

LEXIKON

- Lexikon
- Wort und Begriff
- Repräsentationsschemata; Lexikon-Realisierungen
- Hierarchische Lexika: Lexikalisch-semantische Wortnetze

Lexikon

Lexikon: Zusammenstellung des Wortschatzes einer Sprache

Abstrakte Basiseinheit: *Lexem*

realisiert in verschiedenen Wortformen

z.B. *schreib-* in *schreibst*, *schrieb*, *schrieben*

als Teil anderer Lexeme

z.B. *Schreiberling*, *Geschriebenes*, *vorschreiben*

Trend in der Computerlinguistik: *Lexikalisierung*

Lexikon muss Daten für verschiedene sprachliche Ebenen enthalten:

Wort und Begriff

Begriffe: abstrakte Objekte — extensional oder intensional gegeben

Problem: Beziehung zwischen linguistischem und Weltwissen

Modellierung von Weltausschnitten: "formale Ontologie" nach methodischen

Kriterien: Konsistenz, Expliztheit, Stringenz, Eindeutigkeit

— definitorisch, Wortverwendung in einem (Fach-) Gebiet vorschreibend —

vs.

Modellierung sprachlichen Wissens (Semantik!) nach linguistischen Kriterien

— empirisch, allgemeine Wortverwendung beschreibend — (Grenze fließend)

Schnittstelle Wortsemantik:

Korrespondenz zwischen linguistischer Kategorisierung lexikalischer Elemente und

Typen (Sorten) in der Ontologie

Psycholinguistik: "Parallel interface model" (Friederici)

Zwei kollaterale Module: Linguistik — Weltwissen

Repräsentationsschemata: Entwurf und Realisierung

Gleichgewicht zwischen lexikographischen/ linguistischen Anforderungen, formaler Ausdruckskraft und Effizienz der Verarbeitung

Erfordernisse unterschiedlicher Anwendungen

Informatikaspekte: Effiziente Datenbankstrukturen, Compilation in Netzwerke, Transformation zwischen verwandten Grammatikformalismen

Lexikalische Information

- "Intra-Wort": Kontexte des Wortvorkommens
syntaktisch, semantisch, pragmatisch
- "Inner-Wort": Wortbildung, Synonymie

Organisation

- statische Information
- dynamische Information (Regeln)

Mögliche Realisierungen des Lexikons (Troost)

- *Lexikon als Teil der Grammatik*
- *Vollformenlexikon*
- *Grundformen- oder Lexemlexikon*

(c) Troost

Lexikon als Teil der Grammatik

- Lexikon wird in Grammatik integriert
- Wortformen sind Terminalsymbole der Grammatik
- Daher kann mit Wortformen keine weitere Information assoziiert werden

Nur für formale Sprachen mit kleinem Wortschatz geeignet.

Assoziation von Information in den Grammatikregeln möglich, aber:

Generalisierungen können nicht dargestellt werden.

Vollformenlexikon

- Für jede Wortform ein eigener Lexikoneintrag
- Wortformen sind atomare Einheiten — keine interne Struktur
- Beim Eintrag kann wortformenspezifische und wortspezifische Information gespeichert werden
- Lexikalische Kategorien sind Terminalsymbole der Grammatik
- Ersetzung der lexikalischen Kategorie durch Wortform geschieht durch separaten Lexikonzugriff

Problem: Hohe Redundanz

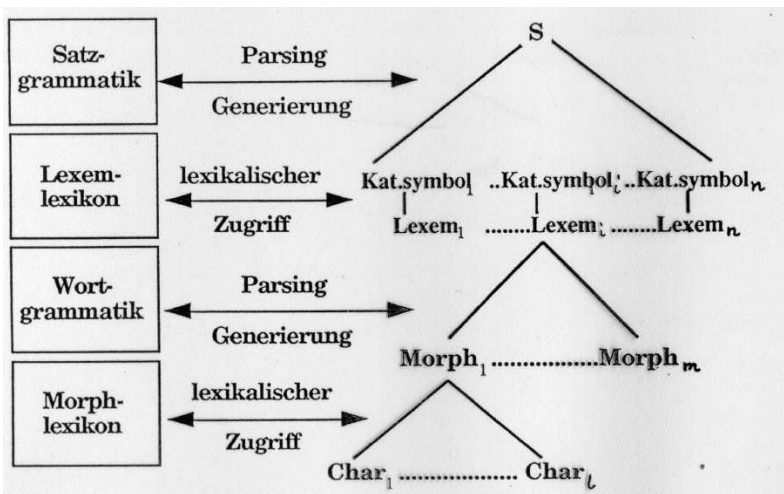
Grundformen- oder Lexemlexikon

- Ein Lexikoneintrag für jedes Wort (Lexem)
- Beim Eintrag kann nur wortspezifische Information gespeichert werden
- Wortformen werden als strukturierte Einheiten betrachtet
- Wortformen werden durch morphologische Komponente auf das zugehörige Lexem abgebildet
- Gleichzeitig muss diese wortformenspezifische Information erzeugen
- Ersetzung der lexikalischen Kategorie durch Wortform geschieht durch morphologische Komponente, die dabei auf Lexikon zugreift

Autonome morphologische Komponente (nach Trost)

- Morphlexikon enthält Einträge für alle Morphe
- Morphologische Komponente bildet Wortformen auf das zugehörige Lexem (Morphkette) ab
- Hierzu wird Morphlexikon und Wortbildungsgrammatik bzw. Lexemlexikon benutzt
- Lexemlexikon enthält Einträge für die Wörter (Lexeme), diese tragen nur wortspezifische Information
- Wortformen werden als aus Morphen aufgebaute komplexe Einheiten betrachtet

Abbildung: Lexikalische Ebenen (Trost)



Thesaurus

Thesaurus: Systematik von Termini (Wörtern)

ROGET: Hervorhebung von *Synonymie* vs. Begriffsumfang (enger/ weiter)

Table 2.6 Roget's system

class	section	given code	
1. abstract relations	existence	1-8	
	relation	9-24	
	quantity	25-57	
	order	58-83	
	number	84-105	
	time	106-139	
2. space	change	140-152	
	causation	153-179	
	space in general	180-191	
	dimensions	192-239	
	form	240-263	
	motion	264-315	
3. matter	matter in general	316-320	
	inorganic matter	321-356	
	organic matter	357-449	
	mind	450-454	
4. intellect (the exercise of the mind)	(1) formation of ideas	455-466	
	general	467-475	
	precursory conditions and operations	476-479	
	materials for reasoning	480-504	
	reasoning processes	505-513	
	results of reasoning	514-515	
	extension of thought		
	creative thought		
	5. volition (the exercise of the will)	(1) individual volition	516-524
		nature of ideas communicated	525-549
		models of communication	550-599
means of communicating ideas			
volition in general		600-619	
prospective volition		620-679	
voluntary action		680-703	
antagonism		704-728	
results of action		729-736	
(2) social volition		737-759	
general social volition		760-767	
special social volition	768-774		
conditional social volition	775-819		
possessive relations			
6. emotion, religion, and morality	general	820-826	
	personal emotion	837-887	
	interpersonal emotion	888-921	
	morality	922-975	
	religion	976-1000	

Thesauri für Fachterminologien

Fachthesauri: Struktur des Themengebiets (Technische Terminologie: Normen)

Relationen zwischen Substantiven

1. Synonyme, Antonyme
2. breitere, engere Termini
3. ähnliche Wörter
4. Relationen: Teil/Ganzes, Ordnung, Ursache/Wirkung, Nachfolge, logische Beziehung, gemeinsame Merkmale
5. Komposita, abgeleitete Wörter

Relationen zwischen Substantiven und Verben

1. Subjekt des Verbs *Ein Hund* bellt.
2. Objekt *ein Buch* kaufen.
3. Nutznießer etwas *einem Hund* geben.
4. Werkzeug, Methode mit *einem Löffel* essen.
5. Ort nach *Erlangen* fahren.
6. Zeit um *fünf Uhr* aufstehen.
7. Ursache wegen *einer Panne* zu spät kommen.
8. Rolle als *Sprecher* amtieren.

Beispiel: INSPEC Thesaurus

Table 2.7 Symbols for expressing the relations among nouns

hierarchical	BT	broader term
	NT	narrower term
generic	BTG	broader term generic
	NTG	narrower term generic
partitive	BTP	broader term partitive
	NTP	narrower term partitive
associative	RT	related term
antonymic	A	
equivalence	USE	use
	UF	used for
	USE+	use in combination

Table 2.8 Partial sample of INSPEC (International Information Services for the Physics and Engineering Communities) thesaurus

integrated circuits		TT networks (circuits)
UF IC		RT integrated-circuit manufacture
microcircuits		integrated-circuit technology
microelectronics		integrated-circuit testing
NT digital-integrated circuits		masks
hybrid-integrated circuits		modules
linear-integrated circuits		radiation hardening (electronics)
microwave-integrated circuits		semiconductor devices
monolithic-integrated circuits		substrates
thick film circuits		thick films
thin film circuits		
BT networks (circuits)	CC B2220 B2570	
	FC b2220-s b2570-r	
	DI January 1973	

TT: top term(s)
 CC: version of the code as used in the INSPEC Magnetic Tape Services
 FC: full form of code as input to the INSPEC database
 DI: date of entry or input

Lexikonkomponente: Separat oder integriert?

- Modulspezifische Lexikonkomponenten (Wartbarkeit?!)
- Zentrales Lexikon; Informationen für verschiedene Beschreibungsebenen
 - Integriert mit Morphologiekomponente
 - Semistrukturierte Repräsentation (SGML / XML mit zugehöriger DTD)
 - Codierung linguistischer Informationen in einem einheitlichen Lexikonmodell (z.B. DATR)
 - (Relationale) Datenbanken als Repositorien; Datenkonvertierung in anwendungsspezifische Formate

Anforderungen an Lexikonformalismen

Allgemeine Anforderungen

- Ausdrucksstärke: Zu beschreibende Phänomene
- Deklarative Beschreibungsmittel; Flexibilität und Effizienz bzgl. Verarbeitungsmodi (v.a. Zugriff, Suche)
- Linguistischer Bezug der Ausdrucksmittel; Theorienneutralität
- Wartbarkeit

Anforderungen aus Lexikon-Erstellung und Pflege

- Redundanzfreiheit, Faktorisierbarkeit, Konsistenzkontrolle
- Dokumentierte Spezifikation der Struktur der Einträge
- Schnittstellen für Import und Export

Lexikon: Modularisierung

- Problem der Domänenabhängigkeit
- Modularisierung ermöglicht teilweise Wiederverwendung
 - Domänenunabhängiger Teil:
Oberflächenstruktur, Valenz, Wortform, Kategorie, lexikalische Semantik
 - Domänenabhängiger Teil:
Präferierte Lesart, Spezifische Semantik: Schnittstelle zur formalen Domänenontologie, Verbindungen zu Domänen-Konzepten

Lexikalisch-semantische Wortnetze: (Euro) WordNet

Hierarchisches Lexikon mit Darstellung der bedeutungstragenden Beziehungen zwischen Wörtern.

Beispiel:

Sitzmöbel
Stuhl
Drehstuhl
Klappstuhl
Kinderstuhl
Bank
Parkbank
Gartenbank
TEILE: Sitzfläche, Lehne

Bildung *lexikalischer Konzepte* durch **Synsets** (Synonym-Mengen); Unterscheidung von Lesarten

Grundbeziehungen in Wortnetzen

- Lexikalische Relationen: **Synonymie**, Synset-intern (*Ruf*, *Leumund*) und **Antonymie** (*Geburt* / *Tod*, *glauben* / *zweifeln*, *schön* / *hässlich*)
- Konzept-Relationen: Bestehen zwischen Konzepten, gelten also für alle Realisierungen innerhalb eines Synset
 - **Hyponymie**, invers: **Hyperonymie**, hierarchiebildend (*Rotkehlchen* und *Vogel*)
 - **Meronymie**, invers: **Holonymie**, Teil-Ganzes-Beziehung (*Dach* und *Gebäude*)
 - **Kausation** (*töten* und *sterben*, *öffnen* und *offen*)
 - “**Entailment**” (*gelingen* und *versuchen*)

Weiterhin: **Kreuzklassifikation** (*Banane* als *Pflanze* und *Nahrungsmittel*)

Subkategorisierungsrahmen: Syntax-Semantik-Schnittstelle

(Euro) WordNet Synsets

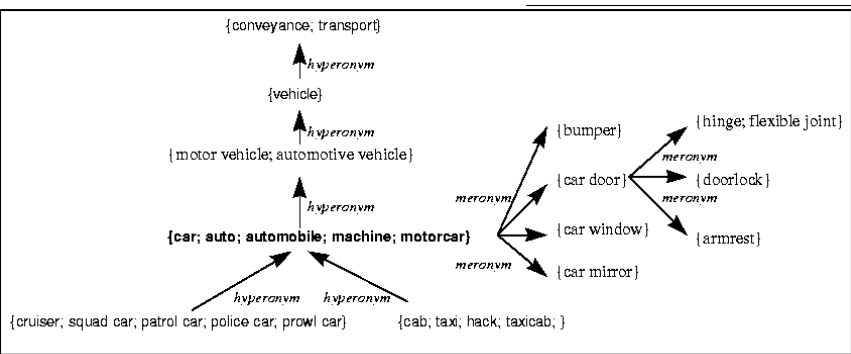


Figure 1: Synsets related to "car" in its first sense in WordNet1.5.

EuroWordNet: Polylinguales Wortnetz

- Kernkomponente: Gemeinsamer *Interlingualer Index (ILI)*, orientiert an WordNet-Synsets
- **Top-Ontologie** mit 63 semantischen Merkmalen
- **Domänen-Ontologien** für semantische Felder

Einzel-sprachliche Wortnetze: Gemeinsame Menge von **Base Concepts** (1000 Nomen, 300 Verben) mit folgenden Eigenschaften:

- Sie dominieren viele Knoten oder haben vielstufige Kette von Unterkonzepten;
- Sie treten häufig in (mindestens) zwei Sprachen auf;
- Sie sind spezifischer als die Konzepte der Top-Ontologie (z.B. *Dynamic, Function, Property*), i.a. Ober-Konzepte der Basic-Level Concepts, z.B. *Werkzeug für Hammer*;
- Sie werden durch Merkmale oder Merkmalskombinationen aus der Top-Ontologie charakterisiert, z.B. *Werkzeug* durch *Artefact, Instrument, Object*.

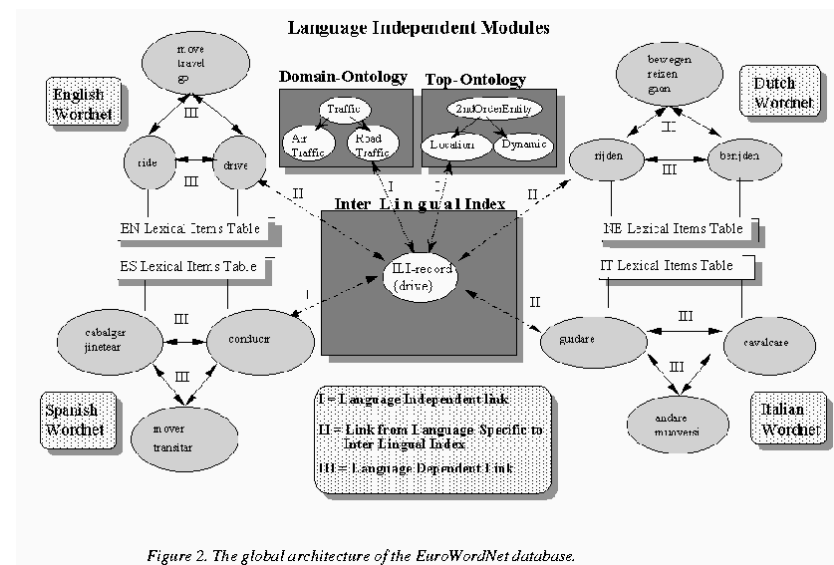


Figure 2: The global architecture of the EuroWordNet database.

EuroWordNet

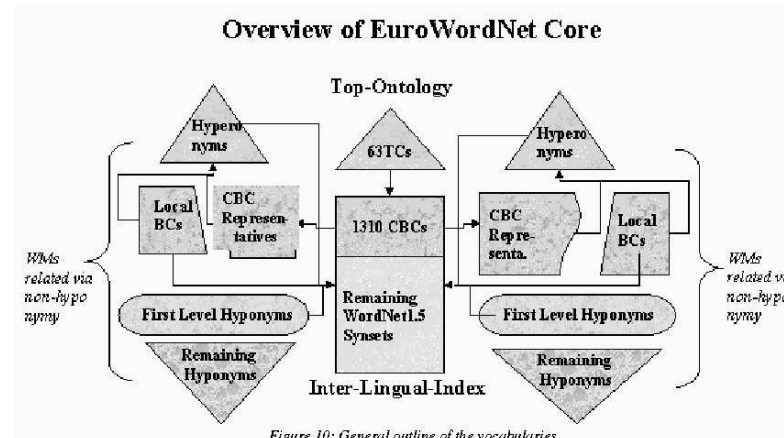


Figure 10: General outline of the vocabularies

Top ⁰	
1stOrderEntity ¹	2ndOrderEntity ²
Origin³ Natural ¹¹ Living ²⁰ Plant ¹⁵ Human ¹⁰⁶ Creature ² Anima ²³	SituationType⁶ Dynamic ¹³⁴ BoundedEvent ⁵³ UnboundedEvent ²⁵ Static ²⁵ Property ⁶¹ Relation ²⁵
Form⁹ Artifact ¹⁴⁴ Substance ²⁷ Solid ⁶³ Liquid ¹³ Gas ¹	SituationComponent⁵ Cause ⁶⁷ Agentive ¹⁷⁰ Phenomena ¹⁷ Stimulating ²⁵
Composition⁹ Object ⁶² Part ²⁶ Group ⁶³	Communication ⁵⁰ Condition ⁶² Existence ²⁷ Experience ⁴³ Location ⁷⁶ Manner ²¹ Mental ¹⁶ Modal ¹⁴⁰ Physical ¹⁴⁰ Possession ²³ Purpose ³⁹ Quantity ¹⁷ Social ¹⁰² Time ²⁴ Usage ⁵
Function³ Vehicle ⁵ Representation ¹² MoneyRepresentation ¹⁰ LanguageRepresentation ³⁴ ImageRepresentation ⁷	
Software ² Place ²⁵ Occupation ³³ Instrument ²⁸ Garment ¹ Furniture ⁶ Covering ⁵ Container ¹² Comestible ²¹ Building ¹³	

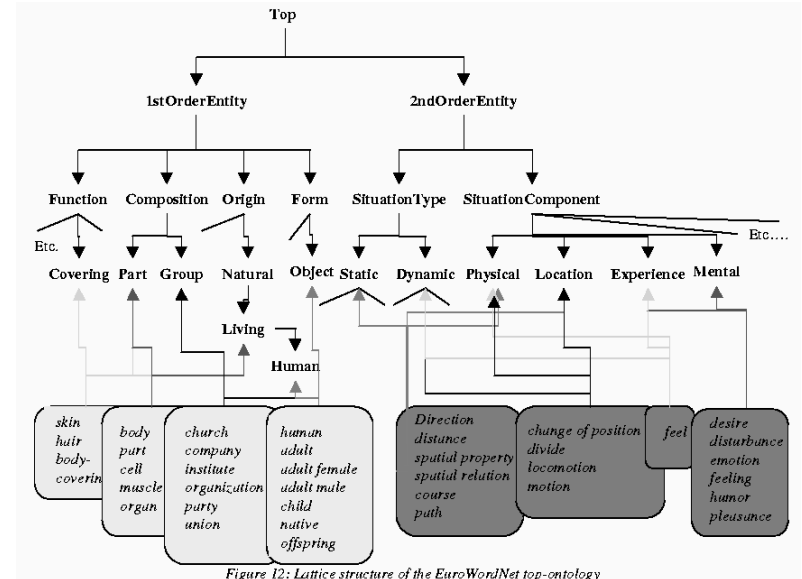


Figure 12: Lattice structure of the EuroWordNet top-ontology

Table 2: Language Internal Relations between synsets in EuroWordNet

Relation Type	Parts of Speech	Labels	Dir. / Syn.
NEAR_SYNONYM	N<N, V>V		Syn <>Syn
XPRES_NEAR_SYNONYM	N<V, N>AdjAdv, V<AdjAdv		Syn <>Syn
HAS_HYPERONYM	N<N, V>V	dis, con	Syn <>Syn
HAS_HYPONYM	N<N, V>V	dis	Syn <>Syn
HAS_XPES_HYPERONYM	N<V, N>AdjAdv, V>AdjAdv, V>N, AdjAdv>N, AdjAdv>V	dis, con	Syn <>Syn
HAS_XPES_HYPONYM	N<V, N>AdjAdv, V>AdjAdv, V>N, AdjAdv>N, AdjAdv>V	dis	Syn <>Syn
HAS_HOLONYM	N<N	dis, con, rev, neg	Syn <>Syn
HAS_HOLO_PART	N<N	dis, con, rev, neg	Syn <>Syn
HAS_HOLO_MEMBER	N<N	dis, con, rev, neg	Syn <>Syn
HAS_HOLO_PORTION	N<N	dis, con, rev, neg	Syn <>Syn
HAS_HOLO_MADEOF	N<N	dis, con, rev, neg	Syn <>Syn
HAS_HOLO_LOCATION	N<N	dis, con, rev, neg	Syn <>Syn
HAS_MERONYM	N<N	dis, con, rev, neg	Syn <>Syn
HAS_MERGO_PART	N<N	dis, con, rev, neg	Syn <>Syn
HAS_MERGO_MEMBER	N<N	dis, con, rev, neg	Syn <>Syn
HAS_MERGO_MADEOF	N<N	dis, con, rev, neg	Syn <>Syn
HAS_MERGO_LOCATION	N<N	dis, con, rev, neg	Syn <>Syn
ANTONYM	N<N, V>V		Syn <>Syn
NEAR_ANTONYM	N<N, V>V		Syn <>Syn
XPRES_NEAR_ANTONYM	N<V, N>AdjAdv, V<AdjAdv		Syn <>Syn
CAUSES	V>V, N<V, N>N, V>N, V>AdjAdv, N>AdjAdv	dis, con, non-f, rev, neg	Syn <>Syn
IS CAUSED BY	V>V, N<V, N>N, V>N, AdjAdv>V, AdjAdv>N	dis, con, non-f, rev, neg	Syn <>Syn
HAS_SUBEVENT	V>V, N<V, N>N, V>N	dis, con, rev, neg	Syn <>Syn
IS SUBEVENT OF	V>V, N<V, N>N, V>N	dis, con, rev, neg	Syn <>Syn
ROLE	N<V, N>N, AdjAdv>N, AdjAdv>V	dis, con, rev, neg	Syn <>Syn
ROLE_AGENT	N<V, N>N	dis, con, rev, neg	Syn <>Syn
ROLE_INSTRUMENT	N<V, N>N	dis, con, rev, neg	Syn <>Syn
ROLE_PATIENT	N<V, N>N	dis, con, rev, neg	Syn <>Syn
ROLE_LOCATION	N<V, N>N, AdjAdv>N, AdjAdv>V	dis, con, rev, neg	Syn <>Syn
ROLE_DIRECTION	N<V, N>N, AdjAdv>N, AdjAdv>V	dis, con, rev, neg	Syn <>Syn
ROLE_SOURCE_DIRECTION	N<V, N>N, AdjAdv>N, AdjAdv>V	dis, con, rev, neg	Syn <>Syn
ROLE_TARGET_DIRECTION	N<V, N>N, AdjAdv>N, AdjAdv>V	dis, con, rev, neg	Syn <>Syn
ROLE_RESULT	N<V, N>N	dis, con, rev, neg	Syn <>Syn
ROLE_MANNER	AdjAdv>N, AdjAdv>V	dis, con, rev, neg	Syn <>Syn
INVOLVED	V>N, N<N, V>AdjAdv, N>AdjAdv	dis, con, rev, neg	Syn <>Syn
INVOLVED_AGENT	V>N, N<N	dis, con, rev, neg	Syn <>Syn
INVOLVED_PATIENT	V>N, N<N	dis, con, rev, neg	Syn <>Syn