Morphological disambiguation in German

Karel Vaculík

Xerox Incremental Deep Parsing System (XIP)

- Disambiguation of noun phrases

- Two types of contextual rules:

  1. Ordinary disambiguation rules,

  2. Double reduction rules

- Syntactic heuristics for refinement

Xerox Incremental Deep Parsing System (XIP)

Ordinary disambiguation rules

- General form:

    readings_filter = |left_context| selected_readings |right_context|

- Example:

    det, pron = det |adj*, noun|

Xerox Incremental Deep Parsing System (XIP)

Double reduction rules

- General form:

    |node_sequence| => boolean_constraints.

- Example:

    |adj*, adj#1, adj*, noun#2| => (#1[agr] :: #2[agr]).

GERTWOL

- System for automatic recognition of German word forms

- Two types of morphology disambiguation:

    1. Local disambiguation,

    2. Contextual disambiguation

GERTWOL

Local disambiguation

- Context is not considered

- Retains only those readings with the fewest suffixes or composition borders

- Example: "<zugriffsbereite>"

  - "zug#riff\s|bereit" A POS SG NOM FEM
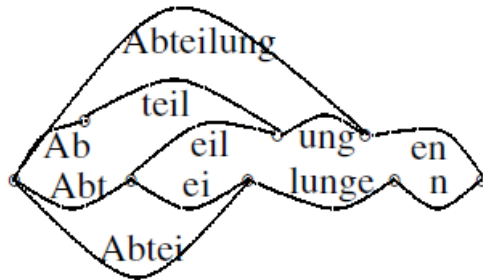 - "zu|griff\s|bereit" A POS SG NOM FEM

GERTWOL

Contextual disambiguation

- Grammatical rules
  - functional area (domain), target, operator and contextual conditions

- Heuristic rules

Other approaches

- Head-lexicalized probabilistic context –free grammar

  - Split words into morpheme sequences using morphology analyzer. Example:

  - Parse with PCFG. Used grammar is quite small and its probabalities are trained on unlabeled data with LoPar parser. It is using the Inside-Outside algorithm which is an instance of the unsupervised EM algorithm.

Other approaches

- Xerox HMM tagger
  - German model is created for this

- SMES – system for information extraction
  - Morphological disambiguation is carried out by combination of Brill-based unsupervised tagger and word-case sensitive rules

References

- Hinrichs, E., Trushkina, J.: Forging Agreement: Morphological Disambiguation of Noun Phrases. In *Proceedings of the First Workshop on Treebanks and Linguistic Theory*. 2002. pp 78—95.

- GERTWOL: http://www2.lingsoft.fi/doc/gertwol/intro/overview.html, http://www2.lingsoft.fi/doc/gercg/NODALIDA-poster.html

- Schmid, H.: Disambiguation of Morphological Structure using a PCFG. In *Proceeding HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 2005. pp 515—522

References

- LoPar: http://www.ims.uni-stuttgart.de/tcl/SOFTWARE/LoPar.html
- Feldweg, H.: Implementation and evaluation of a German HMM for POS disambiguation. In *Proceedings of the EACL SIGDAT Workshop*. 1995.
- Neumann, G. et al.: An Information Extraction Core System for Real World German Text Processing. In *proceedings of ANLP-1997*, Washington, DC, pages 209-216.