

Lexikalische Semantik: elektronische Wörterbücher, Wortbedeutung

Katja Markert, einige Folien von Dan Jurafsky und Yannick Versley

Institut für Computerlinguistik
Uni Heidelberg
markert@cl.leeds.ac.uk

January 9, 2023

- Bis jetzt: Wörter und Wortfolgen
- Uns war egal, was diese Wörter bedeuten
- Jetzt: Übergang zur Wortbedeutung
- Danach: Distributionelle Semantik, Wortvektoren

- 1 Was ist Wortbedeutung?
- 2 Relationale Bedeutungstheorie und Wortfelder; semantische Relationen und WordNet

- 1 Was ist Wortbedeutung?
- 2 Relationale Bedeutungstheorie und Wortfelder; semantische Relationen und WordNet

Lexikalische Semantik

Lehre der Bedeutung von Wörtern (hier = Lexeme/Lemmata/Zitationsformen)

Lemma	Wortform
bank	bank, banks
bank	bank, banking, banks, banked
found	found, founded, finds, founding
find	find, finds, finding, found

Lexem/Lemma

Lexikonform eines Wortes. Abstrakte Einheit, die der Menge von Formen, die ein Wort haben kann, entspricht. Grundeinheit der Bedeutung.

Wortform

Die voll flektierte Oberflächenform eines Wortes in einem Text.

Was ist eine Wortbedeutung? Definitionsmöglichkeiten

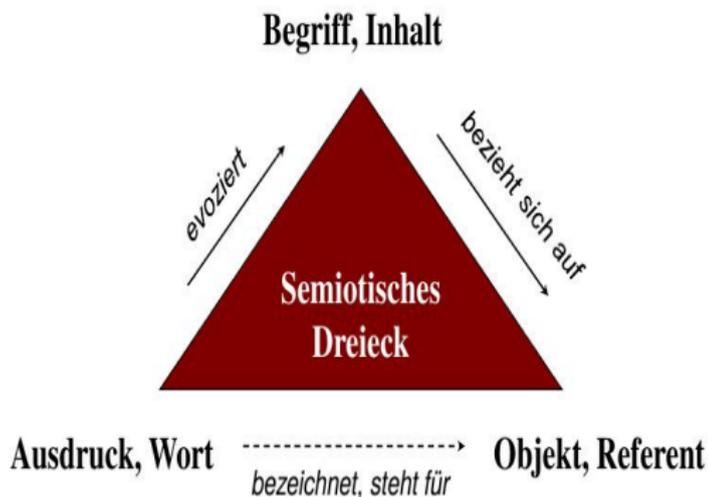
- **Extension:** Bedeutung von / = alle Objekte in der Welt, auf die / zutrifft
- **Denotation/Intension:** Beschreibung der Bedeutung von / via Eigenschaftsliste, oft Genus und Differentiae
- **Merkmalssemantik:** semantische Primitive (nicht mehr teilbare Meinungseinheiten)

	Geschlecht	Alter	Huhn?	Mensch
Henne	+w	+erw	+	-
Hahn	-w	+erw	+	-
Küken	?	-erw	+	-
<hr/>				
Frau	+w	+erw	-	+
Mann	-w	+erw	-	+
Kind	?	-erw	-	+
Säugling	?	-erw	-	+
Mädchen	+w	-erw	-	+

Ein relativ neuer Vorschlag für sehr abstrakte Primitiva: *Wierzbicka A. (1996)*

Semantics: Primes and Universals. Oxford University Press.





- Lexikalische Semantik: Beziehung zwischen Lexem und Begriff/Denotation
- Referentielle Semantik: Beziehung zwischen Wort (token) und Objekt

Wikipedia (abgerufen: 10. 12. 2022)

Bears are *carnivoran mammals of the family Ursidae*. They are classified as *caniforms, or doglike carnivorans*. Although only eight species of bears are extant, they are widespread, appearing in a wide variety of habitats throughout the Northern Hemisphere and partially in the Southern Hemisphere. Bears are found on the continents of North America, South America, Europe, and Asia. Common characteristics of modern bears include *large bodies with stocky legs, long snouts, small rounded ears, shaggy hair, plantigrade paws with five nonretractile claws, and short tails*.

- **Genus**
- Typische Eigenschaften, sogenannte **Differentiae**

Bear may also refer to (mehr als 40 Möglichkeiten)

- Bear, Arkansas, an unincorporated place
- KQBR, Texas, United States (ein Radiosender)
- Chicago Bears, an American football team
- Bear (2010 film), a US natural horror film
- The Bear (opera), by Jamie Smart

Enzyklopädien: oft referentielle Semantik

Collins English Dictionary (British English), abgerufen 10.12.2022

https:

[//www.collinsdictionary.com/dictionary/english/bear](https://www.collinsdictionary.com/dictionary/english/bear)

Noun uses in British English:

- 1 any plantigrade mammal of the family Ursidae: order Carnivora (carnivores). Bears are typically massive omnivorous animals with a large head, a long shaggy coat, and strong claws See also black bear, brown bear, polar bear. Related adjective: ursine
- 2 any of various bearlike animals, such as the koala and the ant bear
- 3 a clumsy, churlish, or ill-mannered person
- 4 a teddy bear
- 5 a speculator who sells in anticipation of falling prices to make a profit on repurchase

+ 17 Verbbedeutungen

WordNet <https://wordnet.princeton.edu>: Thesaurus mit Disambiguierung (synsets), glosses (denotationale Beschreibungen, oft mit Beispielen)

WordNet Search - 3.0 - [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Noun

- **S: (n) bear** (massive plantigrade carnivorous or omnivorous mammals with long shaggy coats and strong claws)
- **S: (n) bear** (an investor with a pessimistic market outlook; an investor who expects prices to fall and so sells now in order to buy later at a lower price)

- Enzyklopädien: Deutsche Wikipedia
- Lexika: www.duden.de
- Wordnet: Germanet und Germanet Rover
<https://uni-tuebingen.de/en/142806>
- Leipzig Wortschatz Projekt (Lexikon mit statistischen Angaben aus Korpora): <http://wortschatz.uni-leipzig.de/>

- Interne Bedeutungsstruktur des Wortes: Genus und Differentiae
- Mehrdeutigkeit/Ambiguität repräsentiert
 - Verschiedene Lexika machen verschiedene Bedeutungsunterscheidungen
 - Nicht möglich, alles Wissen über Bedeutung zu inkludieren: knowledge acquisition bottleneck
 - Definitionen oft zirkulär:

Zirkularität: American heritage dictionary

red (n): the colour of blood or ruby

blood (n): the red liquid that circulates ...

- Interne Bedeutungsstruktur des Wortes: Genus und Differentiae
- Mehrdeutigkeit/Ambiguität repräsentiert
- Verschiedene Lexika machen verschiedene Bedeutungsunterscheidungen
- Nicht möglich, alles Wissen über Bedeutung zu inkludieren: knowledge acquisition bottleneck
- Definitionen oft zirkulär:

Zirkularität: American heritage dictionary

red (n): the colour of blood or ruby

blood (n): the red liquid that circulates ...

Homonyme

Zwei Lexeme mit gleicher Orthografie und Aussprache , aber verschiedenen, unrelationierten Bedeutungen

Beispiele:

- Deutsch: *Steuer, Bank*
- Englisch: *bank, bat*

Homograph: gleiche Schreibweise (*bass/bass*)

Homophone: gleiche Aussprache *write/right, piece/peace*

Probleme durch Homonyme, Homographie und Homophone in NLP

- 1 Information retrieval: BEARS → Google (10.12.2022) : “The official Website of the Chicago Bears”.
- 2 Maschinelle Übersetzung: Bank (dt.) → bank, bench (engl.)
- 3 Sprachsynthese: bass vs. bass

Relationierte Bedeutungen des gleichen Lexems:

Bear:

- any plantigrade mammal of the family Ursidae: order Carnivora (carnivores). Bears are typically massive omnivorous animals with a large head, a long shaggy coat, and strong claws. See also black bear, brown bear, polar bear. Related adjective: , ursine
- a clumsy, churlish, or ill-mannered person

Bank:

- finanzielle Institution
- Gebäude der Institution

Systematische Polysemie

Muster verwandter Bedeutungen, gelten für mehrere Lexeme

- Institution vs. Gebäude vs. Zeitabschnitt: Schule, Uni, Bank
- Tier vs. Fleisch: Lamm, Rind
- Objekt für Besitzer: *der BMW drängelt sich natürlich vor*
- Tier für Mensch: Affe, Schwein, Fuchs

Wie finde ich Wortbedeutungen als Lexikograf?

- Introspektion
- Etymologie
- Andere Wörterbücher
- Korpusevidenz
- Tests:
 - Syntaktisch: *Ich esse Lamm* (ohne Artikel)
 - Zeugma: "Konjunktion" zweier Bedeutungen unmöglich
 - ① * Ich betrat die Bank, auf der ich im Park saß.
 - ② ? Ich betrat die Schule, die mich herausgeworfen hatte
- Ändern sich die Wahrheitsbedingungen?: *Affe*: Tier vs. dummer Mensch

- Schwierigkeiten bei Bedeutungsdefinition sowie Bedeutungsunterscheidungen
- Denotationale und referentielle Bedeutung; Intension und Extension
- Verschiedene Ressourcen haben komplementäres Wissen: Wikipedia, Lexika, Wordnet etc.
- Verschiedene Arten von Ambiguität

- 1 Was ist Wortbedeutung?
- 2 Relationale Bedeutungstheorie und Wortfelder; semantische Relationen und WordNet

Wortfeldtheorie

Gruppen verwandter Worte können in ein semantisches Feld (konzeptuelle Domäne) integriert werden. Lexikalische Bedeutung kann nicht für ein Wort isoliert betrachtet werden, sondern nur für Gruppen verwandter Lexeme.

Semantische Relationen

Relationen zwischen Wortbedeutungen

Beispiele:

- Synonymie
- Antonymie
- Hyperonymie, Hyponymie
- Meronymie
- Troponymie

Synonyme

Zwei Lexeme sind Synonyme, wenn sie füreinander substituierbar sind, in allen (?) Satzkontexten, ohne dass sich die Satzbedeutung (= Wahrheitsgehalt des Satzes) ändert.

- couch/sofa
- flat/apartment
- verscheiden/sterben/abkratzen
- Flüchtling/Asylant/Asylbewerber/Geflüchteter
- big/large → Relation zwischen senses
 - Big plane = large plane
 - Big sister = large sister?

Vollständige Synonyme sind sehr selten: Konnotationen, Genre, Sprachvariationen etc.

Principle of Linguistic Contrast

unterschiedliche Form → Unterschied in Bedeutung

Zudem: Synonymie eigentlich eine Relation zwischen word senses, nicht Wörtern

Konnotationen

Aspekte einer Wortbedeutung, die mit den Meinungen, Emotionen, Evaluationen des Schreibers/Sprechers oder Lesers/Hörers verbunden sind.

Osgood et al 1957: drei Hauptdimensionen:

- Valenz (= pleasantness)
- Arousal: intensity of emotion provoked
- dominance: degree of control/power

	Valence	Arousal	Dominance
courageous	8.05	5.5	7.38
music	7.67	5.57	6.5
heartbreak	2.45	5.65	3.58
cub	6.71	3.95	4.24
life	6.68	5.59	5.89

Antonyme

Lexeme/Senses mit gegensätzlicher Bedeutung

- Skala: lang/kurz
- binär/komplementär: lebendig/tot
- konvers: kaufen/verkaufen
- revers (Anfang eines Geschehens ist Endpunkt des anderen):
beladen/entladen

Oft sehr ähnlich im Verhalten im Kontext: kommen mit den gleichen Worten vor!

Hyponymie/Hyperonymie = Subordinates/Superordinates

Sense X ist Hyponym von Sense Y, wenn X spezifischer als Y ist. Y ist dann Hyperonym von X.

- Auto Hyponym Gefährt
- Mango Hyponym Obst

Auch genannt: X Is-a Y oder Y subsumiert X. Eigenschaften:

- **Extensionen** werden subsumiert. X Hyponym von Y. Dann gilt:
 $\forall x \in X \rightarrow x \in Y$
- **Implikationen:** Bei Hyponymie von X unter Y gilt, wer X ist, ist auch Y.
- **Transitivität:** X is-a Y und Y is-a Z \longrightarrow X is-a Z

Meronyme

X ist Meronym von Y, wenn X ein Teil von Y ist. Y ist dann Holonym von X.

- physischer Part: Rad — Auto
- Member: Baum —Wald
- Substanz: Papier — Buch

Transitiv (?).

WordNet ist ein Thesaurus + eine Datenbank lexikalisch-semantischer Relationen für das Englische (<http://wordnet.princeton.edu/>).
Manuell erstellt seit 1985.

- Nomina, Verben, Adjektive, Adverbien
- Wortbedeutungen sind in **Synsets** organisiert, mit **gloss**
- Synset: {chump1 , fool2, gull1 , mark9, patsy1, fall guy1, sucker1, soft touch1, mug2:} a person who is gullible and easy to take advantage of

Synset

Synset: Grundlegende Organisationseinheit in Wordnet. Eine Menge von Synonymen. {chump1, fool2, gull1, mark9 ... }

Gloss

Textuelle denotationale Beschreibung eines word sense. Eigenschaft eines synsets.

- Many-to-many mapping between words and synsets
- Synsets sind durch semantische Relationen verknüpft
- WordNets für andere Sprachen: GermaNet (deutsch), Cornetto (NL), WOLF (FR) , EuroWordNet:
<https://archive.illc.uva.nl//EuroWordNet/>
- Princeton WordNet wird nicht mehr erweitert: Open English WordNet (McCrae et al. 2020) wird erweitert

WordNet 3.0: 117.000 Synsets

Wortart	Anzahl W	Anzahl Synsets	Senses/Wort
Nomen	117 798	82,115	1.23/2.79
Verben	11 529	13 767	2.16/3.57
Adj	22 479	18 151	1.40/2.71
Adv	4 481	3621	1.25/2.71

Umfang GermaNet (April 2022): 159,514 Synsets mit 205 000
Lexemen

Es gibt 26 Supersenses (lexikographische Kategorien) für Nomen
und 15 für Verben.

Category	Example	Category	Example	Category	Example
ACT	<i>service</i>	GROUP	<i>place</i>	PLANT	<i>tree</i>
ANIMAL	<i>dog</i>	LOCATION	<i>area</i>	POSSESSION	<i>price</i>
ARTIFACT	<i>car</i>	MOTIVE	<i>reason</i>	PROCESS	<i>process</i>
ATTRIBUTE	<i>quality</i>	NATURAL EVENT	<i>experience</i>	QUANTITY	<i>amount</i>
BODY	<i>hair</i>	NATURAL OBJECT	<i>flower</i>	RELATION	<i>portion</i>
COGNITION	<i>way</i>	OTHER	<i>stuff</i>	SHAPE	<i>square</i>
COMMUNICATION	<i>review</i>	PERSON	<i>people</i>	STATE	<i>pain</i>
FEELING	<i>discomfort</i>	PHENOMENON	<i>result</i>	SUBSTANCE	<i>oil</i>
FOOD	<i>food</i>			TIME	<i>day</i>

Figure 18.2 Supersenses: 26 lexicographic categories for nouns in WordNet.

Bild aus Jurafsky und Martin, Ed 3, Figure 18.2

- ① S: (n) bear cub (a young bear)
- ② S: (n) brown bear, bruin, *Ursus arctos* (large ferocious bear of Eurasia)
 - ① S: (n) Syrian bear, *Ursus arctos syriacus* (yellowish-grey Syrian brown bear)
 - ② S: (n) grizzly, grizzly bear, silvertip, silver-tip, *Ursus horribilis*, *Ursus arctos horribilis* (powerful brownish-yellow bear of the uplands of western North America)
 - ③ S: (n) Alaskan brown bear, Kodiak bear, Kodiak, *Ursus middendorffi*, *Ursus arctos middendorffi* (brown bear of coastal Alaska and British Columbia)
- ③ S: (n) bruin (a conventional name for a bear used in tales following usage in the old epic 'Reynard the Fox')
- ④ ...

Nomenrelationen in WordNet

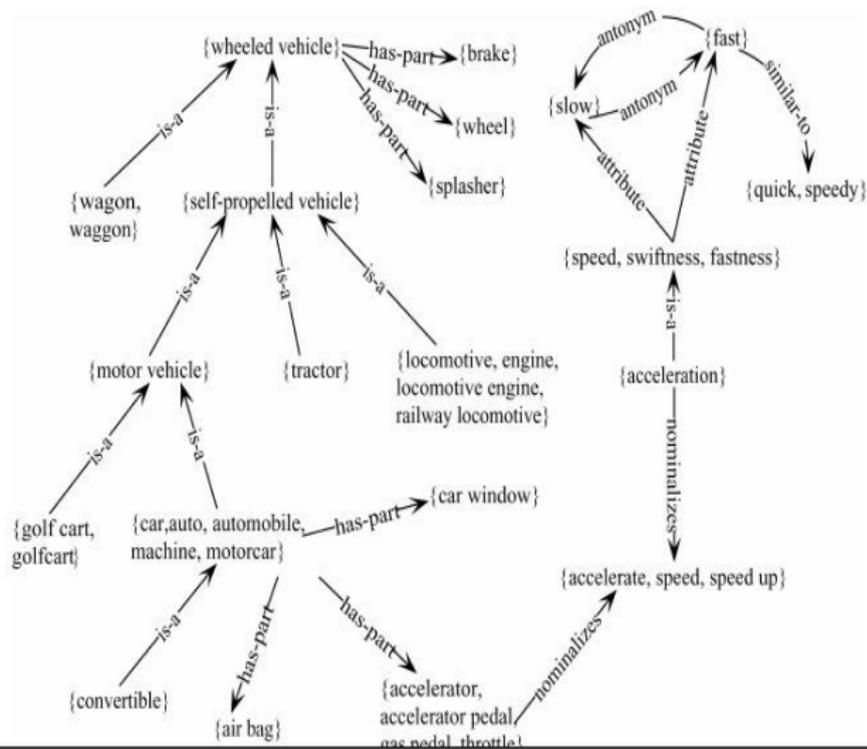
Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	<i>breakfast</i> ¹ → <i>meal</i> ¹
Hyponym	Subordinate	From concepts to subtypes	<i>meal</i> ¹ → <i>lunch</i> ¹
Instance Hypernym	Instance	From instances to their concepts	<i>Austen</i> ¹ → <i>author</i> ¹
Instance Hyponym	Has-Instance	From concepts to concept instances	<i>composer</i> ¹ → <i>Bach</i> ¹
Member Meronym	Has-Member	From groups to their members	<i>faculty</i> ² → <i>professor</i> ¹
Member Holonym	Member-Of	From members to their groups	<i>copilot</i> ¹ → <i>crew</i> ¹
Part Meronym	Has-Part	From wholes to parts	<i>table</i> ² → <i>leg</i> ³
Part Holonym	Part-Of	From parts to wholes	<i>course</i> ⁷ → <i>meal</i> ¹
Substance Meronym		From substances to their subparts	<i>water</i> ¹ → <i>oxygen</i> ¹
Substance Holonym		From parts of substances to wholes	<i>gin</i> ¹ → <i>martini</i> ¹
Antonym		Semantic opposition between lemmas	<i>leader</i> ¹ ⇔ <i>follower</i> ¹
Derivationally Related Form		Lemmas w/same morphological root	<i>destruction</i> ¹ ⇔ <i>destroy</i> ¹

Adaptiert von Jurasky und Martin, Figure 18.3

Relation	Definition	Example
Hypernym	From events to superordinate events	<i>fly</i> ⁹ → <i>travel</i> ⁵
Troponym	From events to subordinate event (often via specific manner)	<i>walk</i> ¹ → <i>stroll</i> ¹
Entails	From verbs (events) to the verbs (events) they entail	<i>snore</i> ¹ → <i>sleep</i> ¹
Antonym	Semantic opposition between lemmas	<i>increase</i> ¹ ⇔ <i>decrease</i> ¹
Derivationally Related Form	Lemmas with same morphological root	<i>destroy</i> ¹ ⇔ <i>destruction</i> ¹

Adaptiert von Jurafsky und Martin, Figure 18.4

WordNet als Graph (Bild aus Navigli (2016))



25

BabelNet

“Largest multilingual encyclopedic dictionary”. Verlinkung von heterogenen Wissensressourcen (Wikipedia, WikiData, WordNet etc) in vielen Sprachen.

- Behält Synsetorganisation mehrsprachig bei: {bear, Bär_{de}, urs_{fr} ... }
- Hat weiterhin semantische Relationen zwischen Synsets
- Aber auch konzeptuelle Relationen wie: Autor-von etc
- 51 Ressourcen in Dutzenden von Sprachen
- 1429 Millionen senses in 20.3 Millionen Synsets, oft mit Bildern

- Wortbedeutung kann oft als semantisches Feld am besten beschrieben werden.
- Wortbedeutungen sind durch semantische Relationen verknüpft.
- WordNet ist der bekannteste Versuch, semantische Felder im großen Stil manuell zu erstellen
- WordNet ist in Synsets organisiert, zwischen denen semantische Relationen bestehen.
- BabelNet vereint referentielle und denotationale Bedeutungsressourcen (WikiData, Wikipedia, WordNets etc) in verschiedenen Sprachen.

- * Jurafsky und Martin, 3rd online edition. Kapitel 6.1 sowie Kapitel 18.1 bis 18.3
- * Fellbaum, C.(2005). WordNet and wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670
- Lemnitzer.L. und Kunze (2007): Computerlexikographie. Kapitel 6.2. https://www.sfs.uni-tuebingen.de/GermaNet/documents/GN_Intro.pdf
- Osgood, W. et al (1957): The Measurement of Meaning. University of Illinois press.
- Clark, E. (1987). The principle of contrast: A constraint on language acquisition. MacWhinney, B.
- Wierzbicka A. (1996) Semantics: Primes and Universals. Oxford University Press.
- Mccrae, J, Rademaker, A., Rudnicka, E. und Bond, F. (2020): English WordNet 2020: Improving and extending a WordNet for English using an open-source methodology. In proc of MMWLREC 2020
- Navigli, R. (2016). Chapter 20. Ontologies. In R. Mitkov, editor, The Oxford handbook of computational linguistics. Oxford University Press. Navigli, R. und S. P. Ponzetto. 2012.
- Navigli, R., Bevilacqua, M., Conia, S., Montagnini, D., und Cecconi, F. (2021). Ten Years of BabelNet: A Survey. In IJCAI (pp. 4559-4567).

- **WordNet:** <https://wordnet.princeton.edu>
- **WordNet in NLTK inkorporiert, sonst auch Tools unter**
<http://wordnetweb.princeton.edu/perl/webwn>
- **GermaNet:** <https://uni-tuebingen.de/en/142806>
- **Französisches WordNet WOLF:** http://almanach.inria.fr/software_and_resources/WOLF-en.html
- **Niederländisches WordNet Cornetto:**
<https://portal.clarin.nl/node/4205>
- **Eurowordnet:**
<https://archive.illc.uva.nl//EuroWordNet/>
- **BabelNet:** <https://babelnet.org/>
- **English WordNet 2020:** <https://en-word.net/>
- **Leipzig Wortschatz:** wortschatz.uni-leipzig.de