The SynDiKATe Text Knowledge Base Generator

Udo Hahn
CLIF Text Knowledge Engineering Lab
Albert-Ludwigs-Universität Freiburg
D-79085 Freiburg, Germany
hahn@coling.uni-freiburg.de

Martin Romacker
CLIF Text Knowledge Engineering Lab
Albert-Ludwigs-Universität Freiburg
D-79085 Freiburg, Germany
romacker@coling.uni-freiburg.de

ABSTRACT

SynDiKATe comprises a family of text understanding systems for automatically acquiring knowledge from real-world texts, *viz.* information technology test reports and medical finding reports. Their content is transformed to formal representation structures which constitute corresponding text knowledge bases. SynDiKATe's architecture integrates requirements from the analysis of single sentences, as well as those of referentially linked sentences forming cohesive texts. Besides centering-based discourse analysis mechanisms for pronominal, nominal and bridging anaphora, SynDiKATe is supplied with a learning module for automatically bootstrapping its domain knowledge as text analysis proceeds.

1. INTRODUCTION

The SynDiKATe system belongs to the broad family of information extraction (IE) systems [1]. Significant progress has been made already, as current IE systems provide robust shallow text processing such that frame-style templates are filled with factual information about particular entities (locations, persons, event types, etc.) from the analyzed documents. Nevertheless, typical MUC-style systems are also limited in several ways. They provide no inferencing capabilities which allow substantial reasoning about the template fillers (hence, their understanding depth is low), and their potential to deal with textual phenomena is highly constrained, if it is available at all. Also novel and unexpected though potentially relevant information which does not match given template structures is hard to account for, since system designers commit to a fixed collection of domain knowledge templates (i.e., they have no concept learning facilities).

With SynDiKATe, we are addressing these shortcomings and aim at a more sophisticated level of knowledge acquisition from real-world texts. The documents we deal with are technical narratives in German language taken from two domains, *viz.* test reports from the information technology (IT) domain as processed by the itSynDiKATe system [8], and finding reports from a medical subdomain (MED), the framework of the medSynDiKATe system [10, 9]. Our first goal is to extract conceptually and inferentially richer forms of knowledge than those captured by standard IE systems such as evaluative assertions and comparisons [25, 24], temporal [26] and spatial information [22]. Second, we also want to dynamically enhance the set of knowledge templates through incremental taxonomy learning devices [12] so that the information extraction capability of the system is increased in a bootstrapping manner. Third, SynDiKATe is particularly sensitive to the treatment of textual reference relations [27, 6, 14]. The capability to properly deal with various forms of anaphora is a prerequisite for the soundness and validity of the knowledge bases we create as a result of the text understanding process and likewise for the feasibility of sophisticated retrieval and question answering applications based on the acquired text knowledge.

2. SYSTEM ARCHITECTURE

The overall architecture of SynDiKATe, an acronym which stands for "SYNthesis of DIstributed Knowledge Acquired from TExts", is summarized in Figure 1. Incoming texts, T_i, are mapped into corresponding *text knowledge bases*, TKB_i, which contain a representation of T_i's content. This knowledge base platform may feed various information services, such as inferentially supported question answering (fact retrieval), text passage retrieval or text summarization [7].

2.1 Sentence-Level Understanding

Grammatical knowledge for syntactic analysis is based on a fully lexicalized dependency grammar [11], we refer to as *Lexicon* in Figure 1. Basic word forms (lexemes) constitute the leaf nodes of the lexicon tree, which are further abstracted in terms of a hierarchy of lexeme class specifications at different levels of generality. The *Generic Lexicon* in Figure 1 contains lexical material which is domain-independent (lexemes such as *move*, *with*, or *month*), while domain-specific extensions are kept in specialized lexicons serving the needs of particular subdomains, e.g., IT (*hard disk*, *color printer*, etc.) or MED (*gastritis*, *surface mucus*, etc.). Dependency grammars capture binary valency constraints between a syntactic head (e.g., a noun) and possible modifiers (e.g., a determiner or an adjective). To establish a dependency relation between a head and a modifier, all the lexicalized constraints on word order, compatibility of morphosyntactic features, and semantic criteria must be fulfilled. This leads to a strictly local computation scheme which inherently lends itself to robust partial parsing [5].

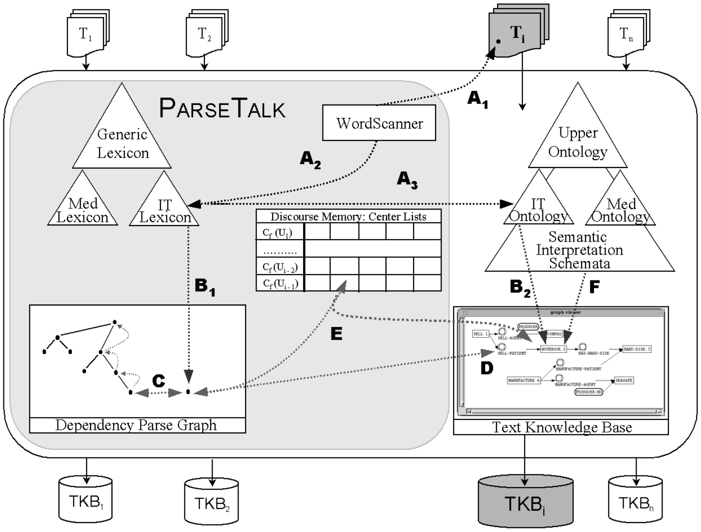

Figure 1: Architecture of a SynDiKATe System

Conceptual knowledge about the different domains is expressed in a Kl-One-like description logic language [28]. Corresponding to the division at the lexical level, the ontologies we provide are split up between one that is used by all applications, the *Upper Ontology*, while several dedicated ontologies account for the conceptual requirements of particular domains, e.g., IT (HardDisk, ColorPrinter, etc.) or MED (Gastritis, SurfaceMucus, etc.).

Semantic knowledge accounts for emerging conceptual relations between conceptual items according to those dependency relations that are established between their corresponding lexical items. Semantic interpretation schemata mediate between both levels in a way as abstract and general as possible [20]. These schemata are applied to *semantically interpretable subgraphs* which are, from a semantic point of view, "minimal" subgraphs of the incrementally built dependency graph. Their bounding nodes contain content words (i.e., nouns, verbs, and adjectives, all of which have a conceptual correlate in the domain ontologies), while all possibly intervening nodes (zero up to four) contain only noncontent words (such as prepositions, articles, auxiliaries, etc., all of which have no conceptual correlates). Semantic interpretation schemata are fully embedded in the knowledge representation model and system (cf. Figure 1).

The ParseTalk system, which comprises the lexicalized grammar and associated dependency parser, is embedded in an object-oriented computation model. So, the dependency relations are computed by lexical objects, so-called *word actors*, through strictly local message passing, only involving the lexical items they represent. To illustrate how a dependency relation is established computationally, we give a sketch of the basic protocol for incremental **parsing** [5]:

- After a word has been read from textual input by the *WordScanner* (step A_1 in Figure 1), its associated lexeme (specified in the *Lexicon*) is identified (step A_2) and a corresponding word actor gets initialized (step B_1). As all content words are directly linked to the conceptual system, each lexical item w that has a conceptual correlate C in the domain knowledge base (step A_3) gets instantiated in the text knowledge base (step B_2). The lexical item `Festplatte` (*hard disk*) with the conceptual correlate Hard-Disk is instantiated, e.g., by Hard-Disk.3, the particular item being talked about in a given text.[1]
- For integration in the parse tree, the newly created word actor searches its head (alternatively, its modifier) by sending parallel requests for dependential government to its left context (step C). The search space is restricted, since these requests are propagated upwards only along the 'right shoulder' of the dependency graph constructed so far. All word actors addressed this way check, in parallel, whether their valency restrictions, i.e., grammatical and conceptual constraints, are met by the requesting word actor. Step D simulates a conceptual check in the text knowledge base, step E illustrates a test in the discourse memory.
- If all required constraints are fulfilled by one of the targeted word actors, an immediate semantic interpretation is performed. This usually alters the conceptual representation structures by way of slot filling (step F).

Semantic interpretation consists of finding a relational link between the conceptual correlates of the two content words bounding the associated semantically interpretable subgraph. The linkage may either be constrained by dependency relations (e.g., the *subject:* relation of a transitive verb such as *"sell"* may only be interpreted conceptually in terms of agent or patient roles), by intervening lexical material (e.g., some prepositions impose special role constraints, such as `mit` (*with*) does in terms of has-part or instrument roles), or it may be constrained by conceptual criteria only (as with the *genitive:* dependency relation, which unlike *subject:* imposes no additional selective conceptual constraints for interpretation). The corresponding knowledge about these language-specific constraints is densely encoded in the *Lexicon* class hierarchy, an approach which heavily relies on the property inheritance mechanisms inherent to the object-oriented paradigm.

2.2 Text-Level Understanding

2.2.1 Referential Text Phenomena

The textual phenomena we deal with in SynDiKATe establish referential links between consecutive utterances in a coherent text such as illustrated by three possible continuations of sentence (1), with three different forms of extrasentential anaphora:

(1) Compaq verkauft *ein Notebook* mit einer Festplatte, die von Seagate hergestellt wird.
(Compaq sells *a notebook* with a hard disk that is manufactured by Seagate.)

(2) **Pronominal Anaphora:**
Es ist mit einer Pentium-III-CPU ausgestattet.
(*It* comes with a Pentium-III CPU.)

(3) **Nominal Anaphora:**
Der Rechner ist mit einer Pentium-III-CPU ausgestattet.
(*The machine* comes with a Pentium-III CPU.)

(4) **Functional Anaphora:**
Der Arbeitsspeicher kann auf 96 MB erweitert werden.
(*The main memory* can be expanded up to 96MB.)

[1]Due to the recognition of referential relations at the text level of analysis this instantiation might be readjusted by subsequent coreference declarations (cf. Section 2.2).

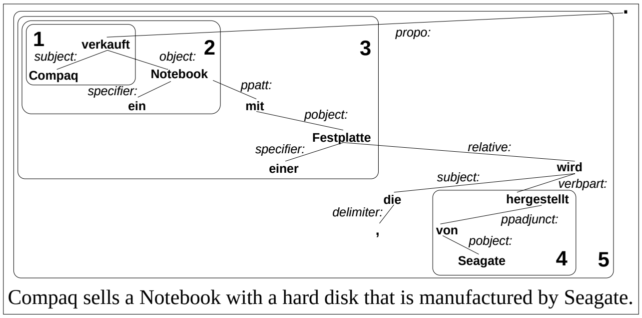

Figure 2: Dependency Parse for Sentence (1)

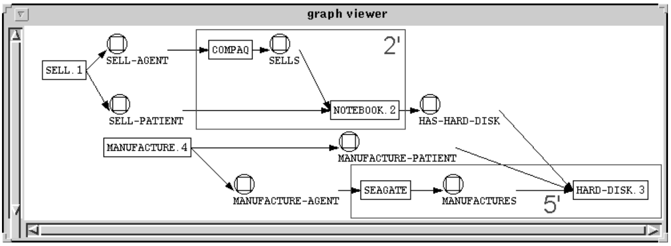

Figure 3: Conceptual Interpretation for Sentence (1)

The results of sentence-level analysis for sentence (1) are given in Figure 2, which contains a syntactic dependency graph (together with five configurations of semantically interpretable subgraphs), and Figure 3, which displays its conceptual representation. For text-level analysis, pronominal anaphora still heavily depend on grammatical conditions – the agreement of the antecedent (*"Notebook"*) and the pronoun (*"Es" (it)*) in gender and number; also conceptual criteria apply insofar as a potential antecedent must fit the conceptual role (or case frame) restrictions when it is integrated in governing structures, say, the head verb of the clause. In general, however, the influence of grammatical criteria gradually diminishes for other types of text phenomena, while the influence of conceptual criteria increases. For nominal anaphora, number constraints are still valid, while a generalization relation between the anaphoric noun (*"Rechner" (machine)*) and its proper antecedent (*"Notebook"*) must hold, in addition. In the case of functional anaphora, no grammar constraints at all apply, while quite sophisticated conceptual role path conditions come into play, e.g., *"Arbeitsspeicher" (main memory)* being a constituent physical part of *"Notebook"*.

The problems text phenomena cause are of vital importance for the adequacy of the representation structures resulting from text processing, and are centered around the notions of incomplete, invalid and incoherent knowledge bases.

Incomplete knowledge bases emerge when references to already established discourse entities are simply not recognized, as in the case of *pronominal anaphora*. Consider the reference relationship between the pronoun *"Es" (it)* in sentence (2) which refers to the noun phrase *"ein Notebook" (a notebook)* in sentence (1). The occurrence of the pronoun is not reflected at the conceptual level, since pronouns (as noncontent words) do not have conceptual correlates. Hence, an incomplete concept graph emerges as shown in Figure 4 — the referent for the pronoun *"Es" (it)*, NOTEBOOK.2, is not linked to PENTIUM-III-CPU.6. An adequate treatment with a properly resolved anaphor is shown in Figure 6, where the representation of the relevant portions of sentence (1) is linked to the one of sentence (2), in particu-

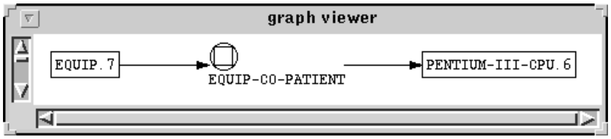

Figure 4: Unresolved Pronominal Anaphor, Sentence (2)

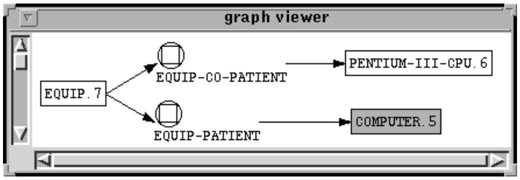

Figure 5: Unresolved Nominal Anaphor, Sentence (3)

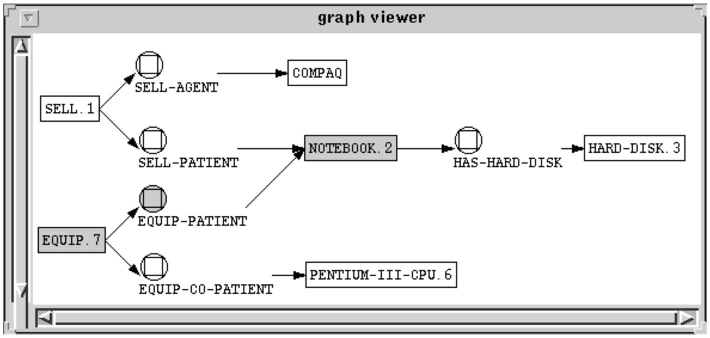

Figure 6: Resolved Anaphors, Sentences (1) and (2)/(3)

lar by determining the EQUIP-PATIENT role between EQUIP.7 and the proper referent, NOTEBOOK.2.

Invalid knowledge bases emerge when each entity which has a different denotation at the text surface is treated as a formally distinct conceptual item at the symbol level of knowledge representation, although all different denotations refer literally to the same conceptual entity. This is the case for *nominal anaphora*, an example of which is given by the reference relation between the noun phrase *"Der Rechner" (the machine)* in sentence (3) and the noun phrase *"ein Notebook" (a notebook)* in sentence (1). An invalid referential description appears in Figure 5, where COMPUTER.5 is introduced as a new entity in the discourse, whereas Figure 6 shows the valid conceptual representation capturing the intended meaning at the representation level, *viz.* maintaining NOTEBOOK.2 as the proper referent (note that pronominal as well as nominal anaphora are two equivalent ways to *co*refer to the discourse entity denoted by NOTEBOOK.2).

Finally, *incoherent* knowledge bases emerge when entities which are linked by nontaxonomic conceptual relations at the knowledge level occur in a text such that an implicit reference to these relations can be made in the text source. Unlike the previously discussed cases of coference, these relations have to be made explicit at the symbol level of the targeted text knowledge base by a search for connecting paths between the concepts involved [6]. This is the basic scenario for *functional (or bridging) anaphora*. Consider, e.g., the relationship holding between the noun phrase *"Der Arbeitsspeicher" (the main memory)* in sentence (4), which refers to the noun phrase *"ein Notebook" (a notebook)* in sentence (1). In Figure 8 the relational link missing in Figure 7 between MAIN-MEMORY.8 and NOTEBOOK.2 is established (via a HAS-PART-type relation, *viz.* HAS-MAIN-MEMORY), and, hence, representational coherence at the symbol level of knowledge representation is preserved.

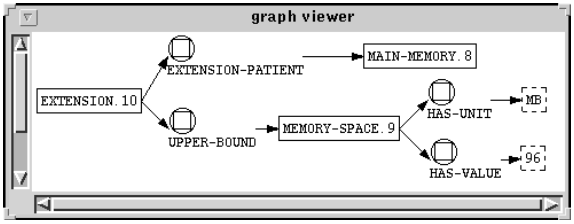

Figure 7: Unresolved Functional Anaphor, Sentence (4)

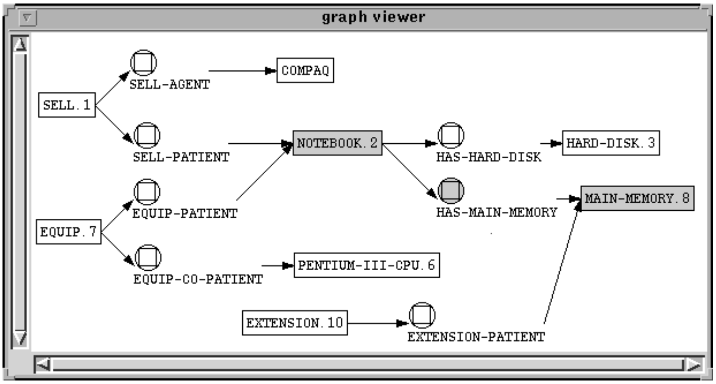

Figure 8: Resolved Functional Anaphor, Sentences (1) and (4)

Disregarding textual phenomena will cause dysfunctional system behavior. A query `Q` such as

```
Q :  (retrieve ?x (Computer ?x))
A-:  (|I| Notebook.2, |I| Computer.5)
A+:  (|I| Notebook.2)
```

triggers a search for all instances of COMPUTER in the text knowledge base. Given an invalid knowledge base (cf. Figures 3 and 5), the incorrect answer (A-) contains two entities, *viz.* NOTEBOOK.2 and COMPUTER.5 — both are in the extension of the concept COMPUTER. If, however, a valid text knowledge base such as the one in Figure 6 or 8 is given, only the correct answer, NOTEBOOK.2, is inferred (A+).

Rendering also quantitative substance to our claims, we analyzed a randomly chosen sample of 100 reports on histological findings with approximately 14,000 text tokens [9]. In IT texts, (pro)nominal anaphora and functional anaphora occur at an almost balanced rate [27]. In the medical texts, however, functional anaphora turn out to be the major glue for establishing local coherence, while anaphora, pronominal anaphora in particular, play a far less important role than in other text genres. The high proportion of functional anaphora (45%) [42%-48%][2] and the remarkable rate of nominal (34%) [31%-37%] compared to extrasentential pronominal anaphora (2%) [1%-3%] is clearly an indication of the primary orientation in medical texts to convey facts in a very compact manner. Two consequences can be drawn from this observation. First, resolution procedures for functional anaphora – supplementing well-researched procedures for (pro)nominal anaphora – have to be provided urgently (cf. [6] for a fully worked out approach). Second, functional anaphora presuppose a considerable amount of deep background knowledge, with emphasis on partonomic reasoning [13], supplementing well-known principles of taxonomic reasoning for text understanding.

[2]For all percentage numbers 95% confidence intervals are supplied in square brackets.

2.2.2 *Centering Model for Anaphora Resolution*

In order to avoid the emergence of incomplete, invalid and incoherent text knowledge bases we consider discourse entities for establishing reference relations with upcoming items from the textual input at a local [27] and at a global level [14] of cohesion. To preserve adequate text representation structures a *centering* mechanism is used. The discourse entities which occur in an utterance U_i constitute its set of *forward-looking centers*, $C_f(U_i)$. The elements in $C_f(U_i)$ are ordered to reflect relative prominence in U_i in the sense that the most highly ranked element of $C_f(U_i)$ is the most likely antecedent of an anaphoric expression in U_{i+1}, while the remaining elements are ordered according to decreasing preference for establishing referential links.

While it is usually assumed (for the English language, in particular) that *grammatical* roles are the major determinant for the ranking on the C_f [4], we claim that for German – a language with relatively free word order – it is the *functional* information structure of the sentence [27]. Accordingly, the constraints on the ordering of entries in $C_f(U_i)$ prefer *hearer-old* (either evoked or unused) elements in an utterance (i.e., those that can be related to previously introduced discourse elements or generally accessible world knowledge) over *mediated* (inferrable) ones, while these are preferred over *hearer-new* (brand-new) elements for anaphora resolution. If two elements belong to the same category, then preference is defined in terms of linear precedence of the discourse units in the source text.

When we apply these criteria to sentence (1), Table 1 depicts the resulting order of forward-looking centers in $C_f(S_1)$. Since we have no discourse-bound elements in the first sentence, textual precedence applies exclusively to the ordering of the center list items. Only nouns and their conceptual correlates are taken into consideration. The tuple notation takes the conceptual correlate of the lexical item in the text knowledge base in the first place, while the lexical surface form appears in the second place.

(1)	**Cf:**	[COMPAQ: Compaq, NOTEBOOK.2: Notebook, HARD-DISK.3: Festplatte, SEAGATE: Seagate]

Table 1: Centering Data for Sentence (1)

Processing of the centering list $C_f(S_1)$ for sentence (3) until the generalization constraint is fulfilled, finally, results in a query whether NOTEBOOK is subsumed by COMPUTER, the conceptual correlate of the lexical item *"Rechner"*. As this relationship obviously holds, in the conceptual representation structure of sentence (3) (cf. Figure 5) COMPUTER.5, the literal instance identifier, is declared coreferent to NOTEBOOK.2, the referentially valid identifier. Instead of having two unlinked sentence graphs, Figures 3 and 5, the reference resolution for (pro)nominal anaphora leads to joining them in a common valid text graph (Figure 6). In particular, NOTEBOOK.2 links to the relation EQUIP-PATIENT, formerly occupied by COMPUTER.5. The corresponding centering list at the end of the analysis of sentence (3) is provided in Table 2 ($C_f(S_1)$ has been updated to reflect the consumption of the antecedent, NOTEBOOK.2, in the processing of $C_f(S_3)$).

(1)	**Cf:**	[COMPAQ: Compaq, ~~NOTEBOOK.2: Notebook,~~ HARD-DISK.3: Festplatte, SEAGATE: Seagate]
(3)	**Cf:**	[NOTEBOOK.2: Rechner, PENTIUM-III-CPU.6: Pentium-III-CPU]

Table 2: Centering Data for Sentences (1) and (3)

2.3 Textual Learning

The approach to learning new concepts as a result of text understanding builds on two different sources of evidence — the prior knowledge of the domain the texts are about, and grammatical constructions in which unknown lexical items occur in the texts. The architecture of SynDiKATe's concept learning component is depicted in Figure 9.

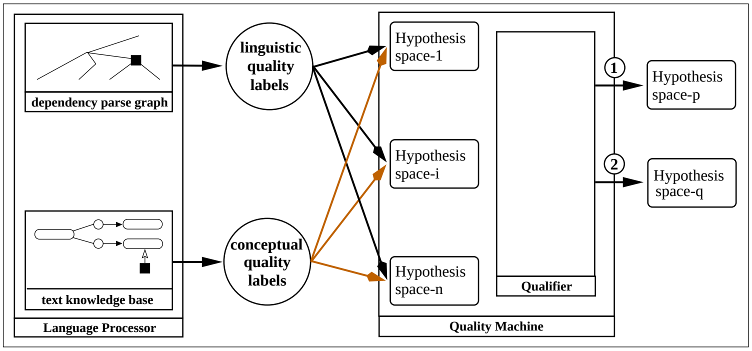

Figure 9: SynDiKATe**'s Learning Component**

The ParseTalk system generates *dependency parse graphs*. The kinds of syntactic constructions (e.g., genitive, appositive, comparative), in which unknown lexical items appear, are recorded and later assessed relative to the credit they lend to a particular concept hypothesis, e.g., high for appositives (*"the notebook X"*), lower for genitives (*"Compaq's X"*). The conceptual interpretation of parse trees involving unknown lexical items in the *text knowledge base* leads to the deduction of *concept hypotheses*. These are further enriched by conceptual annotations which reflect structural patterns of consistency, mutual justification, analogy, etc. relative to already available concept descriptions in the text knowledge base or other hypothesis spaces. Both kinds of evidence, in particular their predictive 'goodness' for the learning task, are represented by corresponding sets of *linguistic* and *conceptual quality labels*.

Alternative concept hypotheses for each unknown lexical item are organized in terms of corresponding *hypothesis spaces*, each of which holds a different conceptual reading. An inference engine embedded in the terminological system, the so-called *quality machine*, determines the overall credibility of single concept hypotheses by taking the available set of quality labels for each hypothesis into account. The *qualifier*, a terminological classifier extended by an evaluation metric for quality classes, computes a preference ranking of those hypotheses which remain valid after the text has been processed completely (cf. [12] for details).

3. COVERAGE AND EVALUATION

SynDiKATe's coverage varies considerably depending on the target domain. The generic lexicon currently includes 3,000 entries, the IT lexicon adds 5,000, while the MED lexicon contributes 70,000 entries each. The Upper Ontology contains 1,200 concepts and roles, to which the IT ontology adds 3,000 and the MED ontology contributes 240,000 items.

The IT domain was chosen as a testbed that can be extended on demand. The MED domain, however, is subject to ontology engineering efforts on a larger scale. In order to cope with the enormous knowledge engineering requirements, we semi-automatically transformed large portions of a semantically weak, yet high-volume medical terminology (UMLS) to a very large terminological knowledge base [21].

Admittedly, SynDiKATe has not yet undergone a thorough empirical evaluation in one of the envisaged application dimensions. We have, however, carefully evaluated its subcomponents. The results can be summarized as follows:

Sentence Parsing. We compared a standard active chart parser with full backtracking capabilities with the parser of SynDiKATe, which is characterized by limited memoization and restricted backtracking capabilities, using the same grammar specifications. On average, SynDiKATe's parser exhibits a linear time complexity the factor of which is dependent on ambiguity rates of input sentences. The active chart parser runs into exponential time complexity whenever it encounters extragrammatical or ungrammatical input, since then it conducts an exhaustive search of the entire parse space. The loss of structural descriptions due to the parser's incompleteness amounts to 10% compared with the complete, though intractable parser [5].

Text Parsing. While with respect to resolution capacity (effectiveness) no significant differences could be determined, the functional centering model we propose outperforms the best-known centering algorithms by a rate of 50% with respect to a measure of computation costs which considers "cheap" and "expensive" transitional moves between utterances to assess a text's coherency. Hence, the procedure we propose is more efficient [27].

Semantic Interpretation. Our group has been pioneering work on the empirical evaluation of meaning representations. We assessed the quality and coverage of semantic interpretation for randomly sampled texts in the two domains we consider. While recall was rather low (57% for MED, 31% for IT), precision peaked at 97% and 94%, respectively [19].

"Heavy" Semantics. We can deal with intricate semantic phenomena for which we have provided the first empirical evaluation data available at all. This relates to the resolution of metonymies, where we have determined a gain in effectiveness that amounts to 16% compared with the best procedures known so far [16], as well as it relates to comparatives and evaluative assertions, where gains in effectiveness were almost tripled [25].

Concept Learning. The performance of the concept learning component has been compared to standard learning mechanisms based on the terminological classifier available in any sort of description logics systems. Our data indicate an increase of performance of 8% (87% accuracy, while that of standard classifiers is on the order of 79%) [12].

Evaluating a text knowledge acquisition rather than an IE system poses hard methodological problems [2]. The main reason being that a gold standard for comparison — what constitutes a canonical, commonly agreed upon interpretation of the content of a text? — is hard to establish, even for technical texts. A follow-up problem is constituted by the lack of a significant amount of annotated text knowledge bases on which comparative analyses might be assessed. MUC-style evaluation metrics, e.g., have already been qualified not to adequately reflect the functionality of less constrained text understanders [29].

4. CONCLUSIONS

A major hypothesis underlying the design of SynDiKATe is that ignoring the referential relations between adjacent utterances will lead to referentially incomplete, invalid, or incoherent text knowledge bases. We determine plausible discourse units for reference resolution using the centering model. This allows us to deal with various forms of pronominal, nominal and functional anaphora in a uniform way.

In order to establish local coherence at the text representation level, single discourse entities related by anaphoric expressions have to be conceptually linked. We claim that only sophisticated knowledge representation languages with powerful terminological reasoning capabilities, such as those from the KL-ONE family, are able to deal with the full range of challenges of referentially adequate text understanding, in particular considering nominal and functional anaphora.

These two types of anaphora pose an enormous burden on the availability of rich domain knowledge. We respond to this challenge in two ways. In a large-scale knowledge engineering effort, we semi-automatically transform a semantically weak though huge thesaurus-style medical knowledge source into a terminological knowledge base. If such a human-made resource is missing, we turn to a purely automatic approach of bootstrapping a given domain knowledge base as part of on-going text understanding processes.

The depth of understanding we provide comes closest to systems such as SCISOR [18], TACITUS [15] or PUNDIT/KERNEL [17], but SYNDIKATE's knowledge acquisition strategies or learning capabilities have no counterpart there. Text understanders which incorporate learning components are even rarer but systems such as SNOWY [3] or WRAP-UP [23] either have a very narrow domain theory and lack robustness for dealing with unseen input effectively, or fail to account for a wide range of referential text phenomena, respectively.

5. ACKNOWLEDGMENTS

The development of the SYNDIKATE system has been supported by various grants from *Deutsche Forschungsgemeinschaft* under Ha 2097/*. SYNDIKATE would not have come to existence without the exciting contributions and enthusiasm of current and former members of the CLIF group, in particular, Steffen Staab, Katja Markert, Michael Strube, Martin Romacker, Stefan Schulz, Klemens Schnattinger, Norbert Bröker, Peter Neuhaus, Susanne Schacht, Manfred Klenner, and Holger Schauer.

6. REFERENCES

[1] Jim Cowie and Wendy Lehnert. Information extraction. *Communications of the ACM*, 39(1):80–91, 1996.

[2] Carol Friedman and George Hripcsak. Evaluating natural language processors in the clinical domain. *Methods of Information in Medicine*, 37(4/5):334–344, 1998.

[3] Fernando Gomez and Carlos Segami. The recognition and classification of concepts in understanding scientific texts. *Journal of Experimental and Theoretical Artificial Intelligence*, 1(1):51–77, 1989.

[4] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.

[5] Udo Hahn, Norbert Bröker, and Peter Neuhaus. Let's PARSETALK: Message-passing protocols for object-oriented parsing. In H. Bunt and A. Nijholt, editors, *Advances in Probabilistic and other Parsing Technologies*, pages 177–201. Kluwer, 2000.

[6] Udo Hahn, Katja Markert, and Michael Strube. A conceptual reasoning approach to textual ellipsis. In *Proceedings of the ECAI'96*, pages 572–576, 1996.

[7] Udo Hahn and Ulrich Reimer. Knowledge-based text summarization: Salience and generalization operators for knowledge base abstraction. In I. Mani and M. Maybury, editors, *Advances in Automatic Text Summarization*, pages 215–232. MIT Press, 1999.

[8] Udo Hahn and Martin Romacker. Content management in the SYNDIKATE system: How technical documents are automatically transformed to text knowledge bases. *Data & Knowledge Engineering*, 35(2):137–159, 2000.

[9] Udo Hahn, Martin Romacker, and Stefan Schulz. Discourse structures in medical reports – watch out! The generation of referentially coherent and valid text knowledge bases in the MEDSYNDIKATE system. *International Journal of Medical Informatics*, 53(1):1–28, 1999.

[10] Udo Hahn, Martin Romacker, and Stefan Schulz. How knowledge drives understanding: Matching medical ontologies with the needs of medical language processing. *Artificial Intelligence in Medicine*, 15(1):25–51, 1999.

[11] Udo Hahn, Susanne Schacht, and Norbert Bröker. Concurrent, object-oriented natural language parsing: The PARSETALK model. *International Journal of Human-Computer Studies*, 41(1/2):179–222, 1994.

[12] Udo Hahn and Klemens Schnattinger. Towards text knowledge engineering. In *Proceedings of the AAAI'98*, pages 524–531, 1998.

[13] Udo Hahn, Stefan Schulz, and Martin Romacker. Partonomic reasoning as taxonomic reasoning in medicine. In *Proceedings of the AAAI'99*, pages 271–276, 1999.

[14] Udo Hahn and Michael Strube. Centering in-the-large: Computing referential discourse segments. In *Proceedings of the ACL'97/EACL'97*, pages 104–111, 1997.

[15] Jerry R. Hobbs, Mark E. Stickel, Douglas E. Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1/2):69–142, 1993.

[16] Katja Markert and Udo Hahn. On the interaction of metonymies and anaphora. In *Proceedings of the IJCAI'97*, pages 1010–1015, 1997.

[17] Martha S. Palmer, Rebecca J. Passonneau, Carl Weir, and Tim Finin. The KERNEL text understanding system. *Artificial Intelligence*, 63(1/2):17–68, 1993.

[18] Lisa F. Rau, Paul S. Jacobs, and Uri Zernik. Information extraction and text summarization using linguistic knowledge acquisition. *Information Processing & Management*, 25(4):419–428, 1989.

[19] Martin Romacker and Udo Hahn. An empirical assessment of semantic interpretation. In *Proceedings of the NAACL 2000*, pages 327–334, 2000.

[20] Martin Romacker, Katja Markert, and Udo Hahn. Lean semantic interpretation. In *Proceedings of the IJCAI'99*, pages 868–875, 1999.

[21] Stefan Schulz and Udo Hahn. Knowledge engineering by large-scale knowledge reuse: Experience from the medical domain. In *Proceedings of KR 2000*, pages 601–610, 2000.

[22] Stefan Schulz, Udo Hahn, and Martin Romacker. Modeling anatomical spatial relations with description logics. In *Proceedings of the AMIA 2000*, pages 779–783, 2000.

[23] Stephen Soderland and Wendy Lehnert. Wrap-up: A trainable discourse module for information extraction. *Journal of Artificial Intelligence Research*, 2:131–158, 1994.

[24] Steffen Staab and Udo Hahn. Comparatives in context. In *Proceedings of the AAAI'97*, pages 616–621, 1997.

[25] Steffen Staab and Udo Hahn. "Tall", "good", "high" – compared to what? In *Proceedings of the IJCAI'97*, pages 996–1001, 1997.

[26] Steffen Staab and Udo Hahn. Scalable temporal reasoning. In *Proceedings of the IJCAI'99*, pages 1247–1252, 1999.

[27] Michael Strube and Udo Hahn. Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344, 1999.

[28] William A. Woods and James G. Schmolze. The KL-ONE family. *Computers & Mathematics with Applications*, 23(2/5):133–177, 1992.

[29] P. Zweigenbaum, J. Bouaud, B. Bachimont, J. Charlet, and J.-F. Boisvieux. Evaluating a normalized conceptual representation produced from natural language patient discharge summaries. In *Proceedings of the AMIA'97*, pages 590–594.