be applied to achieve this task.

The annotations of *morphosyntax and syntax* applied to the PCC follow the conventions of the TIGER corpus [3]. For example (2), shown below with minimal context, Fig. 1 illustrates the mapping of the original syntax annotation (as visualized by TIGERSearch [19]) to the PAULA data model.

(1) *Berlin muss zurück zur Kultur ...*
Berlin needs.to return to.the culture
'Berlin needs to focus again on (lit. to return to) culture ...'

(2) *Viele Kulturschätze sind aber nicht in einem präsentablen Zustand*
many cultural.treasures are but not in a representable state
'Many cultural treasures are, however, not in a representable state.'

(3) *Wir müssen sie restaurieren ...*
we have.to them restore
'We have to restore them ...'

[2] A detailed description of the PCC and an online demo can be found under `http://www.ling.uni-potsdam.de/pcc/pcc.html`. The Potsdam Commentary Corpus is distributed free of charge under an academic licence, inquiries to access the data should be directed to Manfred Stede, `stede@ling.uni-potsdam.de`.
Note that at the moment, we prepare a new and extended release, this paper thus focuses on a subcorpus of the PCC only. An evaluation of the entire PCC with respect to the research queries mentioned above may be conducted after the revision is completed. For this paper, we consider the ProCon10 subcorpus as being sufficient for a proof of principle.

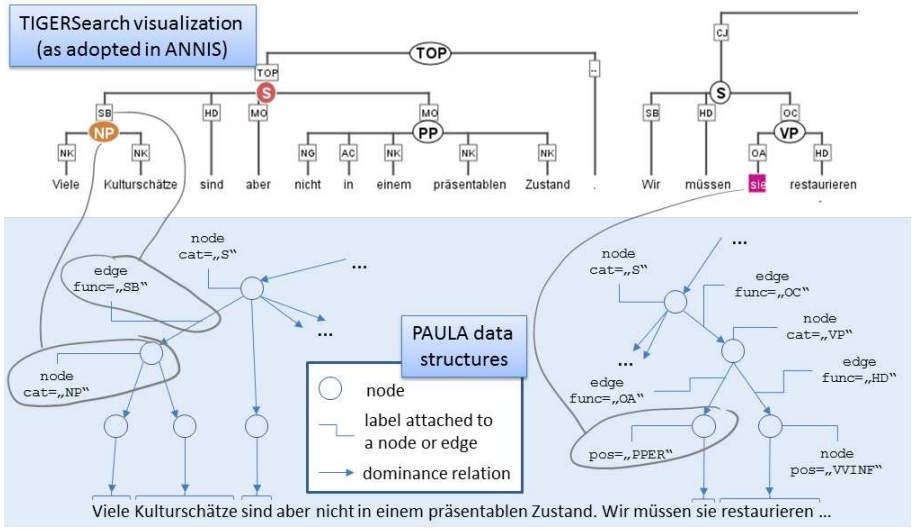

Fig. 1. Syntax annotation of example (2), TIGERSearch visualization and partial mapping to PAULA

The entire corpus is annotated for nominal and pronominal anaphoric *coreference* in accordance to the Potsdam Coreference Scheme [7] using MMAX2 [21]. About half of the corpus (105 texts) have been annotated with MMAX2 for *information structure*, i.e., topic ('what the sentence is about'), focus (the 'new' information in the sentence) and information status (the 'givenness' of the discourse referent at the time of its utterance). For the pronoun *sie* 'they' in ex. (3), Fig. 2 illustrates the mapping of the coreference annotation to PAULA data structures.

The PCC is one of very few corpora available with annotations for *discourse structure*, i.e., the hierarchical and relational structure of a discourse as a whole. At the moment, 86% of the PCC (190 texts) have been annotated in accordance with Rhetorical Structure Theory [20, RST] using the RSTTool [22]. Figure 3 illustrates the mapping of the original RST annotation (as visualized by the RSTTool) to the corresponding PAULA data structures.

Connectives are the most important surface factors for the assignment of RST annotations. As they are, however, often ambiguous, we introduced an independent annotation layer for connectives and their scopes (similar to the approach of the Penn Discourse Tree Bank, [24]) in the PCC. For connective annotation, we developed ConAno,[3] a tool that identifies potential connectives in text and that also makes suggestions for the two arguments of the connective. In preparation for an extended annotation of the PCC, ProCon10 has further been annotated for *illocutions and argumentative structure* using MMAX, that represent another important basis of discourse structure. The mapping of annotations for connec-

[3] The tool is available for download under `http://www.ling.uni-potsdam.de/pcc/Conano-Distrib.zip`.

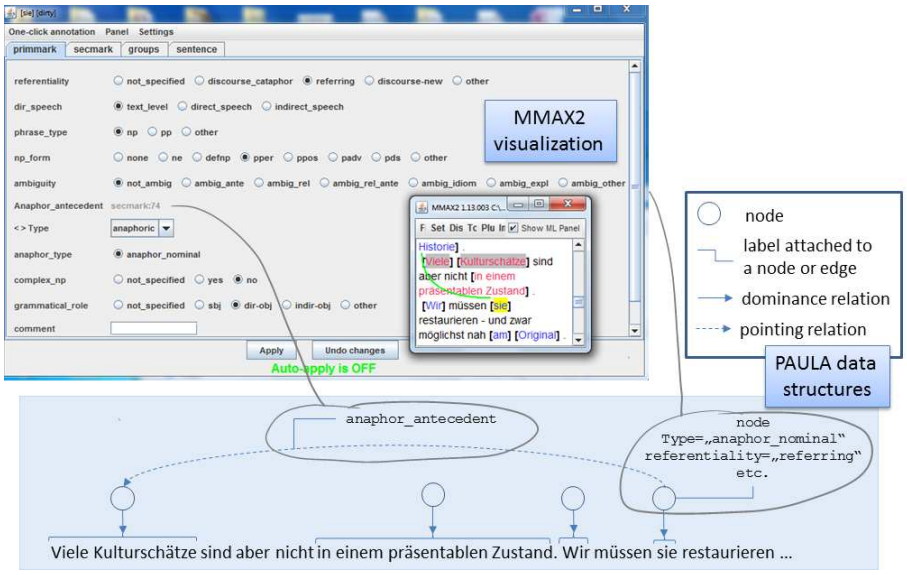

Fig. 2. Coreference annotation of the pronoun 'sie' in example (3), MMAX2 visualization and partial mapping to PAULA

tives and illocutions and argumentative structure is comparable to the mapping of coreference annotations from MMAX.

After conversion, the standoff XML linearization of PAULA allows to merge the different annotations into a single (hyper)graph [10] that represents the data structure of the ANNIS data base. Figure 4 summarizes the merged graph conveying the various annotations from Figs. 1, 2 and 3.

We would like to emphasize that our approach is designed to be generic and not specifically tailored to the aforementioned annotation tools and source formats. Among other resources, we have successfully imported corpora like TüBa-D/Z [28] and OntoNotes [25] into ANNIS, whose anaphoric and syntactic annotations differ greatly from those of the PCC.[4]

4 Querying, Visualizing and Evaluating Multi-Layer Annotations

ANNIS2, the current instantiation of the ANNIS database, consists of a relational database backend and a web interface that allows users to formulate queries, to export the results and to visualize them in a convenient way [30]. After login, the

[4] The ANNIS download page includes further examples for corpora with various other types of annotations, see `http://www.sfb632.uni-potsdam.de/d1/annis/download.html`.

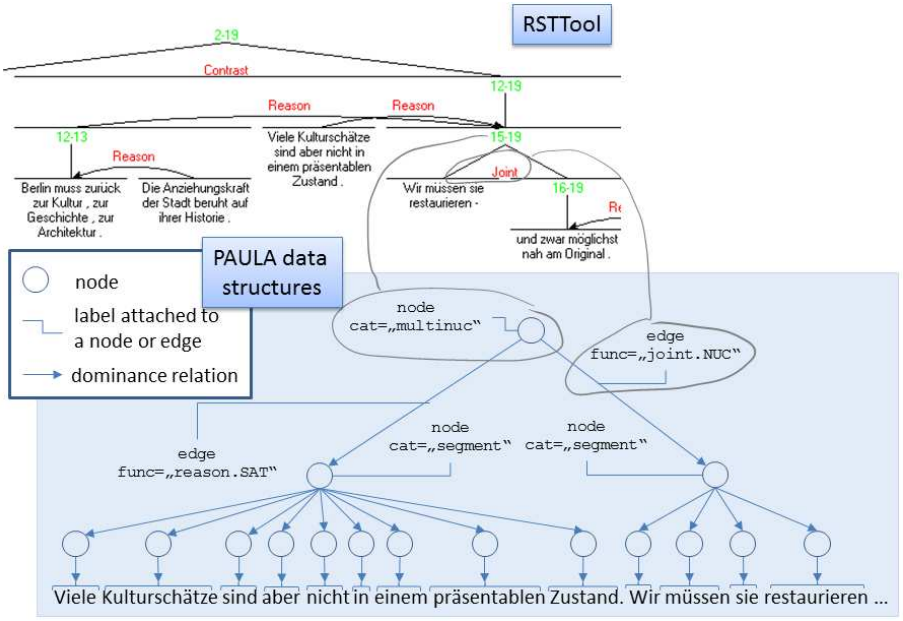

Fig. 3. Discourse structure visualization of example (2) and its context, RSTTool visualization and partial mapping to PAULA.

user may formulate a query (either directly or using a graphical query editor), select one or multiple corpora, and apply his query to these resources.

The primary elements of the query language are, nodes, edges and features (labels) as defined by PAULA:

Nodes can be queried for by their features, e.g., `tiger:cat="S"` retrieves a sentence node from the `tiger` namespace and establishes a variable `#`i (with index i).

Dominance relations can be queried with $>$ (parent-child relationship), with optional constraints for the features attached to it, e.g., `#1 >[func="SB"] #2` for `#2` being the grammatical subject of `#1`.

Pointing relations can be queried with the operator -> with a type attribute, e.g., `#1 ->anaphor_antecedent #2`.

Extensionality operators are typical for multi-layer annotations. They allow to query for nodes that cover the same stretch of primary data (co-extensionality: `#1 _=_ #2`), where the extension of one node covers the extension of another node (inclusion: `#1 _i_ #2`), etc.

With these operators, an example query for the ProCon10 corpus can now be easily formulated, that allows us to study the influence of the grammatical role of the antecedent (±subject_antec) and its discourse-structural subordination (±satellite_antec) on the realization of anaphors (±pronoun_anaph). An example query that retrieves pronouns with discourse-structurally subordinated subject antecedents is shown in Fig. 5. As an illustration for how this query works, the

nodes and edges of ex. (2) that match the corresponding variables are marked in Fig. 4.

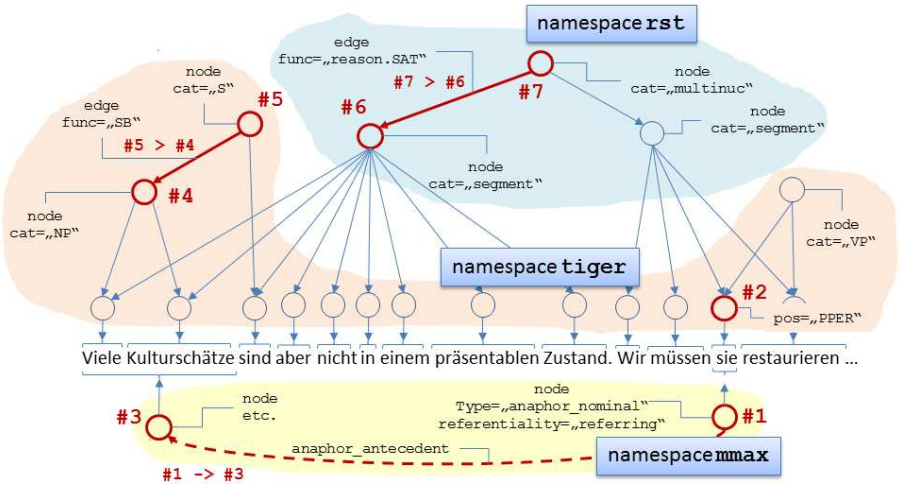

Fig. 4. Partial visualization of the PAULA graph resulting from Figs. 1, 2 and 3 and its mapping to the ANNIS visualizations. Nodes and edges that match the variables of the query in Fig. 5.

```
mmax:node &                                  // #1: anaphor (coreference layer)
tiger.pos=/P.*/ & #1 _=_ #2 &                // #2: anaphor is a pronoun (on syntax layer)
mmax:node & #1 ->anaphor_antecedent #3 &     // #3: antecedent (coreference layer)
tiger:node & #4 _=_ #3 &                     // #4: antecedent (syntax layer)
tiger:cat="S" & #5 >[func="SB"] #4 &         // #5: antecedent sentence, where #4 is grammatical subject
cat="segment" & #6 _i_ #4 &                  // #6: antecedent edu that includes #4
cat & #7 >[func=/.*SAT/] #6                  // #7: parent of antecedent edu, edu is a satellite
```

Fig. 5. Query for pronominal anaphors with subject antecedent in a discourse-structural satellite

Figure 6 shows how ex. (2) and its context is visualized by ANNIS. Aside from the visualization of query results, the screenshot also shows elements of the query interface: The top left text area is the textual query editor (a graphical query editor can be evoked through the “QueryBuilder” dialog below). Under the search form there is an area with metadata about the corpora currently selected by the user, here ProCon10 and PCC2, another demo corpus of the PCC. The two tabs in the field below can be used to define how query matches are to be processed. ANNIS provides different options to export query matches, or to visualize them graphically.

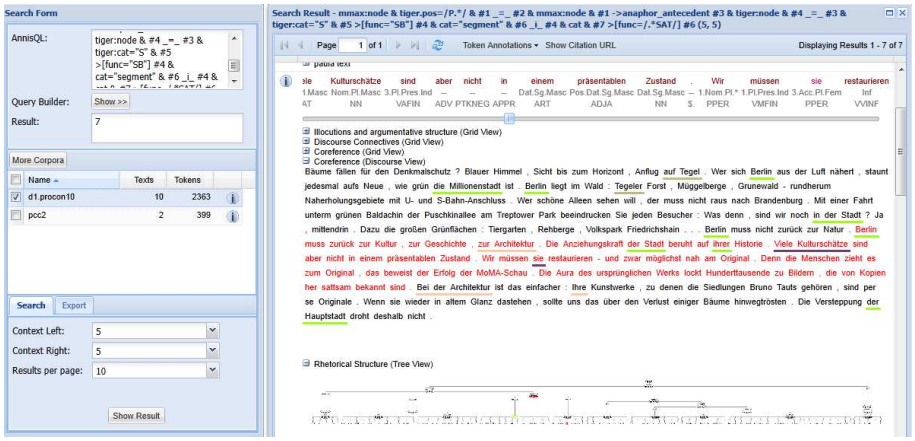

Fig. 6. ANNIS screenshot with query results for the query in Fig. 5.

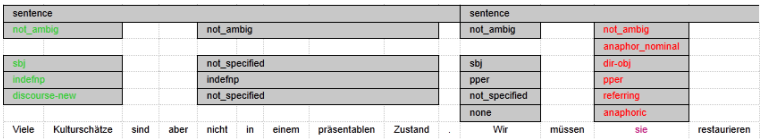

Fig. 7. Screenshot of the grid view visualization of the coreference annotation of ex. (3) and its context.

Query results are displayed in the "Search Result" pane that occupies the largest part of Fig. 6. Every match is represented at first with its various token-level annotations, here morphological and part-of-speech annotations drawn from the original TIGER annotation. More complex visualizations for different annotation layers or parts of annotation layers can be activated by clicking the "+" symbol. In the example, we see the discourse visualization of the coreference annotation that uses different colors to indicate anaphoric chains, and it marks user-selected chains by means of co-highlightning in different colors, if clicked upon. Red font indicates the text matched by the query (here, the extension of variable `#7`). Annotations attached to the original MMAX2 markables are visualized separately in a grid visualization (Fig. 7).

Below the coreference, Fig. 6 also includes a visualization of RST. The current RST visualization employs the visualization of syntax trees familiar from Fig. 1. This is possible as the PAULA data structures (nodes and dominance relations) are identical for both types of trees, only their labels and namespaces are different. This visualization is relatively broad, and hence, unreadable in the screenshot, but in ANNIS, it can be zoomed and studied in detail.

For reasons of space, we restrict ourselves to these visualizations here. The visualization system of ANNIS is modular and extensible, and within our project, we have developed additional visualizations, including a generic graph visualization, several visualizations for dependency trees, etc.

While convenient visualizations allow in-depth qualitative analyses of different layers of annotations and their interdependencies, ANNIS also provides means for quantitative analyses: Query results can be exported and processed with tools for statistical analysis and data mining [9]. Another simple way to evaluate multi-layer annotations is by means of the match count for different queries. The query in Fig. 5 provides us with the frequency of +pronoun_anaph, +subject_antec, +satellite_antec. With similar queries for other parameter combinations,[5] the results in Tab. 1 can be obtained for the ProCon10 corpus. Although such a small corpus does not allow for a proper statistical evaluation, we can already observe some tendencies. From the literature, we would expect that +subject_antec and −satellite_antec correlate with +pron, whereas −pron would be more likely for −subject_antec and +satellite_antec. This general picture seems to be confirmed for ±subject_antec only (20 of 34 matches correspond to the prediction), but not for ±satellite_antec (only 14 of 34 correspond to the prediction, for combinations of ±satellite_antec with ±subject_antec 16 of 34).

An evaluation of the entire PCC by means of these (and related) queries may be conducted after the completion of the currently on-going revision of the corpus, this will be subject to another publication.

[5] For −pronoun_anaph, replace line 2 by `tiger:cat="NP" & #1 _=_ #2`,
for −subject_antec, replace line 5 by `tiger:cat="S" & #5 >[func=/[^S].*/] #4`,
for −satellite_antec, replace line 7 by `cat & #7 >[func=/.*NUC/] #6`.

Table 1. Influence of ±subject_antec (SBJ) and ±satellite_antec (SAT) on ±pronoun_anaph (PRN)

	+SBJ	−SBJ	−SAT	+SAT	+SBJ and +SAT	−SBJ or +SAT
+PRN	17	0	10	7	10	7
−PRN	14	3	13	4	11	6

5 Results and Discussion

We have shown how the linguistic information system ANNIS can be applied to the joint querying of coreference annotations and multiple other layers of linguistic annotation, and propose it as a tool well-suited to conduct qualitative and quantitative corpus studies of reference and related phenomena.

Among corpus search engines with database and graphical visualization that have been applied to the joint querying of relational (especially anaphoric) and hierarchical annotations, ANNIS is outstanding with respect to its genericity. For the query in Fig. 5, for example, other systems would require either a counter-intuitive modelling of linguistic annotations or structural modifications of annotation layers:

- MMAX2 [21] and GATE [11] support only markables, no structs, i.e., these systems lack a formally defined notion of dominance. However, lines 5 and 7 would not be possible if pointing relations would be abused to represent dominance relations.[6]
- Like ANNIS, TrED [23] and the CDTB annotation tool [18] can combine discourse-structural and/or coreference annotations with syntactic annotations. However, it seems that both are not able to support conflicting hierarchies and operators for co-extensionality. Lines 2, 4 and 6 would only be possible if multiple hierarchical annotations are unified into a single hierarchy and the coreference annotation uses the same markables as the syntax.

Subsequent research may include the application of ANNIS for the corpus-driven investigation of the interplay between discourse structure, information structure and coreference. Earlier studies on the Potsdam Commentary Corpus that were conducted by our group, e.g., [8], suffered from the inavailability of a corpus management system capable to process these different layers simultaneously. With the improved functionality of ANNIS, this research can be taken up again. Another obvious extension is the evaluation of further resources that combine coreference and discourse structure, e.g., the intersection between the Penn Discourse Treebank [24] with the OntoNotes corpus [25] or the RST Discourse Treebank [5] with the coreference annotations of the RefRhet corpus [17].

[6] Dominance can be heuristically approximated by pointing relations between markables that include each other. This approximation is, however, non-satisfying if an annotation scheme uses pointing relations between elements that indirectly dominate each other. This occurs, for example, with secondary edges in TIGER [3].

As ANNIS allows to represent and to work with multiple logically independent annotation layers with any kind of hierarchical and relational annotation, it would also be possible to compare different types of discourse structure annotation applied to the same corpus, e.g., the respective intersections between the RST Discourse Treebank, the Penn Discourse Treebank and the Penn Discourse Graph Bank [29].

These are just two of several fields of research where ANNIS could be applied in a fruitful way and that – to our best knowledge – would not have been possible with any other linguistic information system in existence.

References

1. Bird, S., Liberman, M.: A formal framework for linguistic annotation. Speech Communication 33(1-2), 23–60 (2001)
2. Bouma, G.: Querying linguistic corpora with prolog. In: Proceedings of the 10th Conference on Natural Language Processing (KONVENS-2010). p. 9. Saarbrücken, Germany (Sep 2010)
3. Brants, S., Dipper, S., Eisenberg, P., Hansen, S., König, E., Lezius, W., Rohrer, C., Smith, G., Uszkoreit, H.: TIGER: Linguistic interpretation of a German corpus. Research on Language and Computation 2(4), 597–620 (2004)
4. Carletta, J., Evert, S., Heid, U., Kilgour, J., Robertson, J., Voormann, H.: The NITE XML Toolkit: flexible annotation for multi-modal language data. Behavior Research Methods, Instruments, and Computers 35(3), 353–363 (2003)
5. Carlson, L., Marcu, D., Okurowski, M.E.: Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In: van Kuppevelt, J., W. Smith, R. (eds.) Current and New Directions in Discourse and Dialogue, pp. 85–112. Kluwer, Dordrecht (2003)
6. Chiarcos, C., Dipper, S., Götze, M., Leser, U., Lüdeling, A., Ritz, J., Stede, M.: A flexible framework for integrating annotations from different tools and tagsets. TAL (Traitement automatique des langues) 49 (2008)
7. Chiarcos, C., Krasavina, O.: PoCoS - Potsdam Coreference Scheme. In: Proceedings of the Linguistic Annotation Workshop (LAW). pp. 156–163. Prague (2007)
8. Chiarcos, C., Krasavina, O.: Rhetorical Distance revisited - A parameterized approach. In: Benz, A., Kühnlein, P. (eds.) Constraints in Discourse, pp. 97–116. John Benjamins, Amsterdam (2008)
9. Chiarcos, C., Ritz, J.: Qualitative and quantitative error analysis in context. In: Proceedings of the 10th Conference on Natural Language Processing (KONVENS-2010). pp. 111–117. Saarbrücken, Germany (Sep 2010)
10. Chiarcos, C., Ritz, J., Stede, M.: By all these lovely tokens ... merging conflicting tokenizations. Journal of Language Resources and Evaluation (LREJ) (accepted)
11. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A framework and graphical development environment for robust NLP tools and applications. In: Proceedings of ACL-2002. pp. 168–175. Philadelphia, USA (2002)
12. Grosz, B., Sidner, C.: Attention, intentions, and the structure of discourse. Computational linguistics 12(3), 175–204 (1986)
13. Grosz, B., Weinstein, S., Joshi, A.: Centering: A framework for modeling the local coherence of discourse. Computational linguistics 21(2), 203–225 (1995)

14. Ide, N., Suderman, K.: Bridging the gaps: Interoperability for GrAF, GATE, and UIMA. In: Proceedings of the Third Linguistic Annotation Workshop. pp. 27–34. Association for Computational Linguistics (2009)
15. Ide, N., Romary, L., de la Clergerie, E.: International Standard for a Linguistic Annotation Framework. In: Proceedings of HLT-NAACL'03 Workshop on the Software Engineering and Architecture of Language Technology. pp. 25–30. Edmonton, Canada (2003)
16. Ide, N., Suderman, K.: GrAF: A graph-based format for linguistic annotations. In: Proceedings of The Linguistic Annotation Workshop (LAW) 2007. pp. 1–8. Prague (2007)
17. Khudyakova, M.V., Dobrov, G.B., Kibrik, A.A., Loukachevitch, N.V.: Computational modeling of referential choice: Major and minor referential options. In: Proceedings of the CogSci 2011 Workshop on the Production of Referring Expressions. Boston (July 2011)
18. Korzen, I., Buch-Kromann, M.: Anaphoric relations in the Copenhagen Dependency Treebanks. In: Proceedings of Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena. pp. 83–98. Göttingen, Germany (Feb 2011)
19. Lezius, W.: TIGERSearch - Ein Suchwerkzeug für Baumbanken. In: Proceedings of the 6. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2002). pp. 107–114. Saarbrücken, Germany (Oct 2002)
20. Mann, W., Thompson, S.: Rhetorical Structure Theory: Toward a functional theory of text organization. Text 8(3), 243–281 (1988)
21. Müller, C., Strube, M.: Multi-level annotation of linguistic data with MMAX2. In: Braun, S., Kohn, K., Mukherjee, J. (eds.) Corpus Technology and Language Pedagogy. Peter Lang, Frankfurt a.M., Germany (2006)
22. O'Donnell, M.: RSTTool 2.4 – A markup tool for Rhetorical Structure Theory. In: Proceedings of the International Natural Language Generation Conference (INLG'2000). pp. 253–256. Mitzpe Ramon, Israel (2000)
23. Pajas, P., Štěpánek, J.: System for querying syntactically annotated corpora. In: Proceedings of the ACL-IJCNLP 2009 Software Demonstrations. pp. 33–36 (2009)
24. Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., Webber, B.: The Penn Discourse Treebank 2.0. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008). Marrakech, Morocco (2008)
25. R. Weischedel, S. Pradhan, L.R., Micciulla, L.: OntoNotes Release 1.0 (2006)
26. Rehm, G., Schonefeld, O., Witt, A., Chiarcos, C., Lehmberg, T.: SPLICR: A sustainability platform for linguistic corpora and resources. In: Proceedings of the 9th Conference on Natural Language Processing (KONVENS 2008). pp. 86–95. Berlin (Sep 2008)
27. Stede, M.: The Potsdam Commentary Corpus. In: Proceedings of the ACL Workshop on Discourse Annotation. pp. 96–102. Barcelona, Spain (2004)
28. Telljohann, H., Hinrichs, E.W., Kübler, S.: Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Tech. rep., Seminar für Sprachwissenschaft, Universität Tübingen, Tübingen, Germany (2003)
29. Wolf, F., Gibson, E.: Representing discourse coherence: A corpus-based analysis. Computational Linguistics 31, 249–287 (2005)
30. Zeldes, A., Ritz, J., Lüdeling, A., Chiarcos, C.: ANNIS: A search tool for multi-layer annotated corpora. In: Proceedings of Corpus Linguistics 2009. Liverpool, UK (July 2009)