

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

December 18, 2019

- Bis jetzt: Wortfolgen, Grammatik
- Jetzt: Übergang zur Bedeutung
- Jetzt: Wortbedeutung und lexikalische Relationen: Wörterbücher und Wortnetze
- Danach: Distributionelle Semantik, Wortvektoren

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

Abstrakte Einheit, die der Menge von Formen, die ein Wort haben kann, entspricht. Grundeinheit der Bedeutung. Grundeinheit des Lexikons.

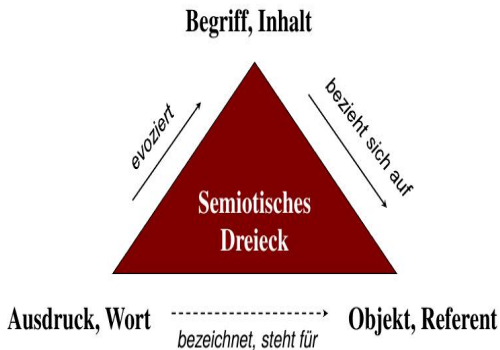
Was ist eine Wortbedeutung? Definitionsmöglichkeiten

- **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: 18.12.2016)

***Bears** are mammals of the family Ursidae. Bears are classified as caniforms, or doglike carnivorans, with the pinnipeds being their closest living relatives. 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.*

- Genus
- Typische Eigenschaften

Bear may also refer to (mehr als 40 Möglichkeiten)

- Bear, Arkansas
- KQBR, Texas, United States (ein Radiosender)
- Chicago Bears
- Bear (2010 film), a US natural horror film
- Bear (song), by indie rock band The Antlers

Enzyklopädien: oft referentielle Semantik

Collins English Dictionary:

- 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 stock exchange: a speculator who sells in anticipation of falling prices to make a profit on repurchase
- 6 a name for Russia

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
<http://www.sfs.uni-tuebingen.de/GermaNet/>
- Leipzig Wortschatz Projekt (Lexikon mit statistischen Angaben aus Korpora): <http://wortschatz.uni-leipzig.de/>

- Interne Bedeutungsstruktur des Wortes: Genus und Differentiae
- Mehrdeutigkeit (Wortbedeutung/word sense) 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 (Wortbedeutung/word sense) 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 (16.12.2016) : “The official Website of the Chicago Bears”
- 2 Maschinelle Übersetzung: Bank (dt.) → bank, bench (engl.)
- 3 Sprachsynthese: bass vs. bass

Relationierte Bedeutungen des gleichen Wortes:

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

- Schwierigkeiten bei Bedeutungsdefinition sowie Bedeutungsunterscheidungen
- Denotationale und referentielle Bedeutung; Intension und Extension
- Verschiedene Arten von Ambiguität

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, Konzeptrelationen

Relationen zwischen Wortbedeutungen

Beispiele:

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

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
- big/large → Relation zwischen senses
 - Big plane = large plane
 - Big sister = large sister?

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

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 mit gegensätzlicher Bedeutung

- Skala: lang/kurz
- binär/komplementär: lebendig/tot
- Konvers: kaufen/verkaufen
- revers: beladen/entladen

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

Hyponymie

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
- **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 dan Holonym von X.

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

Transitiv (?).

WordNet ist ein Thesaurus + Datenbasis 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**
 - 1 "chump(n)" hat gloss: a person who is gullible and easy to take advantage of
 - 2 Synset: chump1 , fool2, gull 1 , mark9, patsy 1, fall guy1, sucker1, soft touch1, mug2
- Synsets sind durch semantische Relationen verknüpft
- WordNets für andere Sprachen: GermaNet (deutsch), Cornetto (NL), WOLF (FR),
...<http://www.casta-net.jp/~kuribayashi/multi/>

WordNet 3.1: 117.000 Synsets

Wortart	Anzahl W	Anzahl Synsets	Senses/Wort
Nomen	117 798	82,115	1.24/2.79
Verben	11 259	13 767	2.17/3.57
Adj	21 479	18 151	1.40/2.71
Adv	4481	3621	1.25/2.71

Umfang GermaNet 8.0: 84584 Synsets

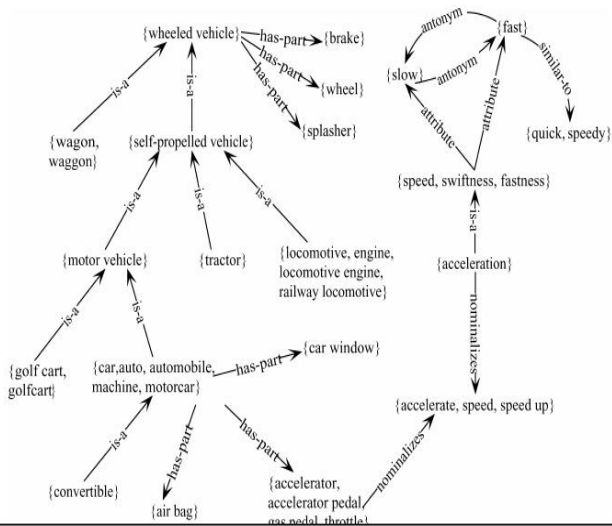
- ① S: (n) bear cub (a young bear)
- ② S: (n) brown bear, bruin, *Ursus arctos* (large ferocious bear of Eurasia)
 - ① S: (n) Syrian bear, *Ursus arctos syriacus* (yellowish-grey Syrian brown bear)
 - ② S: (n) grizzly, grizzly bear, silvertip, silver-tip, *Ursus horribilis*, *Ursus arctos horribilis* (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

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

Wordnet als Graph



25

- Wortbedeutung kann oft als semantisches Feld am besten beschrieben werden
- Dies ist durch semantische Relationen verknüpft
- WordNet ist der bekannteste Versuch, semantische Felder im großen Stil manuell zu erstellen
- Unterschiede Ontologie vs. Wordnet?

- * Jurafsky und Martin, 3rd online edition. Kapitel 6.1 sowie Kapitel 19.1 bis 19.3
- * Fellbaum, Christiane (2005). WordNet and wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670
- Osgood et al (1957): The Measurement of Meaning. University of Illinois press.
- Kapitel 6.3 in Lothar Lemnitzer and Claudia Kunze: Computerlexikographie. Gunter Narr Verlag 2007.

- **WordNet:** <https://wordnet.princeton.edu>
- **WordNet in NLTK inkorporiert, sonst auch Tools unter**
<http://wordnetweb.princeton.edu/perl/webwn>
- **Wordnet: Germanet**
<http://www.sfs.uni-tuebingen.de/GermaNet/>
- **Leipzig Wortschatz:** wortschatz.uni-leipzig.de
- **Sketch engine:** <https://www.sketchengine.co.uk> (benötigt
Registrierung fuer free trial)