

Einführung in die Computerlinguistik

Semantik

WS 2011/2012

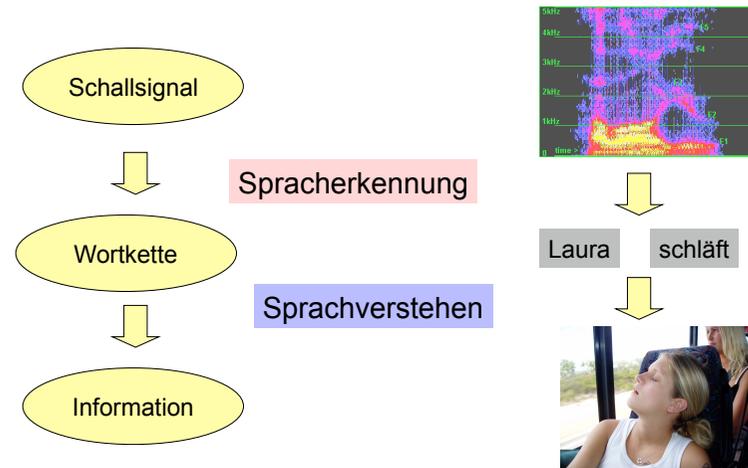
Manfred Pinkal

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS 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" 2011/2012 © M. Pinkal UdS Computerlinguistik



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Ein Beispiel

Sie beginnt um 10.

Vorlesung "Einführung in die CL" 2011/2012 © 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 Semantikonstruktion
- 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

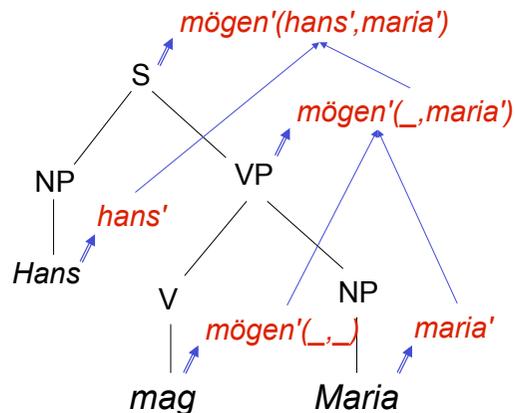
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Wie repräsentieren wir Wort- und Satzbedeutungen?

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

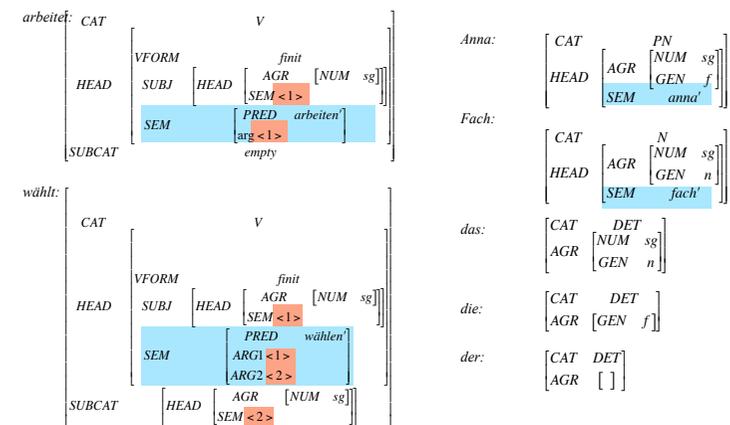
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Semantikonstruktion: Einfache Prädikat-Argument-Struktur



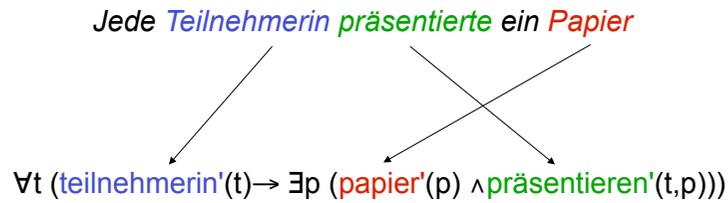
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Semantikonstruktion mit Merkmalsunifikation



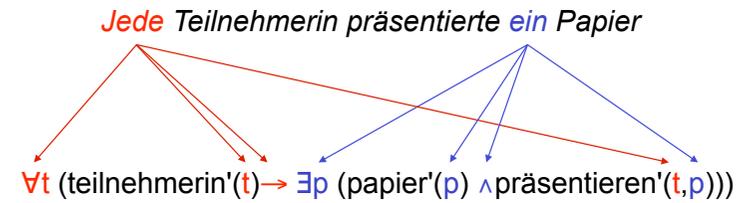
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Semantikkonstruktion: Eine Herausforderung



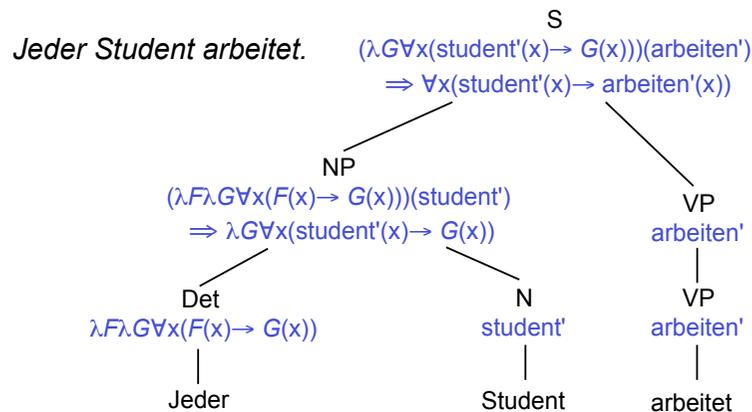
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Semantikkonstruktion: Eine Herausforderung



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Semantikkonstruktion: Typtheorie und Lambda-Kalkül



Vorlesung "Einführung in die CL" 2011/2012 © 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" 2011/2012 © M. Pinkal UdS Computerlinguistik

Was ist Wortbedeutung eigentlich?



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

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.
- Wieviele Luftfrachter verkauft Lockheed an das Vereinigte Königreich?
Der US-Flugzeughersteller Lockheed hat von Grossbritannien den Auftrag fuer 25 Flugzeuge des Typs Hercules C130J erhalten.

Vorlesung "Einführung in die CL" 2011/2012 © 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" 2011/2012 © M. Pinkal UdS Computerlinguistik

Inferenz

- Sind Delfine Fische?
Delfine sind Säugetiere, die ihr ganzes Leben im Meer verbringen ...
 $\forall x(\text{delfin}(x) \rightarrow \text{säugetier}(x)) \models \forall x(\text{delfin}(x) \rightarrow \neg \text{fisch}(x))$
 $\forall x(\text{delfin}(x) \rightarrow \text{säugetier}(x)), \forall x(\text{säugetier}(x) \rightarrow \neg \text{fisch}(x))$
 $\models \forall x(\text{delfin}(x) \rightarrow \neg \text{fisch}(x))$

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Inferenz

- Prädikatenlogik erlaubt uns – über den Folgerungsbegriff – die Modellierung von Inferenz.
- Heute stehen hoch effiziente "Inferenzmaschinen" in der Form von sogenannten **Theorembeweisern** zur Verfügung.
- Wenn wir Sätze in Prädikatenlogik formulieren, bekommen wir die Inferenz sozusagen geschenkt.
- Allerdings nicht ganz: Interessante **Inferenzen benötigen** zusätzlich zu den Regeln der Logik **Wissen** unterschiedlicher Art.
- Von zentraler Bedeutung für Inferenz im Sprachverstehen sind **semantische Relationen**.
- Wenn wir die semantischen Relationen kennen, in denen ein Wort zu anderen Wörtern der Sprache steht, kennen wir nicht "die Wortbedeutung", aber einige wichtige Aspekte der Wortbedeutung.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Semantische Relationen

- **Synonymie**: *Wagen – Auto*
- **Hyponymie/Hypernymie**, die Unter-/Oberbegriffsrelation:
 - *Luffrachter – 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" 2011/2012 © M. Pinkal 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" 2011/2012 © M. Pinkal UoS Computerlinguistik

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").

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal 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" 2011/2012 © M. Pinkal UdS 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" 2011/2012 © M. Pinkal UdS 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" 2011/2012 © M. Pinkal UdS Computerlinguistik

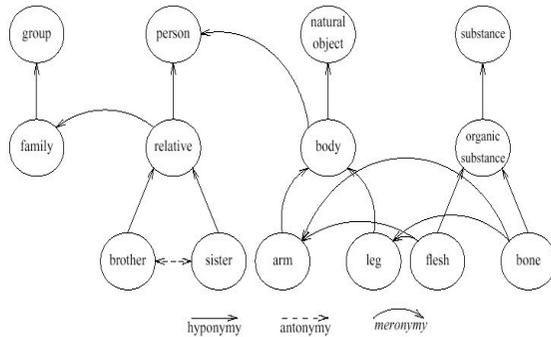
Hyponyme von motor vehicle

- 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 car](#) (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" 2011/2012 © M. Pinkal UdS 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" 2011/2012 © M. Pinkal UdS Computerlinguistik

WordNet-Daten

- Englisches 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
- Englisches WordNet
 - Web Interface: <http://wordnet.princeton.edu/perl/webwn>
 - Allgemeine Information: <http://wordnet.princeton.edu>

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

WordNet und Inferenz

- WordNet-Relationen können als logische Formeln reformuliert werden:
 - $\forall x(\text{family}(x) \rightarrow \text{group}(x))$
 - $\forall x(\text{person}(x) \rightarrow \exists y(\text{substance_m}(y,x) \wedge \text{body}(y)))$
 - $\forall x(\text{body}(x) \rightarrow \exists y(\text{part_m}(y,x) \wedge \text{leg}(y)))$
 - $\forall x(\text{body}(x) \rightarrow \exists y(\text{part_m}(y,x) \wedge \text{arm}(y)))$
- Für die semantische Verarbeitung in einem logischen Framework kann WordNet als große Datenbasis verwendet werden, die zusätzliche Axiome für die Inferenz bereitstellt.

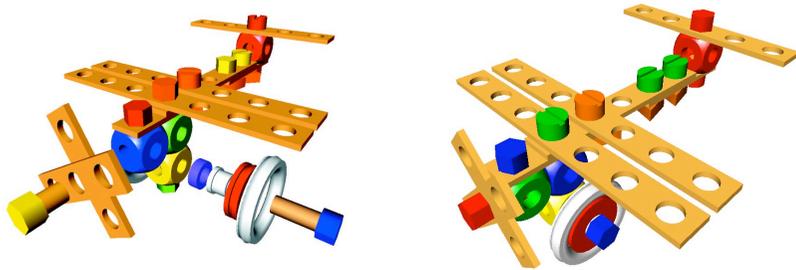
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Was ist Wortbedeutung nun eigentlich?



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Eine Robotik-Anwendung



Sonderforschungsbereich
„Situerte Künstliche Kommunikatoren“
Bielefeld

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Stereotypische Information



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Sprachliche Bedeutung ist vielschichtig und heterogen

- Zur Bedeutung können gehören:
 - **Propositionale/ konzeptuelle Information** – begrifflich erfassbar, in einem logischen Framework darstellbar
 - Visuelle (und andere perzeptuelle) **prototypische Information**
 - Handlungsbezogene Information
 - Stereotypische Information – nur im Regelfall gültig (Default-Information)
- Es gibt keine scharfe Grenze zwischen sprachlicher Bedeutung und nicht-sprachlichem Wissen

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik