

Einführung in die Computerlinguistik

Semantik

WS 2012/2013

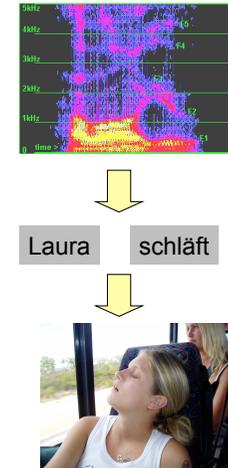
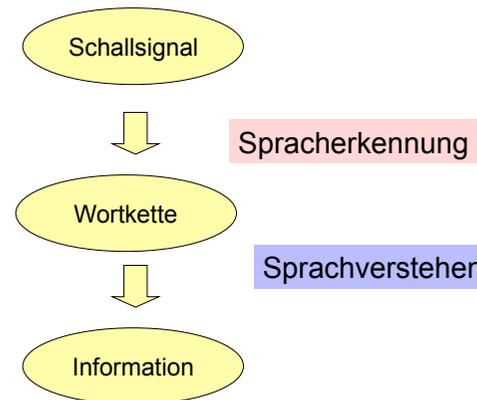
Manfred Pinkal

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Semantik

- Semantik ist der Teilbereich der (Computer-)Linguistik, der sich mit **sprachlicher Bedeutung** befasst.
- Semantik hat einen Doppelcharakter: Sie ist **Teil der Grammatik**, und sie ist **Schnittstelle zwischen Sprache und Welt**:
 - Wie berechnen wir, auf der Grundlage von Lexikon und Satzstruktur, die Bedeutungsrepräsentationen von Sätzen und Texten?
 - Wie kodieren, verarbeiten und kommunizieren wir Wissen (über die Welt) mit Sprache?

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Ein Beispiel

Sie beginnt um 10.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Ebenen der Bedeutung

- Linguistischer Bedeutungsgehalt, das "semantische Potenzial" einer Äußerung
 - Wie repräsentieren wir Wortbedeutung?
 - Wie repräsentieren wir Satzbedeutung?
 - Wie berechnen wir die Satzbedeutung aus der Wortbedeutung (und der syntaktischen Struktur)?
Bedeutungskomposition oder Semantikkonstruktion
- Intendierte Äußerungsbedeutung
 - Wie ermitteln wir die Äußerungsbedeutung auf der Grundlage von Kontextinformation?
Disambiguierung oder Ambiguitätsresolution
- Relevante Äußerungsinformation
 - Wie kommen wir von der Äußerungsbedeutung zur relevanten Äußerungsinformation?
Inferenz

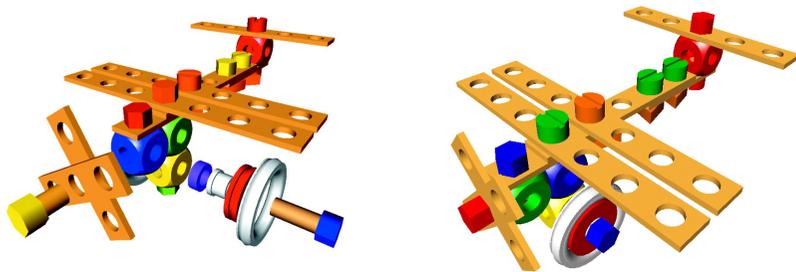
Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Was ist Wortbedeutung eigentlich?



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Eine Robotik-Anwendung



Sonderforschungsbereich
„Situerte Künstliche Kommunikatoren“
Bielefeld

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Stereotypische Information



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Sprachliche Bedeutung ist vielschichtig und heterogen

Zur Bedeutung können gehören:

- Propositionale/ konzeptuelle Information – begrifflich erfassbar, in einem logischen Framework darstellbar
- Visuelle (und andere sensorische) prototypische Information
- Handlungs-/ funktionsbezogene Information
- Stereotypische Information – nur im Regelfall gültig (Default-Information)

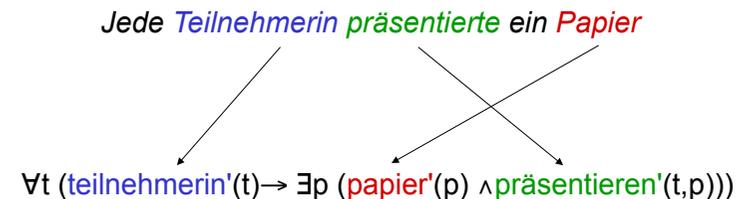
Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Wie repräsentieren wir Wort- und Satzbedeutungen?

- Der Standardformalismus für die Repräsentation satzsemantischer Information ist die **Prädikatenlogik**.
- Inhaltswörter werden als Konstanten repräsentiert.
- Beispiele:
 - *laura'*, *hans'*, *maria'* sind "Individuenkonstanten"
 - *schlafen'* ist einstelliges, *mögen'* zweistelliges Prädikat
- Sätze werden als einfache Prädikat-Argument-Strukturen oder als komplexe logische Formeln präsentiert
- Beispiele
 - Laura schläft* ⇒ *schlafen'(laura')*
 - Hans mag Maria* ⇒ *mögen'(hans', maria')*
 - Jeder Student arbeitet* ⇒ $\forall x(\text{student}'(x) \rightarrow \text{arbeiten}'(x))$
 - Jede Teilnehmerin präsentierte ein Papier* ⇒ $\forall t (\text{teilnehmerin}'(t) \rightarrow \exists p (\text{papier}'(p) \wedge \text{präsentieren}'(t,p)))$

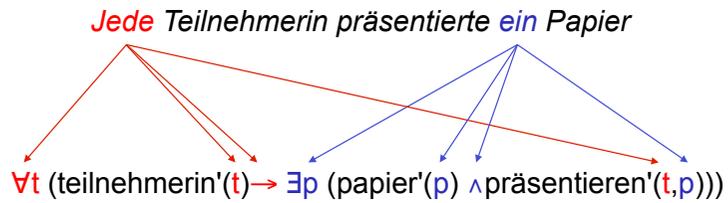
Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Semantikkonstruktion: Eine Herausforderung



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

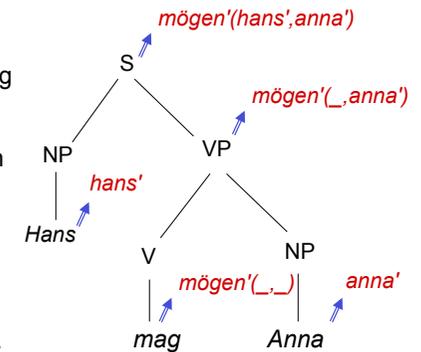
Semantikkonstruktion: Eine Herausforderung



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Semantik-Konstruktion: Der einfache Fall

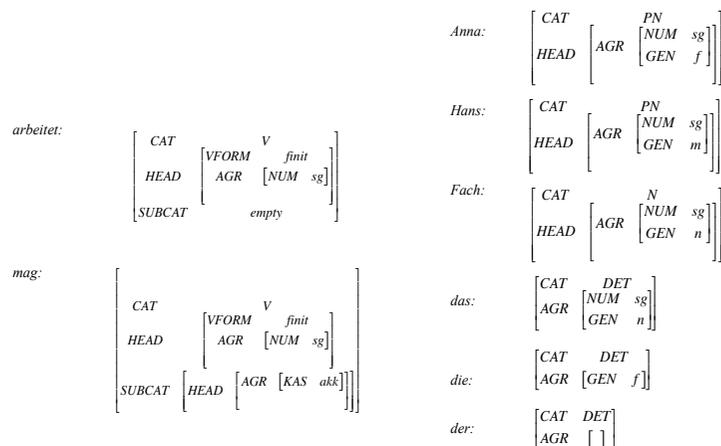
- Semantik-Konstruktion: Wir bauen an der Konstituentenstruktur entlang komplexe semantische Ausdrücke "kompositionell" aus einfacheren Ausdrücken zusammen.



- Wie wird Semantik-Konstruktion konkret durchgeführt? Zum Beispiel mit Unifikation.

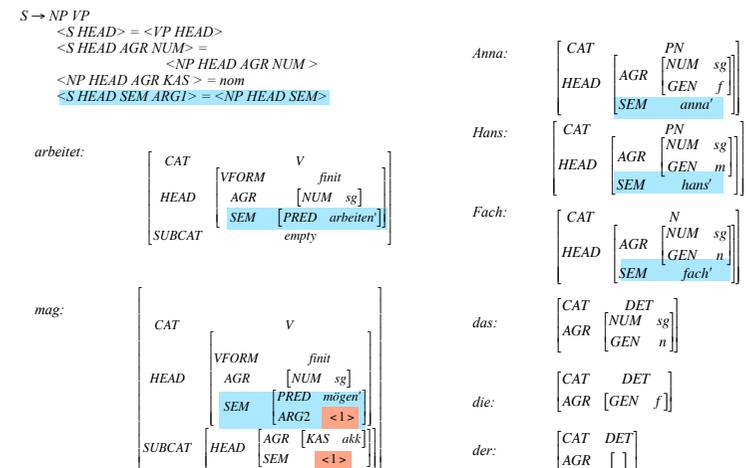
Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Semantikkonstruktion mit Merkmalsunifikation



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Semantikkonstruktion mit Merkmalsunifikation



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Ebenen der Bedeutung

- Linguistischer Bedeutungsgehalt, das "semantische Potenzial" einer Äußerung
 - Wie repräsentieren wir Wortbedeutung?
 - Wie repräsentieren wir Satzbedeutung?
 - Wie berechnen wir die Satzbedeutung aus der Wortbedeutung (und der syntaktischen Struktur)?
Bedeutungskomposition oder Semantikkonstruktion
- Intendierte Äußerungsbedeutung
 - Wie ermitteln wir die Äußerungsbedeutung auf der Grundlage von Kontextinformation?
Disambiguierung oder Ambiguitätsresolution
- Relevante Äußerungsinformation
 - Wie kommen wir von der Äußerungsbedeutung zur relevanten Äußerungsinformation?
Inferenz

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Inferenz

- Um Inferenzen auf sprachlichen Ausdrücken durchzuführen, benötigen wir nicht unbedingt die Repräsentation der vollen Wortbedeutung – nicht einmal die "propositionale" Ebene der Wortbedeutung.
- Hilfreich ist bereits die Kenntnis von semantischen **semantischen Relationen**, d.h., der systematischen Beziehungen, in denen Wörter zueinander stehen.
- Beispiel:
Besitzt Maria einen Wagen?
Maria besitzt ein Cabrio.
Cabrios sind Wagen.
Also besitzt Maria einen Wagen.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Wortbedeutung und Inferenz

- Wollen wir uns um 12 in der Stadt zum Essen treffen?
Meine Vorlesung fängt um 10 an.
- Mag Peter Lammkoteletts?
Peter ist Vegetarier.
- Besitzt Maria einen **Wagen**?
Maria hat sich kürzlich ein **Cabrio** gekauft.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Inferenz

- Das Standardwerkzeug für die Modellierung für Inferenzen ist die Prädikatenlogik.
- Beispiel:
Maria besitzt ein Cabrio.
Cabrios sind Wagen.
Also besitzt Maria einen Wagen.
$$\exists y(\text{cabrio}(y) \wedge \text{besitzen}(\text{maria}, y))$$
$$\forall x(\text{cabrio}(x) \rightarrow \text{wagen}(x))$$
$$\exists x(\text{wagen}(x) \wedge \text{besitzen}(\text{maria}, x))$$

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Semantische Relationen

- **Synonymie:** *Wagen – Auto*
- **Hyponymie/Hypernymie**, die Unter-/Oberbegriffsrelation:
 - *Luftfrachter – Flugzeug*
 - *Dackel – Hund*
 - *Cabrio – Wagen*
- **Antonymie:** Kontrastrelation
 - *teuer – billig*
 - *Freund – Feind*
 - *Säugetier – Fisch*
- **Meronymie/ Holonymie** die Teil-von-Relation:
 - *Ast – Baum*
 - *Holz – Baum*
 - *Baum – Wald*

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinal UoS Computerlinguistik

Mehrdeutigkeit

Lexikalische Mehrdeutigkeit:

Bank, Absatz, Baum

Referenzielle Mehrdeutigkeit:

*er, sie, es, dort, damals, der Präsident, die Vorlesung
Hans mag seinen Hund, obwohl er ihn manchmal beißt*

Strukturelle Mehrdeutigkeit:

*Peter sieht den Mann mit dem Teleskop (Anbindungsambiguität)
Zwei Fremdsprachen spricht jeder Linguist (Skopusambiguität)
1,2 Millionen Besucher tranken 800000 Tassen Kaffee
(distributive und kollektive Interpretation)*

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinal UoS Computerlinguistik

Abreißkalender (m. 3) Kalender mit abreißbaren Blättern
abreiten (V. 196) 1 (V. t.; hat) 1 zu Pferde überprüfen (Gelände, Platz) 2 ein Pferd für eine Prüfung – vorbereitend reiten II (V. i.; ist) wegreiten
Abri (m. 6) steinzeitliche Wohnstelle unter Felsvorsprüngen u. in Höhlen [Fz., Schutz, Unterschlupf]; zu lat. *apricare* „warm halten“
abrichten (V. t.; hat) 1 Tiere ~ dressieren 2 ein Geschütz ~ in Schussrichtung bringen 3 Bretter, Holz ~ durch Glätten in eine exakte Form bringen 4 jmdm. zu etwas ~ unterweisen, wie er sich zu verhalten hat
Abrichter (m. 3; eindeutschend) = Dresseur
Abriechung (f. 20; unz.) das Abrichten, Dressur, Zähmung
Abrieb (m. 3; unz.) 1 Materialabund durch Abreiben bei der Beförderung (z. B. von Kohle) od. durch Abnutzung bei Reibung (z. B. von Gummireifen) 2 das abgebrückelte od. abgeriebene Material [- reiben]
abrieftest (Adj.) widerstandsfähig gegen Abrieb
abriegeln (V. t.; hat) 1 zuriegeln 2 absperren 3 verbarrikadieren • Zugangswege ~ sperren; Truppen ~ von ihren Verbindungen trennen
Abriegelung (f. 20) oV Abriegelung 1 Absperrung 2 Verbarrikadierung
Abriegelung (f. 20) = Abriegelung
abriffeln (V. t.; hat) Samenkapeln ~ von den Leinwandeln trennen I [- Riffel]
abrinde (V. t.; hat) von der Rinde befreien (Baum, Brot)
abringen (V. t. 199; hat) jmdm. etwas ~ etwas von jmdm. gegen seinen anfängl. Widerstand erhalten • jmdm. ein Versprechen, eine Zusage ~; ich habe ihm sein Einverständnis abgerungen; ich habe mir diesen Brief abgerungen er ist mir sehr schwergefallen
Abriß (m. 1) 1 Abruch, etwas Zerfallenes 2 Entwurf, Skizze 3 Schema, kurze Darstellung, wissenschaftliche Übersicht 4 (schweiz.) (schamlose)

abrudern (V. i.; ist/hat) 1 (isruern 4 ate ruagern) durch eine letzte gemeinsame Fahrt ins Verein beschließen
Abruf (m. 3; unz.) das Abrufen • auf ~ sofort, wenn es gebraucht wird; (Kaufmannspr.) nach u. nach (Anweisung zur Teillieferung einer gekauften Warenmenge); etwas auf ~ bereithalten; sich auf ~ bereithalten; eine Existenz auf ~ eine gefährdete Existenz
abrufbar (Adj.) so beschaffen, dass es abgerufen werden kann • ~e Ware W., die jederzeit abgefragt werden kann; ~e Daten
abrufbereit (Adj.) zum Abrufen bereit
abrufen (V. t. 20; hat) 1 wegrufen, wegbegehren 2 (Eisenb.) zur Abfahrt ausrufen (Zug) 3 zur Teillieferung anweisen (Waren) • einen (hohen) Beamten ~ von seinem Posten entfernen; abgerufen werden (poet.) sterben
abführen (V. t.; hat) 1 Brei, Speisen ~ umrühren 2 etwas mit Eiern usw. ~ Eier usw. beimischen, einrühren
abzurunden (V. t.; hat) 1 rund machen 2 ichträgen 3 vervollkommen, ergänzen, kompletieren • eine Darstellung ~ (fig.) (zusammenfassend) ergänzen; eine Zahl ~ die Endziffer(n) durch Null(en) ersetzen; den Geschmack durch Zugabe von Zitronensaft ~; das Bild, der Eindruck ründet sich ab vervollständigt sich, wird zu einem Ganzen
Abzurundung (f. 20; unz.) 1 das Abrunden 2 abgerundete Form
abrupfen (V. t.; hat) abzupfen, abreißen
abrupt (Adj.) 1 abgebrochen, zusammenhanglos 2 plötzlich • die Stimmung schlag ~ um [-lat. abruptus „abgerissen“]
abrüstet (V. i.; hat) 1 (V. i.) Kriegsrüstung abschaffen, vermindern od. begrenzen; Sy demobilisieren II (V. t.) ein Gebäude ~ das Gerüst von ihm wegnehmen
Abrüstung (f. 20; unz.) das Abrüsten; Sy Demobilisierung, Demobilmachung

absaugen (V. t.; hat) 1 durch Absaugen wegnehmen, entfernen (Dampf, Flüssigkeit, Gas, Staub) 2 durch Saugen säubern (Teppich, Polster)
absaugen (V. t.; hat; veraltet) = abstillen
absausen (V. i.; ist; umg.) sausend davonfahren, davonrennen
abschaben (V. t.; hat) 1 durch Schaben entfernen, abkratzen (Schmutz) 2 durch Schaben säubern, glätten (Häute); → a. abgeschabt
abschalfen (V. t.; hat) aufbeben, besetzen (Un-

Die Wort-Bedeutungs-Relation

- Der Wortschatz natürlicher Sprachen ist extrem mehrdeutig. Ein Wort ist in der Regel mit verschiedenen Konzepten assoziiert :
Lexikalische Ambiguität
- Ambiguität zwischen nicht-verwandten Konzepten heißt **Homonymie**
 - *Bank, Absatz, versetzen*
- Ambiguität zwischen semantisch verwandten Konzepten heißt **Polysemie**
 - *Baum, beginnen, leicht*
- Genau genommen bestehen deshalb semantische Relationen nicht zwischen **Wörtern**, sondern zwischen **Konzepten** bzw. **Wortbedeutungen** ("word senses").
- Konzepte können durch unterschiedliche Wörter realisiert werden:
Synonymie

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinal UoS Computerlinguistik

WordNet

- WordNet ist eine große lexikalisch-semantische Ressource: Ein Netzwerk aus semantischen Relationen zwischen Konzepten, mit der Hyponymie-Relation als Kern.
- Konzepte werden als „synsets“ repräsentiert: Mengen von synonymen Wörtern, die sich gegenseitig disambiguieren.
- Außerdem enthält WordNet für alle Synsets „Glossen“ (Umschreibungen) und Anwendungsbeispiele.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Beispiel: Synset, Glosse, Sprachbeispiel

- **car**
 - { car, auto, automobile, machine, motorcar }
 - a motor vehicle with four wheels; usually propelled by an internal combustion engine
 - "he needs a car to get to work"

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Beispiel: Lesarten von car

- S: (n) **car**, [auto](#), [automobile](#), [machine](#), [motorcar](#)
- S: (n) **car**, [railcar](#), [railway car](#), [railroad car](#)
- S: (n) **car**, [gondola](#)
- S: (n) **car**, [elevator car](#)
- S: (n) [cable car](#), **car**

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

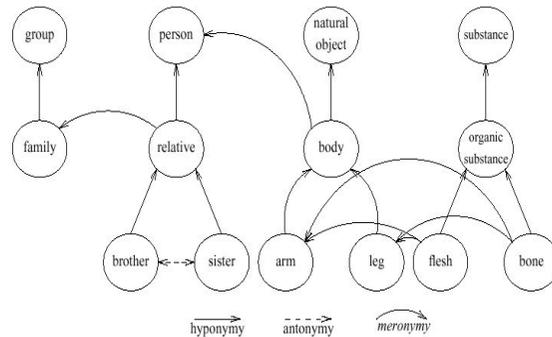
Hyponyme von motor vehicle

- S: (n) **motor vehicle**, [automotive vehicle](#) (a self-propelled wheeled vehicle that does not run on rails)
- *direct hyponym / full hyponym*
 - S: (n) [amphibian](#), [amphibious vehicle](#) (a flat-bottomed motor vehicle that can travel on land or water)
 - S: (n) [bloodmobile](#) (a motor vehicle equipped to collect blood donations)
 - S: (n) [car](#), [auto](#), [automobile](#), [machine](#), [motorcar](#) (a motor vehicle with four wheels; usually propelled by an internal combustion engine) "he needs a car to get to work"
 - S: (n) [doodlebug](#) (a small motor vehicle)
 - S: (n) [four-wheel drive](#), [4WD](#) (a motor vehicle with a four-wheel drive transmission system)
 - S: (n) [go-kart](#) (a small low motor vehicle with four wheels and an open framework; used for racing)
 - S: (n) [golfcart](#), [golf cart](#) (a small motor vehicle in which golfers can ride between shots)
 - S: (n) [hearse](#) (a vehicle for carrying a coffin to a church or a cemetery; formerly drawn by horses but now usually a motor vehicle)
 - S: (n) [motorcycle](#), [bike](#) (a motor vehicle with two wheels and a strong frame)
 - S: (n) [snowplow](#), [snowplough](#) (a vehicle used to push snow from roads)
 - S: (n) [truck](#), [motortruck](#) (an automotive vehicle suitable for hauling)

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UoS Computerlinguistik

Ein kleines Fragment von WordNet

Figure 2. Network representation of three semantic relations among an illustrative variety of lexical concepts



Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

WordNet-Daten

- Englisch WordNet hat zurzeit einen Umfang von
 - 170.000 lexikalischen Einträgen (Wörtern)
 - 120.000 Synsets
- WordNet-Versionen gibt es für etwa 45 Sprachen (für das Deutsche: "GermaNet")
- WordNet wird in vielen sprach- und informationstechnologischen Anwendungen erfolgreich genutzt, insbesondere als Grundlage von Inferenz
- Englisch WordNet
 - Web Interface: <http://wordnet.princeton.edu/perl/webwn>
 - Allgemeine Information: <http://wordnet.princeton.edu>

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

WordNet und Inferenz

- WordNet-Relationen können als logische Formeln reformuliert werden:

$$\forall x(\text{cabrio}(x) \rightarrow \text{wagen}(x))$$

$$\forall x(\text{delphin}(x) \rightarrow \text{säugetier}(x))$$

- Damit können sie unmittelbar für Inferenzprozesse nutzbar gemacht werden.

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik

Ebenen der Bedeutung

- Linguistischer Bedeutungsgehalt, das "semantische Potenzial" einer Äußerung
 - Wie repräsentieren wir Wortbedeutung?
 - Wie repräsentieren wir Satzbedeutung?
 - Wie berechnen wir die Satzbedeutung aus der Wortbedeutung (und der syntaktischen Struktur)?
Bedeutungskomposition oder Semantikkonstruktion
- Intendierte Äußerungsbedeutung
 - Wie ermitteln wir die Äußerungsbedeutung auf der Grundlage von Kontextinformation?
Disambiguierung oder Ambiguitätsresolution
- Relevante Äußerungsinformation
 - Wie kommen wir von der Äußerungsbedeutung zur relevanten Äußerungsinformation?
Inferenz

Vorlesung "Einführung in die CL" 2012/2013 © M. Pinkal UdS Computerlinguistik