Automatic Generation of MultiNet Representations for German Text Corpora

Sven Hartrumpf

Joint work with
Ingo Glöckner, Hermann Helbig, Johannes Leveling, Rainer Osswald

Intelligent Information and Communication Systems (IICS)
http://pi7.fernuni-hagen.de/
FernUniversität in Hagen
58084 Hagen, Germany

2007-04-02, Prague

Overview

1 Motivation

Many success stories for NLP applications relying on shallow methods only

But: there are many more NLP applications which call for deep methods

Here: deep semantic analysis of large text corpora

Many steps and methods spanning the range:

real-world text corpora
↓
formal semantic representations and
↓
their successful exploitation in applications

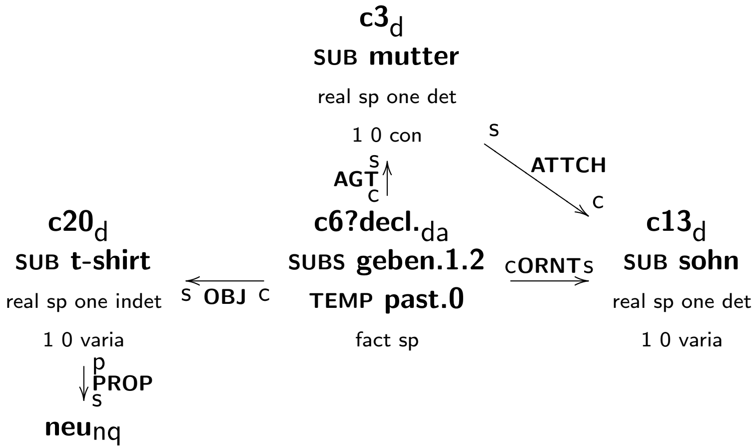
2 Target Representation

The semantic network formalism MultiNet (Helbig, 2006, 2001) provides means for homogenously representing the semantics of:

- words (also: multi-word expressions)
- phrases
- sentences
- texts
- text corpora

Target Representation (contd.)

Figure 1: Small multi-net (= semantic network of the MultiNet formalism) example:
Die Mutter gab ihrem Sohn ein neues T-Shirt.
The mother gave her son a new t-shirt.



3 Parsing

Syntactic-semantic parser (WOCADI, Helbig and Hartrumpf (1997); Hartrumpf (2003)) forms the center of automatic MultiNet construction

3.1 Morpholexical Analysis

Based on different lexica:

- ◇ deep lexicon (HaGenLex, Hartrumpf, Helbig, and Osswald (2003)):
  full information on orthography, morphology, syntactic-semantic valency
  26,000 entries: 13,500 nouns, 7,700 verbs, 3,300 adjectives

- ◇ flat lexicon:
  no semantics, no valency frames; sources: CELEX, UMLS
  47,500 additional entries

- ◇ name lexica: around 100 different lists with 500,000 entries
  for 46 different concept classes like city, country, region
  (some manually maintained; some extracted from Wikipedia etc.)

Figure 2: Abbreviated HaGenLex entry for the verb geben.1.2 used in Figure 1

$$
\begin{bmatrix}
\text{n-sign} & \\
\text{MORPH} & \begin{bmatrix} \text{BASE} & \text{"geben"} \\ \text{INFL-PARA} & \text{i143g} \end{bmatrix} \\
\text{SYN} & \begin{bmatrix} \text{v-syn} & \\ \text{PERF-AUX} & \text{haben} \end{bmatrix} \\
\text{SEMSEL} & \begin{bmatrix}
\text{SEM} & \begin{bmatrix} \text{sem} & \\ \text{ENTITY} & \text{nonment-action} \end{bmatrix} \\
\text{C-ID} & \text{"geben.1.2"} \\
\text{SELECT} & \left\langle \begin{array}{l}
\begin{bmatrix}
\text{REL} & \{\text{agt}\} \\
\text{SEL} & \begin{bmatrix}
\text{SYN} & \begin{bmatrix} \text{np-syn} & \\ \text{CAT} & \text{np} \\ \text{AGR} & \begin{bmatrix} \text{CASE} & \text{nom} \end{bmatrix} \end{bmatrix} \\
\text{SEMSEL} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{sem} & \\ \text{ENTITY} & \begin{bmatrix} \text{object} & \\ \text{LEGPER} & + \\ \text{POTAG} & + \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\begin{bmatrix}
\text{REL} & \{\text{ornt}\} \\
\text{SEL} & \begin{bmatrix}
\text{SYN} & \begin{bmatrix} \text{np-syn} & \\ \text{CAT} & \text{np} \\ \text{AGR} & \begin{bmatrix} \text{CASE} & \text{dat} \end{bmatrix} \end{bmatrix} \\
\text{SEMSEL} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{sem} & \\ \text{ENTITY} & \begin{bmatrix} \text{animal-object} \end{bmatrix} \vee \begin{bmatrix} \text{human-object} \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\begin{bmatrix}
\text{REL} & \{\text{obj}\} \\
\text{SEL} & \begin{bmatrix}
\text{SYN} & \begin{bmatrix} \text{np-syn} & \\ \text{CAT} & \text{np} \\ \text{AGR} & \begin{bmatrix} \text{CASE} & \text{acc} \end{bmatrix} \end{bmatrix} \\
\text{SEMSEL} & \begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{sem} & \\ \text{ENTITY} & \begin{bmatrix} \text{SORT} & \text{abs} \vee \text{co} \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{array} \right\rangle \\
\text{COMPAT-R} & \{\text{dur fin goal instr meth purp strt tlim}\} \\
\text{EXAMPLE} & \langle \text{"(Der Sozialstaat) (gibt) (den Hilfsbedürftigen) (Unterstützung)."} \rangle
\end{bmatrix}
\end{bmatrix}
$$

Morpholexical Analysis (contd.)

Morphology: derived from CELEX

Compound analysis: semantically oriented

Compounds are very productive in German

For better disambiguation: a compound lexicon with 30,000 entries

Specifying only segmentation and semantic class of the compound

Example entries:
("Staubecke.1.1" ("staub.1.1" "ecke.1.1") assoc-compd "SH 2006-06-28")
("Staubecken.1.1" ("stauen.1.1" "becken.1.1") purp-compd "SH 2006-06-28")

Staubecken: *dust corners* or *reservoir*

Semi-automatically maintained

3.2 WCFA

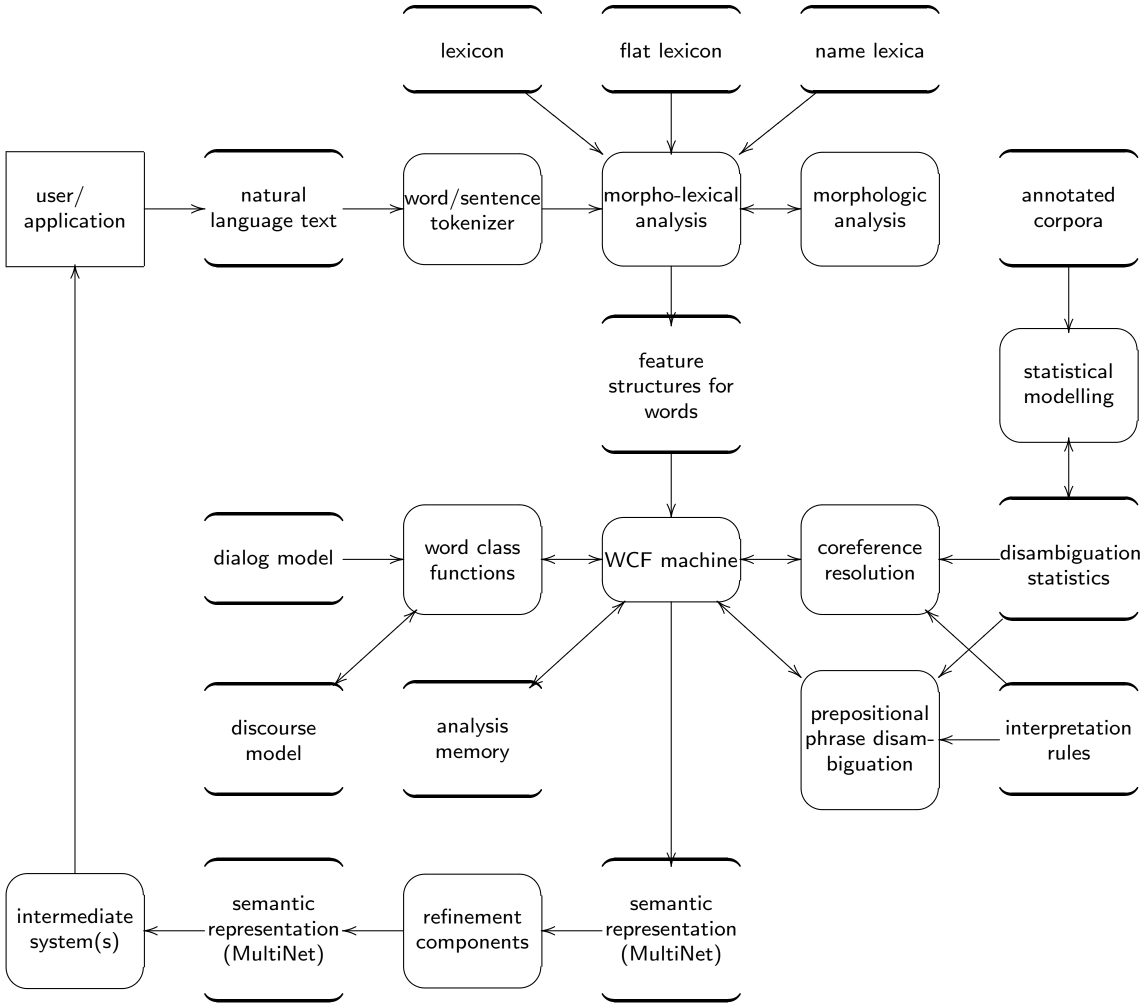
Word Class Function Analysis (WCFA, Helbig (1994, 1986))

No clean separation of parser and grammar; not so fine for maintenance?

Yes, to some extent; but the WCFA parser is

- an *experienced workhorse*
- easily extendible, e.g. by disambiguation modules

New implementation in the Scheme programming language:
WOrd ClAss based DIsambiguating (WOCADI, Helbig and Hartrumpf (1997); Hartrumpf (2003))


lexicon
flat lexicon
name lexica
user/ application
natural language text
word/sentence tokenizer
morpho-lexical analysis
morphologic analysis
annotated corpora
feature structures for words
statistical modelling
dialog model
word class functions
WCF machine
coreference resolution
disambiguation statistics
discourse model
analysis memory
prepositional phrase disambiguation
interpretation rules
intermediate system(s)
semantic representation (MultiNet)
refinement components
semantic representation (MultiNet)

WCFA Characteristics

For each word class (similar to PoS): a word class function (WCF) is implemented

WCFs together with a central control mechanism (WCF machine) model two phases of word activity

Opening act: opening expectations about what might or should follow

Completing act: saturating expectations by connecting complements/adjuncts to heads etc.

$\longrightarrow$ Some similarity to chart parsing

Oriented towards constructing multi-nets; useful by-product: dependency trees

Emphasis on understanding acceptable sentences, not on perfectly grammatical sentences

WCFA analyzes sentences on four levels:

Level 1 elementary kernel (e.g. simple NP)

Level 2 complex kernel (e.g. complex NP)

Level 3 simple proposition (phrase containing only one main verb, i.e. a clause)

Level 4 complex proposition (i.e. a complex sentence)

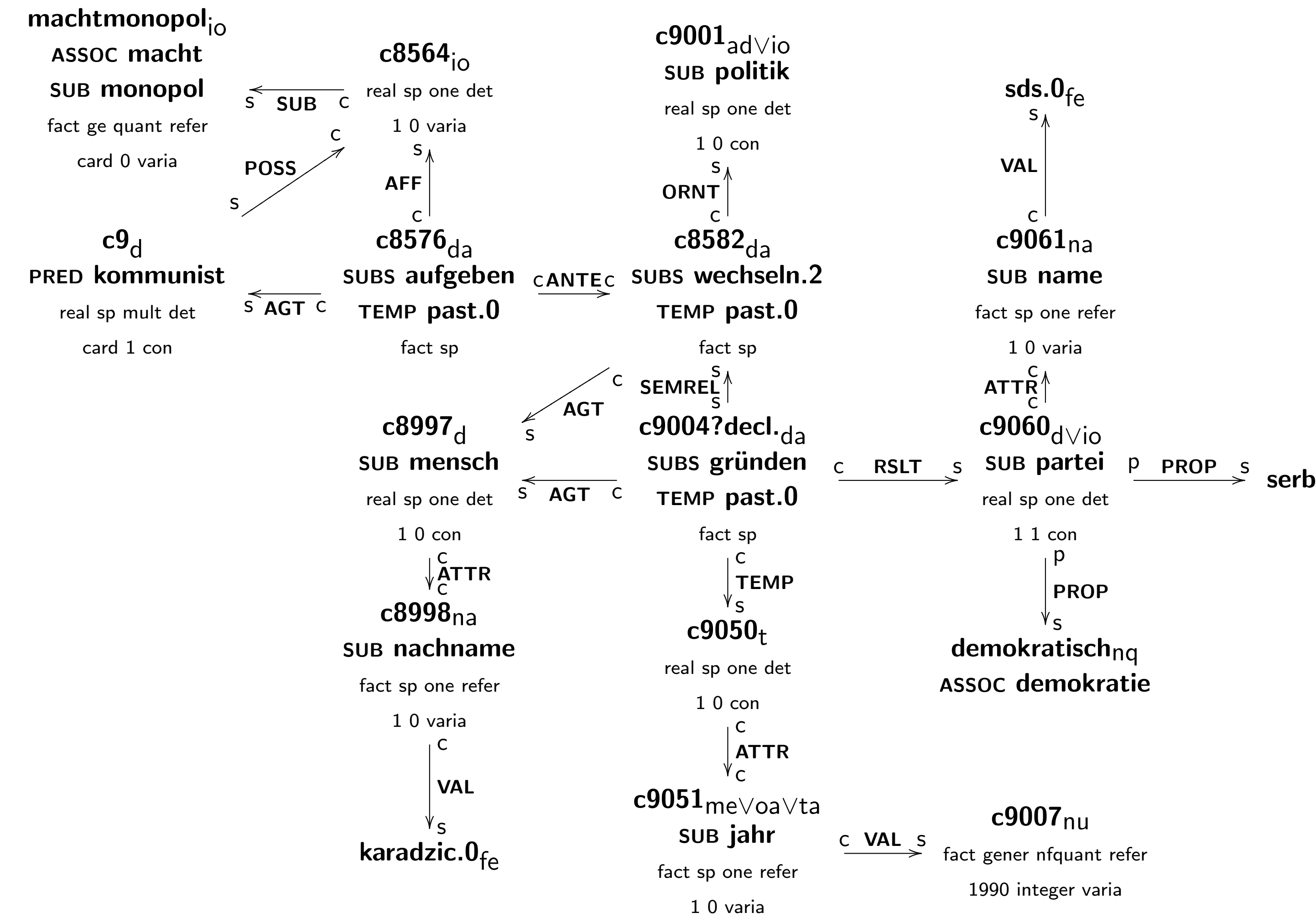
Example of a WCF: Simplified opening act of the WCF ADV

```
next-word-cat ← (fs-value *next-word '(syn cat))
adv-sort ← (fs-value *top '(semsel sem entity sort))
(follow-alternatives ; Each top-level if-clauses can lead to a parse alternative.
  if (and (equal? next-word-cat 'adv)
        (not (member? adv-sort '(l t))))
     ; The adverb modifies another adverb.
     (fs-set-value *top '(open-cat) 'advp) ; expect an adv phrase
  fi
  if (and (equal? next-word-cat 'a)
        (not (member? adv-sort '(ng))))
     ; The adverb seems to modify an adjective.
     (fs-set-value *top '(open-cat) 'ap) ; expect an adjective phrase
     if (and (fs-value=? *top2 '(open-cat) 'advp)
           (fs-value=? *top2 '(complete-cat) 'cat)) ; inside an open advp
        (set-status 'completing)
     elif (not (type-subsumed-d? adv-sort '(gr ql t))) ; adv is not a graduator
        (set-status 'closed)
        (fs-set-value *top '(open-cat) 'advp)
        (fs-set-value *top '(complete-cat) 'advp)
     fi
  fi

⋮
```

3.3 MultiNet Construction in the Parser

Example multi-net for the CLEF corpus sentence:
Nachdem die Kommunisten ihr Machtmonopol aufgegeben hatten, wechselte Karadzic in die Politik und gründete 1990 die Serbische Demokratische Partei (SDS).
After the communists gave up their power monopole, Karadzic moved to politics and founded in 1990 the Serbian Democratic Party (SDS).


machtmonopol$_{io}$
ASSOC macht
SUB monopol
fact ge quant refer
card 0 varia
SUB
POSS
c8564$_{io}$
real sp one det
1 0 varia
AFF
c9001$_{ad \vee io}$
SUB politik
real sp one det
1 0 con
ORNT
sds.0$_{fe}$
VAL
c9$_{d}$
PRED kommunist
real sp mult det
card 1 con
AGT
c8576$_{da}$
SUBS aufgeben
TEMP past.0
fact sp
ANTE
c8582$_{da}$
SUBS wechseln.2
TEMP past.0
fact sp
c9061$_{na}$
SUB name
fact sp one refer
1 0 varia
AGT
SEMREL
ATTR
c8997$_{d}$
SUB mensch
real sp one det
1 0 con
AGT
c9004?decl.$_{da}$
SUBS gründen
TEMP past.0
fact sp
RSLT
c9060$_{d \vee io}$
SUB partei
real sp one det
1 1 con
PROP
serb
ATTR
c8998$_{na}$
SUB nachname
fact sp one refer
1 0 varia
VAL
karadzic.0$_{fe}$
TEMP
c9050$_{t}$
real sp one det
1 0 con
ATTR
c9051$_{me \vee oa \vee ta}$
SUB jahr
fact sp one refer
1 0 varia
VAL
c9007$_{nu}$
fact gener nfquant refer
1990 integer varia
PROP
demokratisch$_{nq}$
ASSOC demokratie

Relational Layer

Where do the multi-net parts come from?

1. WCFs
   e.g. PROP, PROPR, *PMOD from WCF A(djective)

2. Valency frame of verbs, nouns, and adjectives
   Each argument specifies one or more relations leading from the mother node to the argument node
   Set-valued feature SEMSEL|SELECT|*N*|REL
   e.g. verb geben.1.2 (*give*; from Figure 1):
   AGT, ORNT, OBJ

3. Feature SEMSEL|SEM|NET of verbs, nouns, and adjectives
   Used if simple relations do not suffice for connecting arguments
   e.g. if the semantics consists of relations connecting the arguments directly
   e.g. verb bestehen.1.2 (*The table consists of wood.*): NET (origm x1 x2)
   A metaverb because it is the NL expression for a MultiNet primitive

4. Feature SEMSEL|SEM|NET of prepositions (similarly for conjunctions)
   e.g. *the shot in the morning* versus *the shot in the darkness*
   Can also be seen as PP interpretation rules (see Hartrumpf, Helbig, and Osswald (2006))

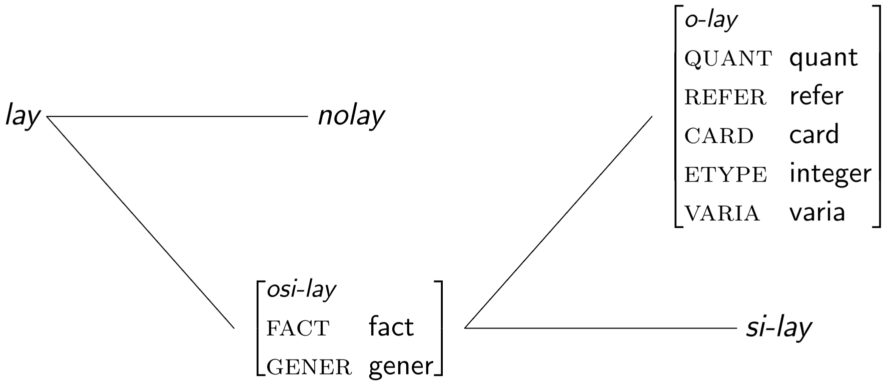
Relational Layer (contd.)

Other HaGenLex features with multi-net content:

- ◇ ENT-NET: filled automatically from semi-formal entailments
- ◇ EQU-NET: equivalences (special case of ENT-NET)
- ◇ LEX-NET: lexical relations to other lexical concepts (SYNO, ANTO, etc.)
- ◇ PRE-NET: presuppositions (planned)
- ◇ SEL-NET: selectional restrictions (planned)

Layer Features of Nodes

MultiNet nodes have layer features depending upon their sort:



Intensional layer features:
gener: genericity (generic vs. specific)
quant: quantification
refer: referentiality (determinate, indeterminate)
varia: variable reference

Extensional layer features:
card: cardinality
etype: extension type (individual, set, set of sets)
fact: facticity (real, hypothetical, nonreal)

Layer Features of Nodes (contd.)

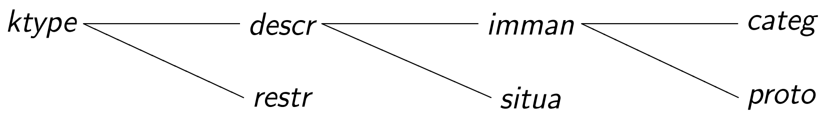
Sources:

- Lexicon
  esp. entries for determiners and quantifiers
  nouns: count nouns versus group nouns
- WCFs
  e.g. WCF n(oun): ETYPE modification depending upon NUM(erus)
- PP interpretation rules: CARD, ETYPE
- **Layer unification principle**: enforced in WCFs participating in NPs (Hartrumpf and Helbig, 2002)

But: some postprocessing needed (e.g. generic uses not handled in parser)
MWR (MultiNet Workbench: layer module)

Layer Features of Relations

Each MultiNet edge characterized by knowledge type (k-type) at start node and end node



Sources:

- ◇ Defaults for each relation (Helbig, 2006)
- ◇ Refinement in some WCFs, mainly refinements to *categ*(orical)
- ◇ Explicit k-types in PP interpretation rules

But: some postprocessing needed ⟶ MWR layer module

3.4 Evaluation of the Parser

Corpus	# Texts	# Sentences	# Words
TüBa-D/Z (Tübingen Treebank of Written German)		27,125	
HaGenLex examples		24,000	
CLEF (Cross-Language Evaluation Forum)	277,000	4,900,000	90,000,000
Wikipedia (German snapshot from 2006-09-25)	475,000	11,000,000	198,000,000

Coverage: around 60% for correct sentences (HaGenLex: near 100%)

Why not higher?

- Not only syntax, but also semantics
- Coverage rates multiply: e.g. $0.8[\text{syntactic coverage}] \cdot 0.8[\text{semantic coverage}] = 0.64$
- Main problems:
  Verb readings missing in the lexicon
  Complex sentences comprising many clauses
  Analysis depth
  Corpus-specific encoding of special characters like quotation marks

Accuracy: only investigated on small samples
But: the need to construct full semantics is often a good filter/disambiguator

Speed: depends on parser parameter *maximum number of WCF calls* $\longrightarrow$ coverage
typically: 1–2 seconds per sentence on current PCs

4 Disambiguation

WOCADI parser uses specialized disambiguation methods:

- ◇ prepositional phrase (PP) attachment and interpretation
- ◇ coreference resolution
- ◇ word sense disambiguation (WSD; in the context of PPs)

4.1 General Hybrid Disambiguation Approach

hybrid method (Hartrumpf, 2003)

1. linguistic rules ⟵ manual
2. disambiguation statistics ⟵ automatic (from annotated corpora)
   new statistical model: multidimensional back-off

Multidimensional Back-off Models

$R \in \mathbf{R}$ representation of an ambiguity problem (instance) as a list of alternatives

$\mathbf{R}$ set of all theoretically possible ambiguity problems

Training Corpus:

(1) $\mathbf{R}_{train}$: list of elements (R_j, i_j) with $R_j \in \mathbf{R}, i_j \in \{0, 1, \ldots, |R_j| - 1\}$

Initial counts:

(2) $C(R_j, i_j) :=$ absolute frequency of (R_j, i_j) in $\mathbf{R}_{train}$

Abstraction function:

(3) $a : \mathbf{L} \times \mathbf{R} \rightarrow \mathbf{A}$

(4) $a(l, R) \;=\; A_l$ with $l \in \mathbf{L}$, $r \in \mathbf{R}$, $A_l \in \mathbf{A}$

Abstraction function from dimension abstraction functions:

(5) $a(l,R) \;:=\; a_0(l,R,a_1,a_2,\ldots,a_d)$

(6) $a_i : \quad \mathbf{L}_i \times \mathbf{R}_i \rightarrow \mathbf{A}_i, \mathbf{L}_i := \{minl_i, minl_i+1, \ldots, maxl_i\}$ for $1 \leq i \leq d$

Back-off levels:

(7) $\mathbf{L} \;:=\; \mathbf{L_1} \times \mathbf{L_2} \times \ldots \mathbf{L_d}$

New compared to standard back-off models: **detail back-off** vs. **alternative back-off** as two dimensions

Example abstractions from coreference resolution

Detail back-off with three levels: $\mathbf{L_1} = \{1, 2, 3\}$
Alternative back-off with two levels: $\mathbf{L_2} = \{1, 2\}$

Level	Abstraction
$(1, 1)$	((ident.n_perspro, entity=substance, case=nom), (ident.n_perspro, entity=substance, case=acc), (ident.perspro_perspro, entity=object, case=nom))
$(1, 2)$	((ident.n_perspro, entity=substance), (ident.n_perspro, entity=substance), (ident.perspro_perspro, entity=object))
$(1, 3)$	((ident.n_perspro, entity=con-object), (ident.n_perspro, entity=con-object), (ident.perspro_perspro, entity=object))
$(2, 1)$	((ident.n_perspro, entity=substance, case=nom), (ident.n_perspro, entity=substance, case=acc))
$(2, 2)$	((ident.n_perspro, entity=substance), (ident.n_perspro, entity=substance))
$(2, 3)$	((ident.n_perspro, entity=con-object), (ident.n_perspro, entity=con-object))

Learning phase: $|\mathbf{L}|$ statistical models by multidimensional back-off

Back-off direction function:

(8) $b : \{0, 1, \ldots, |\mathbf{L}| - 1\} \rightarrow \mathbf{L}$

(9) $\forall j, k \in \{0, 1, \ldots, |\mathbf{L}| - 1\}(j \neq k \longrightarrow b(j) \neq b(k))$

Statistical models:

(10) $C_l(A, i) := \sum_{R \in \mathbf{R} \text{ with } a(l,R)=A \,\wedge\, s(l,R)=\mathit{false}} C(R, i) + acc(l, R, i)$

with $l \in \mathbf{L}$, $A \in \mathbf{A}$, $0 \leq i <$ Zahl der Alternativen in A

acc (accumulation): also count problems with additional alternative

Application phase: find relevant model, relative frequencies, probabilities

Probabilities for first level l_0:

$$(11) \quad f(l_0, R, i) := \frac{C_{l_0}(a(l_0, R), i)}{\sum_{j=0}^{n} C_{l_0}(a(l_0, R), j)} \qquad \text{with } i \in \{0, 1, \ldots, n\}$$

$$(12) \quad p(i|R) \approx f(l_0, R, i)$$

Stopping criterion: absolute threshold t_a

$$(13) \quad c(l, R) := \begin{cases} \mathit{true} & \text{if } \exists i \in \{0, \ldots, m-1\}(C_l(a(l, R), i) \geq t_a) \\ \mathit{false} & \text{otherwise} \end{cases}$$

m: number of alternatives in $a(l, R)$

instead (or in addition): relative threshold

Back-off to next model C_{l_k}:

(14) $$f(l_k, R, i) := \frac{C_{l_k}(A, i)}{\sum_{j=0}^{m-1} C_{l_k}(A, j)}$$

with $A = a(l_k, R), m =$ number of alternatives in $A, i \in \{0, \ldots, m-1\}$

(15) $$p(i|R) \approx f(l_k, R, i)$$

5 Coreference Resolution

WOCADI parses a text sentence by sentence

Resolves only absolutely necessary coreferences (involving relative pronouns or relative determiners)

Coreference resolution afterwards by the coreference module (CORUDIS):
determines antecedent of each mention (markable) as defined in MUC

Leads to additional EQU relations
between nodes of sentence multi-nets or
between nodes of a single multi-net

Trained and evaluated on a corpus annotated by us: GeCCo (German Coreference Corpus)

6 Semantic Assimilation

Further aspects for semantics of a whole text:
during **assimilation** as a post-interpretation step after parsing (refinement components)

Rationale: delay all semantic and pragmatic aspects that need
discourse model, dialog model, or domain knowledge

Only some parts implemented:

- ◇ Turning the EQU relations from CORUDIS into modified multi-nets:
  only one node for all mentions in a coreference chain
- ◇ Deixis resolution: temporal part implemented
  40 MultiNet rules
  Successfully employed for question answering (CLEF 2006)
- ◇ Intertextual coreference resolution (e.g. event tracking): planned

7 Applications

Implemented applications that build on deep semantic representations:

- Question answering (QA system InSicht):
  participating at Cross-Language Evaluation Forum (CLEF) since 2004
  (Hartrumpf, 2005, 2006; Hartrumpf and Leveling, 2006)

- Information retrieval (IR): CLEF track GIRT (German Indexing and Retrieval Test database)
  parsing documents/queries
  (Leveling, 2006b)

- Geographic information retrieval (GIR): CLEF track GeoCLEF
  parsing documents/queries; metonymy recognition for location names
  (Leveling and Veiel, 2006)

- Natural language interfaces (NLIs): for bibliographic databases (Z39.50) and SQL databases
  (Leveling, 2006a)

- Measuring and criticizing readability of NL texts
  (EU project *Benchmarking Tools and Methods for the Web*, BenToWeb)
  e.g. ambiguous pronoun references, distant antecedents, semantic complexity
  (Jenge, Hartrumpf, Helbig, and Osswald, 2006; Hartrumpf et al., 2006)

- Support for semi-automatic knowledge assimilation

Semantics at Work: Details from Question Answering

Question (variant of CLEF question qa06_079):
In welchem Jahr verschied Charles de Gaulle?
In which year did Charles de Gaulle pass away?

Document sentence (SDA.951109.0236.88):
Frankreichs Staatschef Jacques Chirac hat die Verdienste des vor 25 Jahren gestorbenen Generals und Staatsmannes Charles de Gaulle gewürdigt.
France's chief of state Jacques Chirac acknowledged the merits of general and statesman Charles de Gaulle, who died 25 years ago.

Answer from InSicht:
im Jahr 1970
in 1970

Found by comparing semantic networks and inferencing:

- Related verb readings are identified by lexical relations from HaGenLex: verscheiden.1.1 and sterben.1.1
- Temporal deixis is resolved: *vor 25 Jahren* and date of publication (from article start or article metadata)

8 Conclusion and Future Work

Work at the FernUniversität in Hagen (IICS):

- ◇ Deep semantic approach to NL understanding
- ◇ MultiNet as integrating language across all linguistic levels
- ◇ Implemented for use on large corpora
- ◇ Basis for NLP applications, e.g. successful question answering system for QA@CLEF

Plans for the future:

- ◇ Extension of analysis to English:
  English lexicon is being constructed in parallel to German (HaGenLex)
  WCFs are being parametrized for English
- ◇ German: extend coverage
- ◇ Further parts of semantic assimilation, e.g. intertextual coreferences (event and entity tracking)

Questions?

References

Hartrumpf, Sven (2003). *Hybrid Disambiguation in Natural Language Analysis*. Osnabrück, Germany: Der Andere Verlag. 5, 8, 21

Hartrumpf, Sven (2005). Question answering using sentence parsing and semantic network matching. In *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004* (edited by Peters, C.; P. Clough; J. Gonzalo; G. J. F. Jones; M. Kluck; and B. Magnini), volume 3491 of *Lecture Notes in Computer Science*, pp. 512–521. Berlin: Springer. URL http://dx.doi.org/10.1007/11519645_50. 30

Hartrumpf, Sven (2006). Extending knowledge and deepening linguistic processing for the question answering system InSicht. In *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, Revised Selected Papers* (edited by Peters, Carol; Fredric C. Gey; Julio Gonzalo; Gareth J. F. Jones; Michael Kluck; Bernardo Magnini; Henning Müller; and Maarten de Rijke), volume 4022 of *Lecture Notes in Computer Science*, pp. 361–369. Berlin: Springer. URL http://dx.doi.org/10.1007/11878773_41. 30

Hartrumpf, Sven and Hermann Helbig (2002). The generation and use of layer information in multilayered extended semantic networks. In *Proceedings of the 5th International Conference on Text, Speech and Dialogue (TSD 2002)* (edited by Sojka, Petr; Ivan Kopeček; and Karel Pala), volume 2448 of *Lecture Notes in Artificial Intelligence*, pp. 89–98. Brno, Czech Republic. URL http://springerlink.metapress.com/content/qdp316huhhhh/. 18

Hartrumpf, Sven; Hermann Helbig; and Rainer Osswald (2003). The semantically based computer lexicon HaGenLex – Structure and technological environment. *Traitement automatique des langues*, 44(2):81–105. 5

Hartrumpf, Sven; Hermann Helbig; and Rainer Osswald (2006). Semantic interpretation of prepositions for NLP applications. In *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*, pp. 29–36. Trento, Italy. URL http://www.aclweb.org/anthology/W06-2105. 15, 30

Hartrumpf, Sven and 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* (edited by Nardi, Alessandro; Carol Peters; and José Luis Vicedo). Alicante, Spain. URL http://www.clef-campaign.org/2006/working_notes/workingnotes2006/hartrumpfCLEF2006.pdf. 30

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. 8

Helbig, Hermann (1994). Der Einsatz von Wortklassenagenten für die automatische Sprachverarbeitung – Teil II – Die vier Verarbeitungsstufen. Informatik-Bericht 159, FernUniversität Hagen, Hagen, Germany. URL http://pi7.fernuni-hagen.de/papers/helbig-1994-hagen.pdf. 8

Helbig, Hermann (2001). *Die semantische Struktur natürlicher Sprache: Wissensrepräsentation mit MultiNet*. Berlin: Springer. URL http://www.springer.com/sgw/cda/frontpage/0,11855,1-40109-22-2122199-0,00.html. 3

Helbig, Hermann (2006). *Knowledge Representation and the Semantics of Natural Language*. Berlin: Springer. URL http://www.springer.com/sgw/cda/frontpage/0,11855,1-40109-22-72041224-0,00.html. 3, 19

Helbig, Hermann and Sven Hartrumpf (1997). Word class functions for syntactic-semantic analysis. In *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing (RANLP'97)*, pp. 312–317. Tzigov Chark, Bulgaria. 5, 8

Jenge, Constantin; Sven Hartrumpf; Hermann Helbig; and Rainer Osswald (2006). Automatic control of simple language in web pages. In *Proceedings of the 10th International Conference on Computers Helping People with Special Needs (ICCHP 2006)* (edited by Miesenberger, Klaus; Joachim Klaus; Wolfgang Zagler; and Arthur Karshmer), volume 4061 of *Lecture Notes in Computer Science*, pp. 207–214. Berlin: Springer. URL http://dx.doi.org/10.1007/11788713_31. 30

Leveling, Johannes (2006a). *Formale Interpretation von Nutzeranfragen für natürlichsprachliche Interfaces zu Informationsangeboten im Internet*. Der andere Verlag, Tönning, Germany. 30

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* (edited by Nardi, Alessandro; Carol Peters; and José Luis Vicedo). Alicante, Spain. URL http://www.clef-campaign.org/2006/working_notes/workingnotes2006/levelingCLEF2006.pdf. 30

Leveling, Johannes and Dirk Veiel (2006). University of Hagen at GeoCLEF2006: Experiments with metonymy recognition in documents. In *Results of the CLEF 2006 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2006 Workshop* (edited by Nardi, Alessandro; Carol Peters; and José Luis Vicedo). Alicante, Spain. URL http://www.clef-campaign.org/2006/working_notes/workingnotes2006/veielCLEF2006.pdf. 30