Automatic Semantic Analysis for NLP Applications

INGO GLÖCKNER, SVEN HARTRUMPF, HERMANN HELBIG,
JOHANNES LEVELING & RAINER OSSWALD

Abstract

In this article, we describe a long-term enterprise at the FernUniversität in Hagen to develop systems for the automatic semantic analysis of natural language. We introduce the underlying semantic framework and give an overview of several recent activities and projects covering natural language interfaces to information providers on the web, automatic knowledge acquisition, and textual inference.

Keywords: Natural Language Processing, Semantic Analysis, Inference, Text Retrieval, Question Answering

1. Introduction

Recent years have witnessed a rapid growth of electronically available textual information, strongly influenced by the advent of large publicly available resources on the World Wide Web (e. g. Wikipedia). As a consequence, there is an increasing demand for systems that support the retrieval of information in a flexible and user-friendly way, which ideally means that the user can submit his or her queries in natural language (NL). Current natural language processing (NLP) systems for text retrieval and question answering often rely on so-called shallow methods, which do not aim at a full syntactic or semantic analysis; in addition, they employ statistical measures defined on the surface structure of language, e. g. co-occurrence measures of various kinds. In spite of the astonishing success of these approaches for text retrieval, we believe that they are insufficient in the long run for any sort of non-trivial question answering task. The reason is that bridging the distance between a question and the text containing the answer usually calls for lexical and world knowledge as well as for advanced inference techniques. Moreover, any such system has to cope with all kinds of semantic phenomena occurring in unregimented NL.

Zeitschrift für Sprachwissenschaft 26 (2007), 241–266 0721-9067/2007/026–0241
DOI 10.1515/ZFS.2007.019

Within the formal semantics community, there is a large number of highly elaborated approaches for treating intricate semantic phenomena such as presuppositions, generics, pluralities or modalities within logical settings; see, for instance, the overview articles in van Benthem & ter Meulen (1997). However, a "grand unified" formal semantic theory covering all or at least most of these phenomena is far from being available, not to mention parsers that automatically generate such formal semantic representations from NL texts.[1] The logical frameworks currently studied in the knowledge representation community, on the other hand, are thoroughly investigated with respect to their logical properties but are far too restricted to cover a sufficient portion of the above mentioned semantic phenomena. In particular, this is true of the various versions of Description Logic (Baader et al. 2004) − which is not to say that they are not useful for strongly regimented NL; see e. g. Bernardi et al. (2007).

Requirements for application-oriented semantic analysis
If one aims at building NLP systems suitable for the deep semantic analysis of non-regimented, unrestricted text, then the underlying semantic representation framework has to satisfy a number of general criteria:[2] first of all, the semantic framework should be *universal* in that it is neither dependent on a specific language nor on a specific domain of discourse. Moreover, the framework should be (*descriptively*) *complete*, i. e. sufficiently expressive to cover all aspects of NL semantics (up to a certain degree of granularity). Since NLP systems for deep semantic interpretation combine several resources (lexicons, axiom sets) and methods (disambiguation, reasoning), the underlying semantic framework should be *homogeneous* and *interoperable;* that is, the same representational means should be employed for lexical, sentential, and textual semantics as well as for formalizing world knowledge, and they should be processable by a semantic parser, an inference engine, and an NL generator. Another key requirement is *communicability* because developing large NLP systems is a highly collaborative enterprise, especially with regard to the creation of lexical-semantic resources or knowledge bases; in other words, the semantic formalism should be intuitively intelligible and therefore should give rise to a high level of inter-coder agreement. Finally, the semantic representation formalism must allow for an efficient implementation.

1. An interesting case study in this context is Pulman (2007), which shows the typical problems one encounters when trying to implement a formal semanticist's solution to a specific phenomenon, in this case, adjectival comparatives, in a computational semantics framework
2. See Helbig (2006: 4 ff.) for a more elaborate exposition.

The MultiNet approach: a first glimpse
The approach to computational NL semantics presented in this article is based on the so-called *MultiNet* (*Multilayered Extended Semantic Networks*) paradigm, a formalism developed both for representing the semantics of NL expressions and for knowledge representation in general (Helbig 2006). The MultiNet formalism, described in more detail in Section 2., is intended to satisfy the above criteria to a large degree.

MultiNet represents the semantics of NL expressions by means of semantic networks, where nodes represent concepts and edges represent relations between concepts. To give an introductory example, consider the (simplified) MultiNet graph in Figure 1, which shows the meaning of the sentence *Der Prüfling beantwortete jede Frage des Professors* (*The examinee answered every question of the professor*). The example illustrates various key elements of the MultiNet formalism: there is a distinction between generic concepts and instantiated concepts, which are related by *subordination* (SUB or SUBS). For example, the instantiated concept c_2, which corresponds to the expression *der Prüfling* (*the examinee*), is subordinated to the generic concept *Prüfling* (*examinee*). The relations between the situation (represented by) c_1 and its participants c_2 and c_3 are represented by semantic roles, here AGT (agent) and OBJ (neutral object). Concepts are labeled with an *ontological sort*; the concept c_2, for instance, is of the sort *d* (discrete object). Moreover, concepts are specified with respect to several *layer attributes* whose values indicate, e. g., the facticity of a concept (FACT), its determination of reference (REFER), and its quantificational content (QUANT).

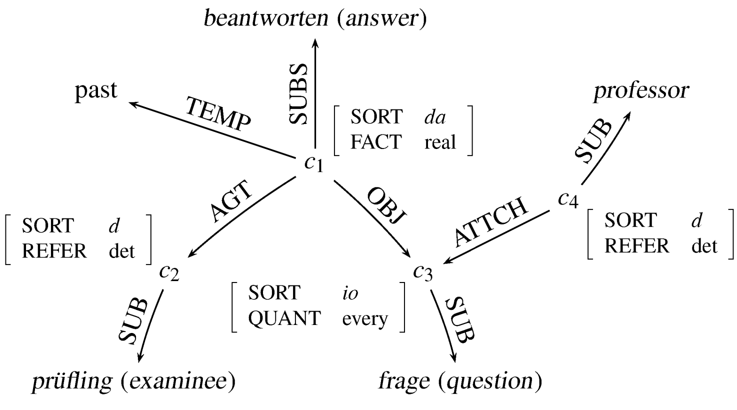

Fig. 1: *Simplified MultiNet representation for the sentence* Der Prüfling beantwortete jede Frage des Professors (The examinee answered every question of the professor)

Question answering as an integrated task for semantic NLP
Probably the most comprehensive NLP task integrating all aspects of NL understanding is question answering based on deep semantic analysis. In the literature one finds different uses of the term *question answering,* which is applied – among other things – to systems relying on text retrieval methods (Hovy et al. 2001), to ontology-based search on the web (Vargas-Vera et al. 2004), and also to systems based on linguistic methods with different depth of understanding (Harabagiu et al. 2001, Hartrumpf & Leveling 2006). Here, we define a *question answering* (*QA*) *system* as a knowledge-based system that can be accessed using NL questions. In contrast to information retrieval systems, the information contained in the knowledge base (KB) of the QA system and used for logical answer-finding is originally given in NL and has to be gathered (assimilated) into the KB by means of NLP techniques that transform NL expressions into a concept-oriented semantic representation. The answers to the questions are typically generated from this semantic representation. The semantic formalism of a QA system should fulfill the criteria formulated above.

To create such a QA system is the long-term goal of all the research using MultiNet. It is also the general setting within which the linguistic work connected with MultiNet is embedded. This is an important decision since the proper interoperation of all components of a large-scale NLP system whose KB contains millions of facts, whose computational lexicon contains tens of thousands of lexemes, and whose inference system has to deal with thousands of axioms, can only be shown in the proper functioning of a large application such as QA which integrates all of these parts. The MultiNet framework is strongly committed to this overall goal and the NLP resources, methods, and applications described in this article can be seen as milestones on this road.

In view of the vast amount of work on QA and semantic information processing reported in literature, the following brief overview of related work is necessarily selective. QA systems using an integrated approach include QUETAL (Frank et al. 2005), JAVELIN (Nyberg et al. 2003), COGEX (Moldovan et al. 2003), PowerAnswer (Bowden et al. 2006), START (Katz 1997), QRISTAL (Laurent et al. 2005), and TACITUS (Hobbs et al. 1993). Each of them has its special merits, but all of them use a shallow semantic representation such as, for instance, Minimal Recursion Semantics (Copestake et al. 2006) with QUETAL. Many of them draw on WordNet (Fellbaum 1998) as a lexical resource; COGEX, for instance, uses an enriched form of WordNet, called XWordNet (Moldovan & Rus 2001). Moreover, there are systems with a strong syntactic component and broad coverage but still shallow semantic representations which are being embedded in a QA system setting. For example, the PARGRAM system (Butt et al. 2002) uses a semantic representation drawn from Lexical-Functional Grammar (LFG) structures (Bobrow et al. 2005) and the QA system of Bos & Nissim (2006) translates NL expressions into Discourse Representation Structures by employing a parser based on Combinatory Categorial Grammar (CCG). However, both approaches seem to have neglected word sense disambiguation up to now, which is problematic if one aims at semantically deep representations.

To the best of our knowledge, none of the existing approaches seems to push both the integral aspect and the depth of semantic analysis so far as the MultiNet paradigm without compromising coverage. The following sections describe the resources and NLP techniques related to the Multi-Net paradigm and applications built on top of them. The exposition has to be reasonably brief and the reader will be referred to more specialized publications for further reading.

2. Meaning Representation with MultiNet

Semantic networks are one of the prominent knowledge representation paradigms for meaning representation in NLP (see Lehmann 1992 for an overview). In MultiNet, as in other semantic network (SN) approaches, concepts are represented by nodes, and relations between concepts are represented as edges between these nodes. Every node of a MultiNet SN is classified according to a predefined conceptual ontology forming a hierarchy of sorts and every edge is labeled by a member of a fixed set of relations and functions. Characteristic features of the MultiNet approach are a distinction between an *intensional* and a *preextensional level* of representation and a *stratification* of different conceptual aspects such as facticity and determination of reference into different representational layers (see below).

Conceptual Ontology of Sorts

The classification of SN nodes, i. e. of the semantic representatives of concepts, by ontological sorts is an important basis for the definition of the domain and value restrictions of relations and functions that establish the interconnections between the nodes. The upper part of the conceptual ontology used in MultiNet is shown in Figure 2.[3] The ontological sorts are also an important source for the semantic interpretation of NL constructs (e. g. prepositional phrases). This especially holds for the semantic disambiguation of relations underlying an NL construct. For example, in the phrase *the holidays in spring,* the preposition *in* must be interpreted by the temporal relation TEMP (and not, for instance, by a local relation), since the semantic representative of the phrase *in spring* bears the sort *t* (a temporal interval).

3. A complete description of the MultiNet sorts can be found in Helbig (2006: 17.1).

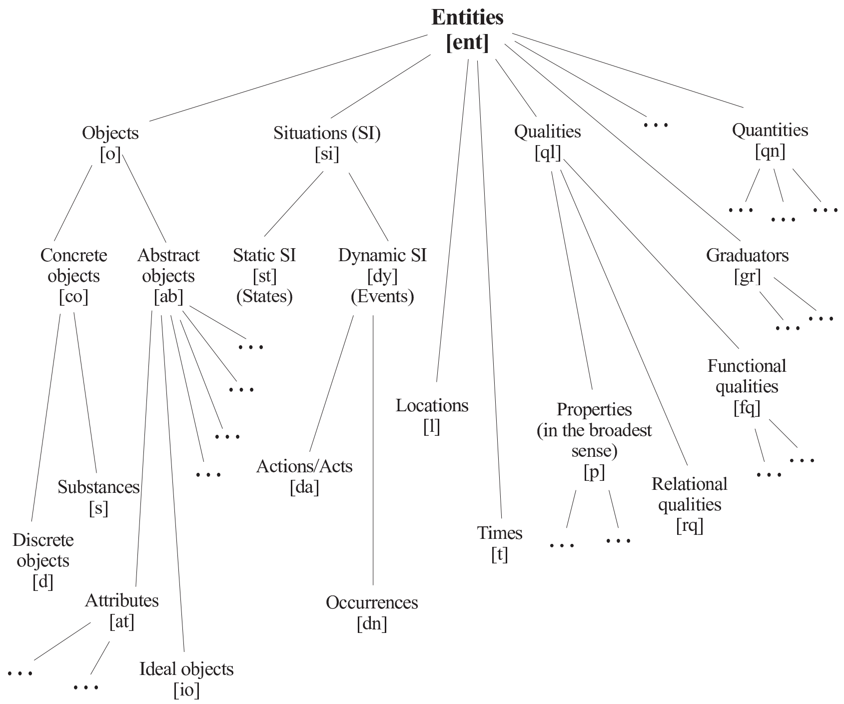

Fig. 2: *Upper part of the MultiNet sort hierarchy*

Relations and Functions

MultiNet provides about 140 basic semantic relations and functions, which can be classified into the following three groups: relations and functions at the intensional level, those at the preextensional level, and lexical relations. On the intensional level, relations and functions are used to describe conceptual objects and situations with their inner structure and their relationships to other entities. Objects are typically characterized by their material structure (by their material origin, ORIGM, or their parts, PARS) or their qualitative characterization (by means of properties or attribute-value specifications). Situations are typically characterized by means of the semantic roles of their participants (like agent, AGT, or experiencer, EXP) and by their spatio-temporal embedding (by means of local and temporal relations). Relations and functions of the preextensional level allow the modeling of sets and extensional representatives, which have to be part of the knowledge representation to deal properly with the meanings of constructs involving sets (such as *three of them*). Finally, lexical relations describe connections between

Table 1: *Strongly simplified description of MultiNet relations used in this article*

Relation	Signature	Short Characteristics
AGT	$si \times co$	Agent
ANTE	$t \times t$	Temporal succession
ATTCH	$o \times o$	Attachment of objects
CIRC	$si \times si$	Relation between situation and circumstance
EXP	$si \times o$	Experiencer
LOC	$[si \cup o] \times l$	Location of an object or situation
MANNR	$si \times [ql \cup st]$	Manner of a situation
OBJ	$si \times [o \cup si]$	Neutral object
ORIGM	$co \times co$	Relation of material origin
ORNT	$si \times o$	Orientation of a situation towards something
PARS	$[o \times o] \cup [l \times l]$	Part-whole relationship
RSLT	$si \times [o \cup si]$	Result of an event
SCAR	$st \times o$	Carrier of a state
SUB	$o \times o$	Conceptual subordination (objects)
SUBS	$si \times si$	Conceptual subordination (situations)
TEMP	$si \times [si \cup t]$	Temporal specification of a situation

generic concepts and play an important role in the specification of lexical entries. This group includes relations specifying synonyms or antonyms, converse concepts and complementary concepts. In addition, there are lexical relations characterizing a change of sorts from one concept to another, such as the relation between an event, e. g. *produce,* and its corresponding abstract situation, e. g. *production.*[4]

All MultiNet relations and functions are systematically described by means of the following characteristic components (see Figure 3 for an example): a short caption with an expressive name; the algebraic signature of the relation or function depending on the MultiNet hierarchy of sorts; a verbal characterization of the relation or function; a mnemonic hint supporting the communicability; patterns of queries aiming at the relation; a detailed description explaining how to use the relation or function and what logical axioms define its inferential properties.

The relations and functions (i. e. the labels of the edges at the concept level) can be regarded as nodes at a meta level that are connected by axiomatic rules. These axioms, which have the form of logical implications, are of central importance to inference processes working over a MultiNet knowledge base. There are two general types of axioms: *B-axioms* and *R-axioms.*[5] B-axioms connect relations and functions with

4. See Helbig (2006: Part II) for the full specification of all MultiNet relations and functions. The MultiNet relations used in this article are listed in Table 1.
5. See Helbig & Glöckner (2007) for a more fine-grained distinction between different types of axioms.

PARS: Relation between Part and Whole

PARS: $[o \times o] \cup [l \times l]$

Definition The relation (o_1 PARS o_2) specifies that o_1 is a part of o_2, in other words, o_2 is the whole which contains o_1 as component. The relation PARS is restrictedly transitive, asymmetric and not reflexive.

Mnemonics pars (Lat.) - (En: part; Ge: Teil)
(x PARS y) – [x is a part of y]

Question patterns What does $\langle o_2 \rangle$ consist of? Which parts does $\langle o_2 \rangle$ {[consist of]/[have]/…}?
Which part(s) does $\langle o_2 \rangle$ {[possess]/[have]/…} ?
To $\langle$WM$\rangle$ does $\langle o_1 \rangle$ belong?
From which object is $\langle o_1 \rangle$ a {part/component}?

Commentary In the case of concrete objects, PARS is used to specifiy the physical components of an object. The relation may be also used to characterize that an abstract object contains another object with sort *ab* (this holds for instance for the parts of a plan), or that locations are contained in each other. Because of the formal properties of PARS, this relation gives rise to hierarchies of conceptual representatives within a semantic network. The inheritance mechanisms connected with them can be described by axioms of the following kind:

$$(o_1 \text{ PARS } o_2) \wedge (o_2 \text{ ORIGM } s) \rightarrow (o_1 \text{ ORIGM } s)$$

The postulation of an unrestricted transitivity for the relation PARS would be problematic. Even if it were formally correct to progress the transitivity of the PARS relation without restriction, this would not conform to the natural language use where there are clearly limits with regard to that.

The inheritance of the part-whole relation in a SUB hierarchy is described by the following axiom:

$$(o_1 \text{ SUB } o_2) \wedge (o_3 \text{ PARS } o_2) \rightarrow \exists\, o_4\, [(o_4 \text{ SUB } o_3) \wedge (o_4 \text{ PARS } o_1)]$$

Fig. 3: *Abbreviated description of the part-whole relationship*

representatives of NL concepts. The following example axiom establishes a connection between the relation PARS and the concept of (*possessing* a *component*):

$$\begin{aligned}\exists p[(p \text{ SUBS } \mathit{possess}) \wedge (p \text{ SCAR } k_1) \wedge (p \text{ OBJ } k_2)\\ \wedge (k_2 \text{ SUB } \mathit{component})\\ \wedge (k_1 \text{ ATTCH } k_2)] \rightarrow (k_2 \text{ PARS } k_1)\end{aligned}$$

R-axioms connect relations and functions with each other and do not contain NL concepts. An example is the last axiom in Figure 3 describing the inheritance of the part-whole relationship in a hierarchy of subordinated concepts.

Stratification

One of the aims of the MultiNet design is to overcome the limitations of network paradigms that force qualitatively different aspects of meaning into one flat structure. For this purpose, the nodes and edges of MultiNet are embedded in a multidimensional space of so-called *layer attributes.*

The layer specifications for edges are expressed by the attribute K-TYPE, which signifies the type of knowledge represented by the edge. An edge connected to a concept c which represents categorical knowledge about c is labeled [K-TYPE = *categ*]. Edges which express prototypical knowledge (default knowledge) about the concept are marked [K-TYPE = *proto*]. Finally, edges connected to c which do not affect its basic meaning but rather indicate participation in certain situations are labeled [K-TYPE = *situa*] to mark them as situational. Categorical and prototypical knowledge together form the *immanent knowledge* which − in contrast to the situational knowledge − characterizes a concept inherently.[6]

The nodes are specified with respect to the following layer attributes:

GENER: The *degree of generality* indicates whether a conceptual entity is generic (*ge*) or specific (*sp*). Examples: a) *A car* [GENER = *ge*] *is a useful means of transport.* b) (*This car*) [GENER = *sp*] *is a useful means of transport.*

FACT: This attribute describes the *facticity* of an entity, i. e. whether it really exists (*real*), whether it does not exist (*nonreal*), or whether it is only hypothetically assumed (*hypo*). Examples: a) *John* [FACT = *real*] *thought that* (*he was ill*) [FACT = *hypo*]. b) *John* [FACT = *real*] *remembered that* (*he was ill*) [FACT = *real*].

REFER: This attribute specifies the *determination of reference,* i. e. whether there is a determined object of reference (*det*) or not (*indet*). This type of characteristic plays an important part in NLP in the phase of knowledge assimilation and especially in the resolution of references. Example: (*The passenger*) [REFER = *det*] *observed* (*an accident*) [REFER = *indet*].

QUANT: The intensional *quantification* represents the quantitative aspect of a conceptual entity; whether it is a singleton (*one*) or a multitude (*two, three, … several, many, … most, all*). Example: *Many villages* [QUANT = *many*] *had been flooded.*

6. The distinction between immanent and situational knowledge in MultiNet roughly corresponds to the distinction between definitional and assertional knowledge drawn, e. g., in Allgayer & Reddig-Siekmann (1990).

ETYPE: This attribute characterizes the *type of extensionality* of an entity with values: 0 − individual that is not a set (e. g. ⟨*Elizabeth I*⟩), 1 − entity with a set of elements from type [ETYPE = 0] as extension (e. g. ⟨*many houses*⟩, ⟨*the family*⟩), 2 − entity with a set of elements from type [ETYPE = 1] as extension (e. g. ⟨*many families*⟩), etc.

CARD: The *cardinality* as characterization of a multitude at the pre-extensional level is the counterpart of the attribute QUANT at the intensional level specifying the cardinality of sets.

VARIA: The *variability* describes whether an object is conceptually varying (*var*) − i. e. it is a so-called parameterized object − or constant (*con*). Example: (*This policeman*) [VARIA = *con*] *checked* (*every passport*) [VARIA = *var*].

3. Automatic Semantic Analysis

3.1. Lexicon

The central lexical resource for the NLP applications described in this article is the computational lexicon HaGenLex (Hagen German Lexicon). HaGenLex is a general domain lexicon for German, which, at the time of the writing, contains about 26,000 lexical units (13,500 nouns, 7500 verbs, 3300 adjectives, and 600 adverbs) with detailed morphological, syntactic, and semantic specifications.[7] The semantic specification of HaGenLex entries rests on the MultiNet formalism: entries are assigned an ontological sort of the MultiNet sort hierarchy and valency frames are equipped with MultiNet semantic roles. In addition, a set of *binary semantic features* such as HUMAN, ARTIF(ICAL), GEOGR(RAPHICAL), and INFO(RMATION) is employed in the lexicon for subclassifying the ontological sorts. The main reason for introducing these features is to allow the efficient checking of semantic compatibility by feature unification. This method is used, for instance, to verify that selectional restrictions in valency frames are satisfied by the respective complements. The combination of sorts and features also helps to disambiguate references during assimilation (see Section 3.3.).

Figure 4 sketches the case frame description in HaGenLex for the two verbs *erfinden* (*invent*) and *beantworten* (*answer*). All valency slots in the case frame are characterized by their semantic roles, their selectional restrictions, and their morpho-syntactic realizations. Figure 5 shows the

7. See Hartrumpf et al. (2003) for a more thorough description of HaGenLex.

erfinden (*to invent*)				*beantworten* (*to answer*)			
action, MENTAL −				*action*, MENTAL −			
AGT	HUMAN	+	NP [NOM]	AGT	HUMAN	+	NP [nom]
RSLT	ARTIF	+	NP [acc]	ORNT	HUMAN	+	NP [dat] optional
				OBJ	INFO	+	NP [acc]

Fig. 4: *Sketch of semantic class and case frame for two example verbs*

actual feature structure representation of the entry for *erfinden* in abbreviated form. In addition to simple role frames, HaGenLex allows more expressive semantic specifications in terms of general MultiNet expressions. This is especially useful for verbs of higher arity, often verbs with prepositional complements, which typically describe complex relationships between participants not expressible by semantic roles alone (see Osswald et al. 2006). The regular adjunct semantics of prepositions is treated within HaGenLex by context-dependent interpretation rules (Hartrumpf et al. 2006b). Since many adjuncts (such as directional and durational adverbials) are restricted to certain types of verbs, the semantic relations compatible with the semantics of a verb are explicitly listed in the entry. For instance, directional and durational adjuncts are excluded for state and punctual verbs, respectively.

$$
\begin{bmatrix}
\text{MORPH} & \begin{bmatrix}\text{BASE} & \textit{“erfinden”}\end{bmatrix}\\
\text{SYN} & \begin{bmatrix}\textit{v-syn} & \\ \text{V-TYPE} & \textit{main}\\ \text{PERF-AUX} & \textit{haben}\end{bmatrix}\\
\text{SEMSEL} & \begin{bmatrix}
\text{SEM} & \begin{bmatrix}\textit{sem} & \\ \text{ENTITY} & \textit{nonment-action}\end{bmatrix}\\
\text{SELECT} & \left\langle \begin{matrix} \begin{bmatrix}\text{REL} & \{\textit{agt}\}\\ \text{SEL} & \begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{CAT} & \textit{np}\\ \text{AGR} & \begin{bmatrix}\text{CASE} & \textit{nom}\end{bmatrix}\end{bmatrix}\\ \text{SEMSEL} & \begin{bmatrix}\text{SEM} & \begin{bmatrix}\text{ENTITY} & \begin{bmatrix}\text{HUMAN} & +\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\\ \begin{bmatrix}\text{REL} & \{\textit{rslt}\}\\ \text{SEL} & \begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{CAT} & \textit{np}\\ \text{AGR} & \begin{bmatrix}\text{CASE} & \textit{acc}\end{bmatrix}\end{bmatrix}\\ \text{SEMSEL} & \begin{bmatrix}\text{SEM} & \begin{bmatrix}\text{ENTITY} & \begin{bmatrix}\text{ARTIF} & +\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{matrix}\right\rangle\\
\text{EXAMPLE} & \left\langle \textit{“(Der Student) (erfand) (ein Berechnungsverfahren).”}\right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Fig. 5: *HaGenLex example entry as typed feature structure (abbreviated)*

3.2. Syntactic-Semantic Parser

To automatically derive SNs for German sentences (or phrases), the parser WOCADI (WOrd ClAss based DIsambiguating parser; Hartrumpf 2003) has been developed which follows the word-class function analysis (WCFA) paradigm of Helbig (1986). The grammatical knowledge of WOCADI is encoded in *word-class functions* (WCFs). For each word-class (syntactico-semantic categories comparable to a specific set of parts of speech), a WCF is implemented. The WCFs together with a central control mechanism model two phases of activity during analysis (or understanding) of the current word in a sentence: opening expectations about what might or should follow and saturating these expectations by connecting complements and adjuncts to their governors etc. The parser employs a morpho-lexical analysis that uses HaGenLex (see Section 3.1.), an additional flat lexicon without semantics and valency frames, several name lexicons, a morphological analyzer, and a compound analyzer. The WCFs construct an SN during parsing mainly by SN unification of the semantic components from lexical entries. WOCADI is oriented towards constructing SNs but also builds labeled dependency trees. The emphasis is on understanding acceptable sentences, not on modeling grammaticality distinctions.

To disambiguate syntactic and semantic alternatives, a hybrid *rule-statistical* approach is employed: symbolic, hand-written rules license possible alternatives, one of which is selected based on statistics derived from annotated corpora. To deal with *sparse data problems* a backoff model is applied, which distinguishes different dimensions of description such as detail of alternative description and number of alternatives. Examples of disambiguation problems tackled by WOCADI are interpretation and attachment of prepositional phrases (with a combined accuracy of 82%; Hartrumpf 2003) as well as coreference resolution (see Section 3.3.).

The parser analyzes large corpora including newspaper corpora and encyclopedias like Wiki-pedia; in total, it has already parsed around 30 million different sentences. Its coverage depends on the corpus: on average, complete SNs are delivered (at a speed of 2000–4000 sentences per hour) for 60% of all sentences, and an additional 25% receive an SN containing only nominal chunks. WOCADI is a central building block for the applications described in Section 4.

3.3. Assimilation

Assimilation (or *semantic integration*) is the process of integrating the meaning representations of individual sentences into a coherent representation of the whole text (*intratextual* assimilation) − and also of integrating the meaning of a text into the overall knowledge base (KB) of an NLP system (*intertextual* or *global assimilation*). One of the main tasks of this process is detecting explicitly or implicitly introduced identical discourse entities occurring in different text parts, which must then be merged into a single semantic representative. The difficulties that arise in this context will be illustrated by the following example. The semantic representation of (S1) is assumed to be the basic information, represented already in the KB, while (S2a) and (S2b) are assumed as possible text continuations of (S1).

(S1) *Familie Beier hat im vergangenen Jahr ein Haus gebaut.* (*Last year, the Beier family built a house.*)

(S2a) *Bald danach waren sie über die Qualität des Gebäudes zerstritten.* (*Soon afterwards, they had a quarrel about the quality of the building.*)

(S2b) *Der Keller wurde beim diesjährigen Hochwasser vollständig überflutet.* (*The basement was completely inundated by this year's floods.*)

The meaning representations of (S1) and (S2b) viewed as isolated sentences are shown in Figure 6; these analyses were generated by the parser described in Section 3.2. Figure 7 represents the assimilation result after joining the SNs of sentences (S1) and (S2b) into one KB.

Resolution of direct references

The most important types of reference are anaphoric (backward pointing), cataphoric (forward pointing), and deictic (pointing to the situational context). We will first consider references to an entity mentioned in the text. Such direct references are often expressed by proforms (pronouns and proadverbs). Consider *sie* (*they*) in (S2a), which refers to *Familie Beier* (*Beier family*) in (S1). To resolve this reference, one must know that *family* represents a collection (expressed by the layer attribute constraint [ETYPE = 1]; see the right part of Figure 7). This information is provided by HaGenLex.

Non-pronominal direct references in a text are often characterized by the use of hypernyms or synonyms to remention entities already introduced in the discourse. We call such references *ontological references* since they depend on relationships which are typical of ontologies (e. g. conceptual subordination as expressed by SUB). An example occurs in (S2a): the phrase *des Gebaudes* (*of the building*) points to the house introduced in (S1). Such a reference (also called *inclusion*) is often mediated over several steps in the subordination hierarchy.

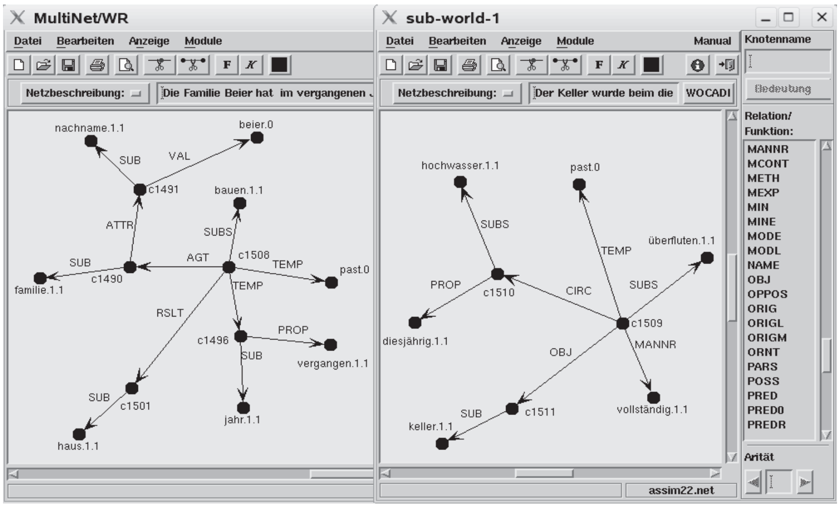

Fig. 6: *The relational layer of semantic representations of sentences (S1) and (S2b) before assimilation − screenshot from the knowledge engineering toolkit MWR (Gnörlich 2002)*

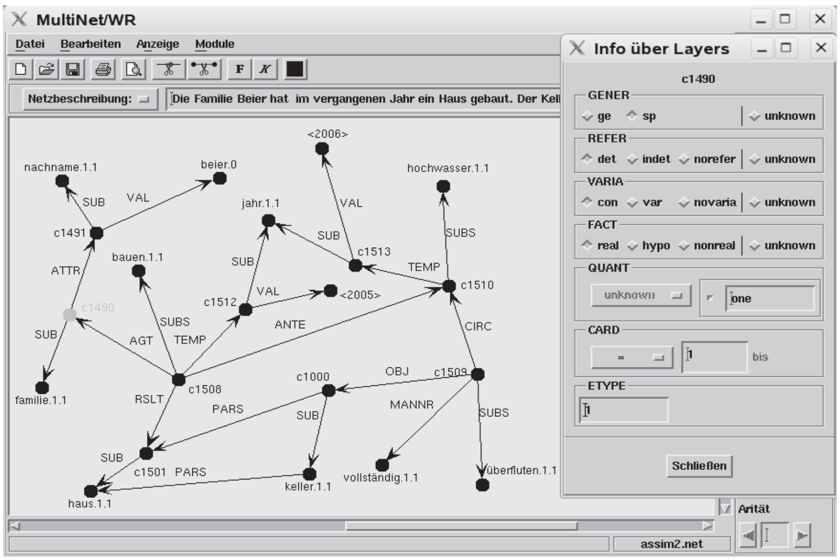

Fig. 7: *The relational layer of the semantic representation of (S1) and (S2b) after assimilation*

The interpretation of proadverbs, like *here* (local deixis), and semantically related expressions, like *last year* (temporal deixis) as in (S1), often depends on the spatio-temporal anchoring of the text. By decrementing the current year *Y*, which can be determined from a dialog model or meta-knowledge about the publication time of the text, *last year* is correctly interpreted as the year $Y - 1$ (see node $c1512$ in Figure 7). More details about the resolution of temporal deixis in the MultiNet framework and an extrinsic evaluation in a QA system can be found in Hartrumpf & Leveling (2006). If the situational context is described by a larger MultiNet network, the agreement of sorts (sort t – *temporal entity* – for temporal deixis, sort l – *local entity* – for local deixis) or even semantic features, like [GEOGR +] or [HUMAN +], attached to concepts in the KB can be of great help to disambiguate between multiple reference candidates (e. g. for finding referents of pronouns like *he* or *she* which impose a selectional restriction on their antecedents).

Automatic coreference resolution

The coreference phenomena mentioned so far are handled by CORUDIS (Hartrumpf 2001, 2003), a hybrid rule-based/statistical coreference resolver which works on the SNs and dependency trees of the text sentences. CORUDIS (COreference RUles with DIsambiguation Statistics) covers all *mentions* (*markables*) that are NPs.

Two kinds of data are required by CORUDIS: rules defining whether two mentions *can* corefer and a corpus annotated with coreference information. The rules license possible coreferences, while the corpus is used for scoring alternatives with estimated probabilities. Each rule consists of a premise specifying constraints for the anaphor and an antecedent candidate and a conclusion with a small SN expressing the coreference (e. g. an EQU relation). On the syntactic side, coreference rules contain predicates that test syntactic configurations like the presence or absence of dependencies and the unifiability of syntactic agreement features. On the semantic side, rules can refer to knowledge encoded in MultiNet, such as hyponymy and meronymy, and can test for unifiability of semantic representations of the mentions involved.

The other kind of data in CORUDIS are corpus annotations which give rise to scores for candidate antecedents. These local scores are optimized in a heuristic search for a global optimum for coreferences in a given text. On a German newspaper corpus with 452 coreferences, CORUDIS achieves 87.8 % precision and 66.7 % recall (75.8 % F-score). The possible coreference alternatives determined by CORUDIS can serve as a starting point for the deductive techniques described below.

Logical recurrence and bridging references
Non-pronoun references including indirect references to entities not explicitly mentioned in the text can be handled by *reference resolution by deduction.* This method, which is part of our knowledge engineering workbench MWR (MultiNet-Wissensrepräsentation; Gnörlich 2002), rests on the following idea.[8] Consider a sentence S, whose semantic interpretation sem(S) has to be assimilated into an already existing KB K. MultiNet employs the layer attribute constraint [REFER = *det*] for an SN node $c_r \in$ sem(S) which signals that c_r stems from a definite description in S and must hence be resolved from K (or from the dialog model). The node c_r is characterized by its definitional knowledge $D(c_r)$ which comprises all edges needed to express what the node stands for.[9] The logical expression describing $D(c_r)$ is interpreted as a question to be answered or, in technical terms, a theorem to be proved over K. In this inference process, c_r is a variable to be substituted by a known node c_a (the antecedent) contained in K. The central step in assimilating sem(S) into K is the identification of the nodes c_r and c_a and merging them into one node c of the extended KB arising from K and sem(S). In this sense the assimilation is a function $A : \mathcal{K} \times \mathcal{T} \rightarrow \mathcal{K}$ mapping an old $K \in \mathcal{K}$ and a meaning representation sem(S) $\in \mathcal{T}$ to another (extended) KB $K' \in \mathcal{K}$

For *bridging* references (Asher & Lascarides 1999, Clark 1977), the antecedent node c_a is not explicitly contained in K, but only implicit from the knowledge provided by axioms. In order to cope with these references, coreference resolution must therefore incorporate background knowledge and logical inference. A typical example is given by sentence (S2b), where meronymic knowledge is needed to find the antecedent c_a for $c_r = c1511$ described by *der Keller* (*the basement*). The semantic description $D(c_r)$ of this phrase involving the variable c_r is given by (c_r SUB *basement*). This is also the theorem to be proved from the SN sem($S1$) shown in Figure 6 (left side). The background knowledge (basementPARS *building*) and the last axiom of Figure 3 are needed to identify c_r with the basement of building $c1501$ (represented by a fresh constant $c1000$), and also infers the part-whole relationship between basement and building.

The coreference resolver CORUDIS, methods for resolving temporal deixis, and the logical assimilation stage built on top of these methods

8. In the following, the assimilation process is denoted by A; $\mathcal{K}$ is the set of KBs successively generated by A; $\mathcal{T}$ is the set of all meaning representations of isolated sentences to be assimilated into that $K \in \mathcal{K}$ which was last generated by A; $K \in \mathcal{K}$ consists of the knowledge assimilated so far, the general background knowledge, and the axioms.
9. For quantified nodes, the defining edges correspond to the restriction of the quantifier, while the non-defining or *assertional* edges form its nuclear scope.

make it possible to integrate the MultiNet representations of individual sentences into a richer representation of the whole text which reveals the connections between sentences sharing discourse entities. The resulting semantic representations on the text level form the basis for knowledge processing in some of the NLP applications described in the next section.

4. Applications

4.1. Retrieval of Bibliographic Information and Text Retrieval

Finding information on the World Wide Web is getting more and more difficult, since the amount of information is steadily increasing. Therefore, easy-to-use interfaces will become even more important in the future. NLI-Z39.50 (Leveling 2006a) is a natural language interface (NLI) for information providers on the web. Its target systems include all large German library union catalogues such as the HBZ (*Hochschulbibliothekszentrum Nordrhein-Westfalen*) and the GBV (*Gemeinsamer Bibliotheksverbund der sieben norddeutschen Bundesländer*), the German national library DDB (*Die Deutsche Bibliothek*), foreign libraries such as the Library of Congress, and commercial information providers such as *Amazon.com.* User interfaces for traditional library systems such as OPACs (Online Public Access Catalog) differ among providers, so that using an OPAC has to be learned for each new system. In contrast, the NLI offers a uniform, easy-to-use interface for parallel access to information providers. Furthermore, it solves known problems with OPACs such as correct usage of the Boolean operators `AND` and `OR` in queries (see Borgman 1996 for a general overview of user's problems with OPACs).

The NLI-Z39.50 system aims at *a full understanding* of user queries. Therefore, NL queries are analyzed and represented in form of SNs corresponding to the MultiNet paradigm. A typical user query for the NLI is: *Wo finde ich Bücher von Peter Jackson, die in den letzten zehn Jahren bei Springer und Addison-Wesley veröffentlicht wurden?* (*Where do I find books by Peter Jackson which were published by Springer and Addison-Wesley in the last ten years?*)

NL queries are processed as follows (numbers correspond to the modules in Figure 8): (1) The User client/Controller controls the communication between the other modules. It is accessible via a web browser, accepts NL queries, and displays results in a readable form. (2) The WOCADI parser (see Section 3.2.) performs a syntactico-semantic analysis of the input text and returns its representation as an SN. An advantage of deep NLP is that decomposition of nominal compounds for query expansion is produced quasi as a side-effect. In statistical approaches for retrieving information (such as *n*-gram based approaches), these methods have to be developed separately or are not applicable at all. Further-

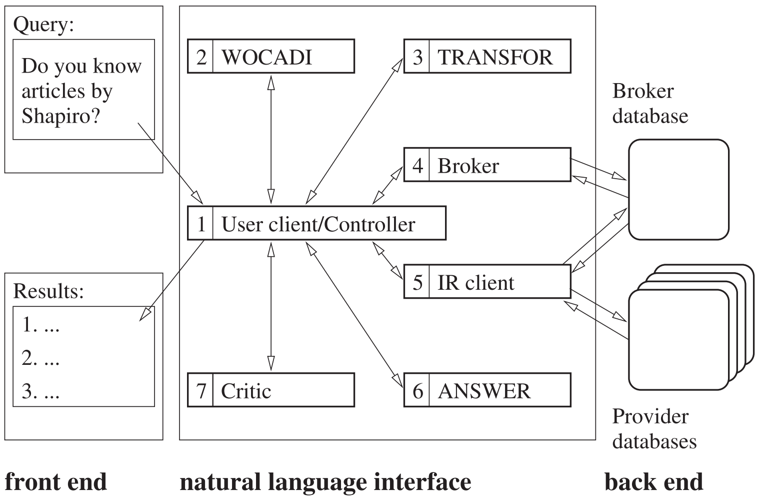

Fig. 8: *Architecture of the NLI-Z39.50* (*arcs indicate data flow*)

more, information in the lexicon can be exploited by including synonyms and hypernyms of search terms in a query. In comparison to traditional OPACs or to other NLIs for bibliographic retrieval, NLI-Z39.50 provides a better interface for bibliographic data (Stenglein 2006). (3) TRANSFOR transforms the meaning structures into an intermediate query representation (Leveling 2006a), which captures the core meaning of the user's information need. Concepts, semantic relations, and functions of the query SN are thus mapped to attributes, types, and search terms of the intermediate representation as indicated by the following example:[10]

```
(bibcode              = (char "b"))
((OR author editor)   = (name "jackson" "peter"))
(publisher            = (OR (name "springer")
                            (name "addison-wesley")))
(publication_time    >= (date 1996))
```

(4) The Broker recommends target systems for a given query. It considers user preferences, properties and supported search features of a target system as well as more pragmatic aspects (such as proximity to a library) as criteria for selecting a target database. For the query *Gibt es hier in der Nähe Bücher über KI?* (*Are there any books about AI nearby?*), librar-

10. Note that the word *and* in the query is interpreted correctly as a logical OR.

ies within a distance of about 30 km around the user's location will be preferred. (5) An Information Retrieval client controls access to the databases and adapts a database query to the target profile of a database before a search is initiated. The target profile includes information about search features that are supported by a database. Thus, a query for an `editor` will be transformed into a query for a `person` if the target does not support a search for editors. Results (in this case, bibliographic records) are then transmitted from the databases to the NLI. (6) The answer module ANSWER processes results from various bibliographic formats (USMarc, MAB, XML, etc.) for presentation. It converts them into a generic intermediate format and creates HTML fragments for their presentation. (7) During query processing, intermediate results are sent to a specialized analytic module, the so-called Critic. The Critic identifies problematic situations and generates corresponding feedback for the user, which will help to solve typical problems encountered (Leveling 2004). For instance, the Critic answers with suggestions for spelling correction if the query contains errors (*Did you mean* …?) or provides semantically related search terms for expanding the query if the search was not successful.

NLI-Z39.50 has been evaluated several times since 2001, using criteria such as precision, recall, form of result presentation, and user effort required for retrieving results. The system performance was investigated in an external evaluation performed by the University of Leipzig in 2003, and a user study with about 300 participants (Stenglein 2006). With regard to precision and recall, the NLI was evaluated once per year between 2003 and 2006 in the domain-specific track of the CLEF (Cross-Language Evaluation Forum) campaign. These evaluations show that the methods employed in the NLI provide results superior to that achievable with traditional OPAC systems. In the domain-specific retrieval experiments at CLEF 2003, the NLI achieved a mean average precision (MAP) of 0.2064. In comparison, an experiment using Boolean retrieval yielded 0.0809 MAP (Leveling 2006a). By 2006, performance improved to 0.3539 MAP (Leveling 2006b).

4.2. Question Answering over Semantic Networks

As emphasized in the introduction, QA is an ideal application to test the interoperation of deep semantic analysis, inferences, and NL generation because all three components are central to a successful open-domain QA system. A QA system delivers in response to a user's question not just a ranked list of relevant texts as in conventional information retrieval, but a concise and precise answer. InSicht is a QA system for German which works on SNs. It implements a deep approach in that it delivers answers by matching SNs from documents with the SN of the question and its deductive expansions.

InSicht has six main processing phases: (1) *Document processing*: All documents from a given collection are transformed into a standard XML format (XCES) with word, sentence, and paragraph boundaries marked up. All preprocessed documents are parsed by WOCADI yielding a syntactic dependency structure and an SN representation of the MultiNet formalism for each sentence. (2) *Query processing*: WOCADI parses the question; determining the sentence type (here, often a subtype of *question*) is especially important because it controls two later steps: query expansion and answer generation. (3) *Query expansion*: SNs equivalent or similar to the original SN of the question are derived by employing lexico-semantic relations from HaGenLex, equivalence rules, and inference rules for MultiNet. The inference rules are axioms for Multi-Net relations and meaning postulates for lexical concepts from HaGenLex like entailments for situations (e. g. linking *töten* (*kill*) and *sterben* (*die*)).[11] The results are disjunctively connected SNs that try to cover the representations of many different sentences that (explicitly or implicitly) contain an answer to the question. (4) *Semantic network matching*: All document sentences matching an SN from query expansion are collected. A two-level approach is chosen for efficiency. First, an index of HaGenLex concepts is consulted with the concepts from the question networks. Second, the SNs of the retrieved documents are compared sentence network by sentence network to find a match with a question network. (5) *Answer generation*: NL generation rules are applied to SNs matching the query and an NL answer is generated. Therefore answers from InSicht can differ from the exact form found in the documents. The sentence type and the SN control the selection of answer rules. The rules also act as a filter for uninformative or bad answers. The results are tuples containing a generated answer string, one or more supporting snippets, an answer score (a number between 0 and 1), the supporting document ID, and the supporting sentence ID. InSicht currently employs answer-oriented rules for generating NL answers from found answers in SN form. It can later be replaced by a more sophisticated generation system if needed. (6) *Answer selection*: To deal with multiple candidate answers from answer generation, a selection step is required. It employs answer clustering and is driven by a preference for more frequently occurring answers and a preference for more elaborate answers. The best answer(s) and the supporting sentences (and the IDs of supporting sentences or documents) are presented.

11. Some entailments come from an automatic formalization of GermaNet glosses (Glöckner et al. 2005).

In the QA tracks at CLEF 2004, 2005, and 2006, InSicht was one of the best German QA systems. Its orientation towards precision allows promising combinations with systems following other QA approaches (Glöckner et al. 2007).

4.3. MAVE: A Tool for Enhancing Answer Quality of QA Systems

Unlike InSicht, most existing QA systems rely on shallow NLP methods, placing emphasis on recall. From the perspective of a user seeking information, this means that a large number of wrong results must be inspected in order to eventually find a correct answer. The MAVE system (Multinet-based Answer VErification) was built to support a user seeking information by automatizing this process of answer validation (Glockner 2006). MAVE employs deep linguistic analysis and robust techniques for knowledge processing to achieve this.

(1) The inputs to MAVE, i. e. the question, the answer to be inspected, and a supporting text snippet provided by the QA system, are first subjected to a *deep linguistic analysis* (Section 3.2.), which results in corresponding MultiNet representations. This process includes disambiguation of word meanings, semantic role labeling, a facticity labeling, and assimilation as described in Section 3.3.(2) Post-processing involves a *normalization* of the generated SNs. Synonymy relationships from HaGenLex and other sources are used for replacing lexical concepts with a canonical synset representative (e. g. *ereignen.1.1* ↦ *eintreffen.1.2* for *occur*). Not only the question and snippet nets, but also the background knowledge is automatically normalized in this way. (3) Next comes *hypothesis construction,* i. e. the MultiNet representation of question and answer is translated into a conjunction of literals which represents the actual hypothesis expressed by the answer. Note that different queries will be constructed depending on the parse quality of the NL expressions involved; for details see Glöckner (2006). (4) In order to gain robustness of knowledge processing against knowledge gaps and errors of semantic analysis, the logical prover used by MAVE is embedded into a *constraint relaxation loop.* If a proof of the hypothesis fails, the system repeatedly eliminates critical literals from the hypothesis until a proof of the simplified hypothesis succeeds. The number of skipped literals is used as a robust indicator of entailment strength. (5) MAVE also checks further logical criteria, like the triviality of an answer (provability of an answer over the original question) and circular answers (where the definiendum is part of the definiens). These logical criteria are complemented by several shallow indicators of answer quality which serve to detect false positives. (6) The results of the robust entailment check and the heuristic

error indicators are aggregated and the considered answer is accepted if the combined score is better than a given quality threshold.

With this basic approach, MAVE achieved the best answer filtering results for German in the CLEF 2006 answer validation exercise (58.4% precision, 50.6% recall, 54.2% F-score; see AVE06 overview in Penas et al. 2006). Recently MAVE was modified such that it no longer treats each answer candidate and supporting text snippet in isolation, but rather determines the joint evidence of all snippets supporting a given answer (Glockner et al. 2007). This allows for an even more purposive filtering: when selecting only a single best answer for each question based on the aggregated evidence, MAVE still covers 90 of the 122 AVE06 test questions with a positive answer (in particular, MAVE outperforms any individual QA system represented in AVE06). The user no longer sees all 1,443 answer candidates in this case, but only one answer for each question. While maintaining high relative recall, MAVE then eliminates 953 out of 1,064 wrong answers, which means a 91.5% efficiency in eliminating incorrect candidates.

4.4. Further Applications

Question Answering on the Web (IRSAW)

IRS AW (Intelligent Information Retrieval on the Basis of a Semantically Annotated Web)[12] is a QA framework that integrates syntactico-semantic analysis, the combination of different *answer streams* (i. e. methods for producing answer candidates), logical answer validation (see Section 4.3.), and NL generation. Currently, IRSAW employs answer streams resulting from three different methods (Glöckner et al. 2007): a pattern matching approach, an information retrieval approach, and the InSicht system (Section 4.2.). The answer candidates produced by the different streams are automatically merged and validated and the highest ranked candidate is selected as the answer.

In contrast to other QA systems, IRSAW aims at: a) using a full semantic interpretation of questions and web documents on which logical inferences are based, b) investigating linguistic phenomena such as idioms, metonymy, and temporal and spatial aspects in questions and documents (e. g. deictic expressions), and c) generating NL answers (instead of text extraction).

---

12. The IRSAW project is funded by the DFG (Deutsche Forschungsgemeinschaft), contract number (LIS 4 − 554975(2) Hagen, BIB 48 HGfu 02-01).

Rating and controlling text readability (DeLite)
The DeLite system is a tool for automatically rating the readability of German texts.[13] Since DeLite is able to spot readability problems in texts, the system can also be employed as an authoring tool for writing readable texts. To achieve this, DeLite computes more than forty readability indicators at all levels of linguistic analysis, primarily using the results of the WOC ADI parser (see Section 3.2.). Examples for such indicators are the number of primitives within a compound noun (morphological level), the number of syntactic dependents per noun phrase (syntactic level), the number of propositions per sentence (semantic level), and the number of possible antecedent candidates (discourse level). DeLite determines a global readability score by a weighted aggregation of the indicators. The weights are learned from a training set of texts, whose readability have already been rated in a controlled experimental setup. DeLite overcomes the well-known limitations of traditional readability formulas since its powerful NLP back-end is able to provide diagnostic information of syntactic and semantic complexity. For a detailed description of the DeLite architecture see Hartrumpf et al. (2006a).

Automated knowledge acquisition from texts (MACQUIK)
MACQUIK (Multinet ACQUIres Knowledge) aims at the automated extraction of knowledge bases from documents. This process involves all methods explained in this article, in particular deep semantic analysis and subsequent assimilation of semantic representations into a coherent knowledge base. Our current focus in this long-term research effort is on improving the methods for intratextual assimilation (explained in Section 3.3.) and adding functionality for identifying objects and events across documents as needed for intertextual assimilation. Moreover, techniques for automatic relation extraction are being adapted to the MultiNet framework in order to abstract background knowledge from episodic content.

5. Conclusion

This article makes a strong point for an integrated approach to NLP in general and to question answering in particular. It is argued that the semantic representation formalism used in such systems should fulfill certain criteria such as universality and homogeneity. Question answer-

13. Part of the work on DeLite is financially supported by the EU project *Benchmarking Tools and Methods for the Web* (BenToWeb, FP6-004275).

ing systems based on a deep semantic representation are considered an appropriate test bed for the correct functioning and proper interplay of all parts of an integrated NLP system.

The automatic creation of large knowledge bases from NL texts (be it lexical knowledge or world knowledge) plays a crucial role in building a question answering system. Such knowl edge bases are also important for cross-validation, consistency checking between all parts of knowledge, and further automatic knowledge acquisition by bootstrapping. Without a sufficient amount of basic knowledge it is impossible to automatically gather further knowledge from texts and embed it properly into existing knowledge bases. To this end, deep semantic analysis is needed. For even if background knowledge is gathered from texts by using statistical methods, it has to be connected to a conceptual world which already exists. If we succeed in implementing a deep semantic analysis within question answering over the Word Wide Web, we will be able to take the next step of a pure ontology-based search giving precise answers instead of whole web pages or text snippets. This could be the right way to build a real Semantic Web.

Eingereicht: 2. Februar 2007 *Hagen*
Überarbeitete Fassung eingereicht: 11. Juni 2007

References

Allgayer, Jürgen & Carola Reddig-Siekmann (1990). What KL-ONE Lookalikes Need to Cope with Natural Language − Scope and Aspect of Plural Noun Phrases. In *Sorts and Types in Artificial Intelligence*, Karl Hans Bläsius, Ulrich Hedtstück & Claus-Rainer Rollinger (eds.), 240−285. Berlin: Springer.

Asher, Nicholas & Alex Lascarides (1999). Bridging. *Journal of Semantics* 15: 83−113.

Baader, Franz, Ian Horrocks & Ulrike Sattler (2004). Description Logics. In *Handbook on Ontologies*, Steffen Staab & Rudi Studer (eds.), 3−28. Berlin: Springer.

Bernardi, Raffaella, Diego Calvanese & Camilo Thorne (2007). Lite Natural Language. In *Proc. of the 7th International Workshop on Computational Semantics*, Jeroen Geertzen, Elias Thijsse, Harry Bunt & Amanda Schiffrin (eds.), 4−15. Tilburg.

Bobrow, Daniel G., Cleo Condoravdi, Richard Crouch, Valeria de Paiva, Ronald M. Kaplan, Lauri Karttunen, Tracy Holloway King & Annie Zaenen (2005). A basic logic for textual inference. In *Proc. of the AAAI Workshop on Inference for Textual Question Answering.* Pittsburgh, PA.

Borgman, Christine L. (1996). Why are Online Catalogs *still* Hard to Use? *Journal of the American Society for Information Science* 47 (7): 493−503.

Bos, Johan & Malvina Nissim (2006). Cross-Lingual Question Answering by Answer Translation. In *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop*, Alessandro Nardi, Carol Peters & José Luis Vicedo (eds.). Alicante.

Bowden, Mitchell, Marian Olteanu, Pasin Suriyentrakorn, Jonathan Clark & Dan Moldovan (2006). LCC's PowerAnswer at QA@CLEF 2006. In *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop*, Alessandro Nardi, Carol Peters & José Luis Vicedo (eds.). Alicante.

Butt, Miriam, Helge Dyvik, Tracy H. King, Hiroshi Masuichi & Christian Rohrer (2002). The Parallel Grammar Project. In *Proc. of COLING-2002 Workshop on Grammar Engineering and Evaluation*, 1–7.

Clark, Herbert H. (1977). Bridging. In *Thinking: Readings in Cognitive Science,* Philip N. Johnson-Laird & Peter C. Wason (eds.), 411–420. Cambridge: Cambridge University Press.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag (2006). Minimal Recursion Semantics: An Introduction. *Research on Language and Computation* 3 (4): 281–332.

Fellbaum, Christiane (ed.) (1998). *Wordnet. An Electronic Lexical Database.* Cambridge, MA: MIT Press.

Frank, Anette, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg & Ulrich Schäfer (2005). Querying Structured Knowledge Sources. In *Proceedings of AAAI-05, Workshop on Question Answering in Restricted Domains*, 10–19. Pittsburgh, PA.

Glöckner, Ingo (2006). University of Hagen at QA@CLEF 2006: Answer Validation Exercise. In *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop*, Alessandro Nardi, Carol Peters & José Luis Vicedo (eds.). Alicante.

Glockner, Ingo, Sven Hartrumpf & Johannes Leveling (2007). Logical Validation, Answer Merging and Witness Selection – A Study in Multi-Stream Question Answering. In *Proceedings of the RIAO 2007.* Pittsburgh, PA.

Glöckner, Ingo, Sven Hartrumpf & Rainer Osswald (2005). From GermaNet Glosses to Formal Meaning Postulates. In *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen – Beiträge zur GLDV-Tagung 2005 in Bonn*, Bernharn Fisseni, Hans-Christian Schmitz, Bernhard Schröder & Petra Wagner (eds.), 394–07. Frankfurt: Peter Lang.

Gnörlich, Carsten (2002). *Technologische Grundlagen der Wissensverwaltung für die automatische Sprachverarbeitung.* Ph.D. thesis, FernUniversität Hagen.

Harabagiu, Sanda, Dan Moldovan, Marius Paşca, Rada Mihalcea, Mihai Surdeanu, Răzvan Bunescu, Roxana Gîrju, Vasile Rus & Paul Morărescu (2001). The Role of Lexico-Semantic Feedback in Open-Domain Textual Question-Answering. In *Proceedings of the ACL 2001,* 274–281. Toulouse.

Hartrumpf, Sven (2001). Coreference Resolution with Syntactico-Semantic Rules and Corpus Statistics. In *Proceedings of the Fifth Computational Natural Language Learning Workshop (CoNLL-2001)*, 137–144. Toulouse.

Hartrumpf, Sven (2003). *Hybrid Disambiguation in Natural Language Analysis.* Osnabrück: Der Andere Verlag.

Hartrumpf, Sven, Hermann Helbig, Johannes Leveling & Rainer Osswald (2006a). An Architecture for Controlling Simple Language in Web Pages. *eMinds: International Journal on Human-Computer Interaction* 1 (2): 93–112.

Hartrumpf, Sven, Hermann Helbig & Rainer Osswald (2003). The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement automatique des langues* 44 (2): 81–105.

Hartrumpf, Sven, Hermann Helbig & Rainer Osswald (2006b). Semantic Interpretation of Prepositions for NLP Applications. In *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions,* 29–36. Trento.

Hartrumpf, Sven & Johannes Leveling (2006). University of Hagen at QA@CLEF 2006: Interpretation and Normalization of Temporal Expressions. In *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop*, Alessandro Nardi, Carol Peters & José Luis Vicedo (eds.). Alicante.

Helbig, Hermann (1986). Syntactic-semantic Analysis of Natural Language by a new word-class controlled functional analysis. *Computers and Artificial Intelligence* 5 (1): 53–59.

Helbig, Hermann (2006). *Knowledge Representation and the Semantics of Natural Language.* Berlin: Springer.

Helbig, Hermann & Ingo Glöckner (2007). Reasoning over Semantic Networks – A Typology of Axioms for Natural Language Inference. In *Workshop on Modeling and Representation in Computational Semantics (MRCS-07)*, Gautam Sengupta, Soma Paul, Renjini Narendranath & Rajasekhar Batchu (eds.), 30–34. Hyderabad.

Hobbs, Jerry R., Mark E. Stickel, Douglas Appelt & Paul Martin (1993). Interpretation as Abduction. *Artificial Intelligence* 63 (1–2): 69–142.

Hovy, Eduard, Laurie Gerber, Ulf Hermjakob, Michael Junk & Chin-Yew Lin (2001). Question Answering in Webclopedia. In *Proceedings of the TREC-9 Conference.*

Katz, Boris (1997). Annotating the World Wide Web Using Natural Language. In *Proceedings of the RIAO 1997.* Montreal.

Laurent, Dominique, Patrick Séguéla & Sophie Nègre (2005). Cross Lingual Question Answering using QRISTAL for CLEF 2005. In *Working Notes for the CLEF 2005 Workshop.* Wien.

Lehmann, Fritz (ed.) (1992). *Semantic Networks in Artificial Intelligence.* Oxford: Pergamon Press.

Leveling, Johannes (2004). Feedback Mechanisms for a Natural Language Interface: An Application of the Critic Paradigm. In *Proceedings of the RIAO 2004,* 431–46. Avignon: C.I.D.

Leveling, Johannes (2006a). *Formale Interpretation von Nutzeranfragen für natürlichsprachliche Interfaces zu Informationsangeboten im Internet.* Tönning: Der andere Verlag.

Leveling, Johannes (2006b). University of Hagen at CLEF 2006: Reranking Documents for the Domain-specific Task. In *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop*, Alessandro Nardi, Carol Peters & José Luis Vicedo (eds.). Alicante.

Moldovan, Dan, Christine Clark, Sanda Harabagiu & Steve Maiorano (2003). COGEX: A logic prover for question answering. In *Proceedings of the NAACL-HLT 2003,* volume 1, 87–93. Edmonton.

Moldovan, Dan & Vasile Rus (2001). Explaining Answers with Extended Wordnet. In *Proceedings of the ACL 2001.* Toulouse.

Nyberg, Eric, Teruko Mitamura, James P. Callan, Jaime G. Carbonell, Robert E. Frederking, Kevin Collins-Thompson, Laurie Hiykumoto, Yifen Huang, Curtis Huttenhower, Scott Judy, Jeongwoo Ko, Anna Kupść, Lucian Vlad Lita, Vasco Pedro, David Svoboda & Benjamin Van Durme (2003). The JAVELIN Question-Answering System at TREC 2003: A Multi-Strategy Approach with Dynamic Planning. In *Proceedings of the TREC 12 Conference.*

Osswald, Rainer, Hermann Helbig & Sven Hartrumpf (2006). The Representation of German Prepositional Verbs in a Semantically Based Computer Lexicon. In *Proceedings of the LREC 2006*, 461–64. Genoa.

Peñas, Anselmo, Álvaro Rodrigo, Valentin Sama & Felisa Verdejo (2006). Overview of the Answer Validation Exercise 2006. In *Working Notes for the CLEF 2006 Workshop.* Alicante.

Pulman, Stephen G. (2007). Formal and Computational Semantics: a Case Study. In *Proceedings of the 7th International Workshop on Computational Semantics*, Jereon Geertzen, Elias Thijsse, Harry Bunt & Amanda Schiffrin (eds.), 181–196. Tilburg.

Stenglein, Brigitte (2006). *Evaluierung des natürlichsprachlichen Interfaces NLI-Z39.50.* Master's thesis, FernUniversität Hagen.

van Benthem, Johan & Alice ter Meulen (eds.) (1997). *Handbook of Logic and Language.* Amsterdam: North-Holland.

Vargas-Vera, Maria, Enrico Motta, Raul Monroy & Gustavo Arroyo-Figueroa (2004). AQUA: Ontology-based question answering system. In *MICAI 2004: Advances in Artificial Intelligence*, 468–477. Berlin: Springer.