

GERMANET

GermaNet is a lexical-semantic net that relates German nouns, verbs, and adjectives semantically by grouping lexical units that express the same concept into synsets. Conceptual and lexical relations are defined within the network. For each of the word categories the semantic space is divided into a number of semantic fields. GermaNet can be viewed as an on-line thesaurus or a light-weight ontology.

GermaNet has been developed and maintained within various projects by Prof. Dr. Erhard W. Hinrichs and his research group for General and Computational Linguistics, Department of Linguistics at the University of Tübingen since 1997. It has been integrated into the EuroWordNet, a multilingual lexical-semantic database.

WHAT IS THE IDEA BEHIND GERMANET?

GermaNet is being developed to serve as an electronic lexicographic reference database for German word senses. It is primarily intended to serve as a resource for word sense disambiguation which is crucial for natural language applications like information retrieval, the construction of various language technology tools, and the annotation of corpora.

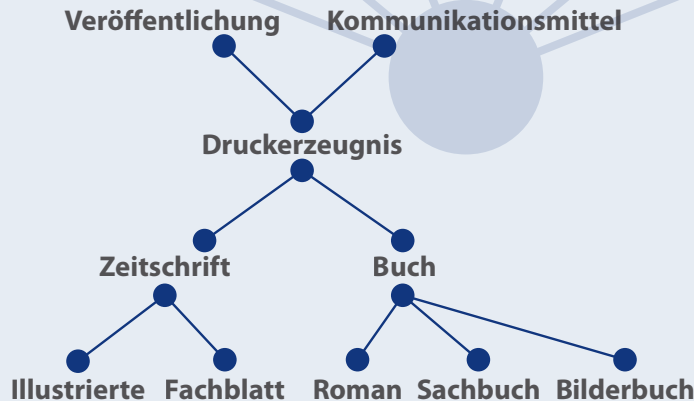
GERMANET & WORDNET®

The English WordNet® served as a basis for GermaNet. Thus these two wordnets share the same idea and have much in common. Nevertheless there are some fundamental differences; e.g. the following applies to GermaNet:

- Adjectives are ordered hierarchically
- GermaNet is represented as a completely connected graph
- Causation is encoded as a cross-class relation
- Uniform treatment of meronymy

THE HIERARCHICAL STRUCTURE

The hyperonymy-hyponymy relation constitutes the hierarchical structure of GermaNet as shown in the following figure, which reveals a part of the hierarchy.



SYNSETS, LEXICAL UNITS & RELATIONS

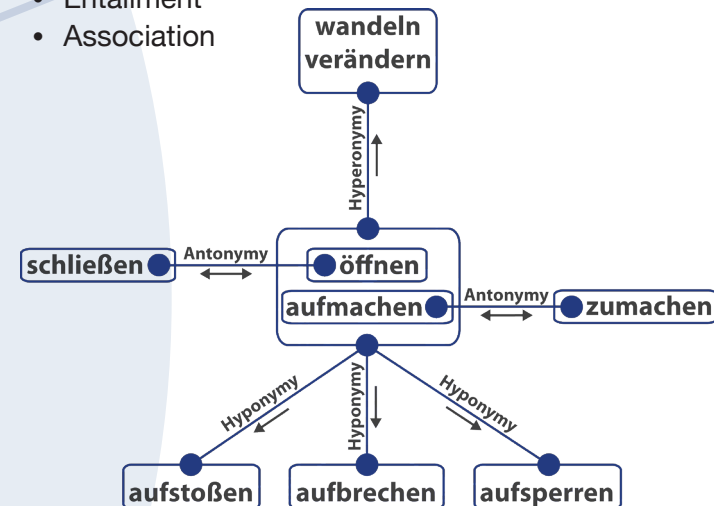
Synsets constitute the basic units within GermaNet. A synset is a set of lexical units (i.e. lexical realizations) that express the same concept. Thus all lexical units that are grouped in the same synset are synonyms.

There are two types of relations: conceptual and lexical. Lexical relations hold between lexical units. Besides the synonymy relation, the following lexical relations appear in GermaNet:

- Antonymy
- Pertonymy
- Participle

Conceptual relations connect synsets. The following conceptual relations are encoded in GermaNet:

- Hyperonymy/hyponymy
- Meronymy/holonymy
- Causation
- Entailment
- Association



COVERAGE

In its initial version GermaNet covers the German base vocabulary. The base vocabulary is determined by statistical analysis of a reference corpus. Since lexicographic work is done with respect to semantic fields, not by decreasing frequencies of lemmas, it is not always easy (nor appropriate) to omit words with frequencies below a certain threshold. For this reason we expect GermaNet to cover more than just the base vocabulary.

The following is an up-to-date statistics of GermaNet's version 5.2 contents (release December 2009):

- Number of literals: 76981
- Readings per literal: 1,10
- Number of synsets: 61575
- Number of lexical units: 84859
- Lexical units per synset: 1,38

APPLICATIONS & TOOLS

There are several applications and tools for GermaNet. See our website for further details.

- Application Programming Interfaces: are available for Java and for Perl.
- GermaNet-Explorer: A software to visualize GermaNet.
- GernEdiT - The GermaNet Editing Tool: is a graphical editor that allows maintaining the GermaNet data in a user-friendly way. It is used by the lexicographers to enter lexicographic data into the GermaNet database.
- Web Interface: to access GermaNet over the internet.

LICENSE

There are three types of license agreements for different applications and users:

- Academic Research License Agreement
- Research & Development License Agreement
- Commercial License Agreement

Please contact our homepage (<http://www.sfs.uni-tuebingen.de/GermaNet/licenses.shtml>) for further details.

CONTACT

Prof. Dr. Erhard W. Hinrichs
Eberhard Karls Universität Tübingen
Seminar für Sprachwissenschaft
Wilhelmstr. 19
72074 Tübingen
Germany

Phone: +49 7071 2974279

Fax: +49 7071 295214

Mail: germanetinfo@uni-tuebingen.de

Web: www.sfs.uni-tuebingen.de/GermaNet/



seminar für sprachwissenschaft



A LEXICAL-SEMANTIC WORDNET FOR GERMAN

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

