

# Einführung in die Computerlinguistik

## Semantik

WS 2009/2010

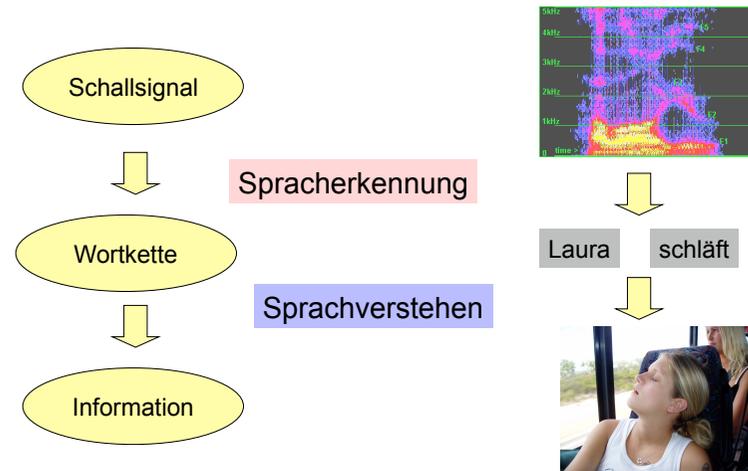
Manfred Pinkal

Vorlesung "Einführung in die CL" 2009/2010 © 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.

Vorlesung "Einführung in die CL" 2009/2010 © M. Pinkal UdS Computerlinguistik



Vorlesung "Einführung in die CL" 2009/2010 © M. Pinkal UdS Computerlinguistik

- Semantik ist Teil der Grammatik: Wie berechnen wir, auf der Grundlage von Lexikon und Satzstruktur, semantische Repräsentationen von Sätzen (bzw. erschließen sie mit statistischen Modellen)?
- Semantische Verarbeitung modelliert die Art und Weise, wie Information über die Welt kodiert und (in Texten und Dialogen) kommuniziert wird.

Vorlesung "Einführung in die CL" 2009/2010 © M. Pinkal UdS Computerlinguistik

## Ein Beispiel

*Sie beginnt um 10.*

5

## 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.
- *Natürliche Sprache ist extrem mehrdeutig und ambivalent.*

6

## Mehrdeutigkeit und Vagheit

Lexikalische Mehrdeutigkeit:

Homonymie (*Bank, Absatz, ergeben*)

Polysemie (*Absatz, beginnen, schnell*)

Vagheit

Referenzielle Mehrdeutigkeit:

*er, sie, es, dort, damals, der Präsident, die Vorlesung*

*Mehrdeutigkeit auf Satzebene:*

*Peter sieht den Mann mit dem Teleskop*

*Zwei Fremdsprachen spricht jeder Linguist*

*Zwei Teilnehmer halten ein Referat*

*1,2 Millionen Besucher tranken 800000 Tassen Kaffee*

7

## 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.
- *Natürliche Sprache ist extrem mehrdeutig und ambivalent.*
- *Die Bedeutung eines natürlich-sprachlichen Ausdrucks hängt massiv vom Kontext seiner Äußerung ab.*

8

## Kontextabhängigkeit

*Jeder Student kennt die Prädikatenlogik.*

*Überall grünt und blüht es.*

*Peter kommt immer zu spät.*

*Peter hat sich einen teuren Wagen gekauft.*

*Bitte noch eins!*

*Hans hat Peter nicht begrüßt. Er ist beleidigt.*

*Hans mag seinen Hund, obwohl er ihn manchmal beißt.*

## 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.
- Natürliche Sprache ist extrem mehrdeutig und ambivalent.
- Die Bedeutung eines natürlich-sprachlichen Ausdrucks hängt massiv vom Kontext seiner Äußerung ab.
- Semantik hört nicht auf der Satzebene auf: Wir unterscheiden Wortsemantik, Satzsemantik und Diskurssemantik
- **Semantik hört nicht mit der Ermittlung der kontextspezifischen Äußerungsinformation auf: Inferenz**

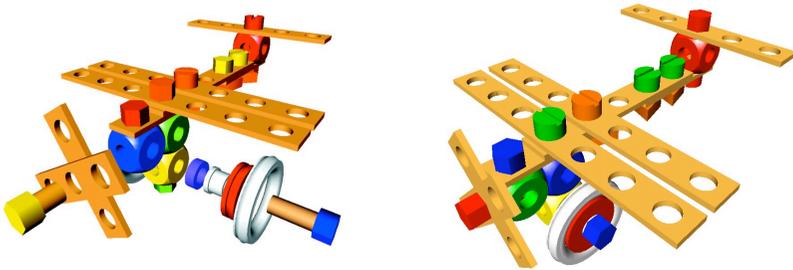
## Inferenz

- Semantisches Potenzial oder linguistischer Bedeutungsgehalt durch semantische Komposition oder Semantikonstruktion.
- Intendierte Äußerungsbedeutung durch Disambiguierung oder semantische Resolution.
- Relevante Äußerungsinformation durch Inferenz.

## Was ist Bedeutung eigentlich?



## Eine Robotik-Anwendung



Sonderforschungsbereich  
„Situerte Künstliche Kommunikatoren“  
Bielefeld



## 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.
- Natürliche Sprache ist extrem mehrdeutig und ambivalent.
- Die Bedeutung eines natürlich-sprachlichen Ausdrucks hängt massiv vom Kontext seiner Äußerung ab.
- Semantik hört nicht auf der Satzebene auf: Wir unterscheiden Wortsemantik, Satzsemantik und Diskurssemantik
- Semantik hört nicht mit der Ermittlung der kontextspezifischen Äußerungsinformation auf: Inferenz
- Der Gegenstand der Semantik ist abstrakt: Auf Wörter und Sätze können wir zeigen, aber was ist Bedeutung?
- Sprachliche Bedeutung ist vielschichtig und heterogen.

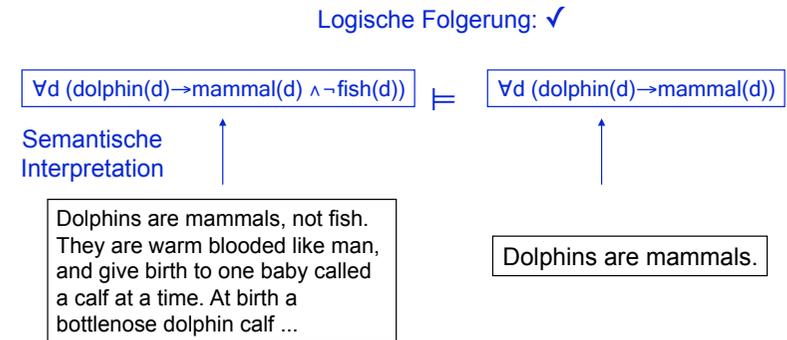
## Sprachliche Bedeutung ist vielschichtig und heterogen

- 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

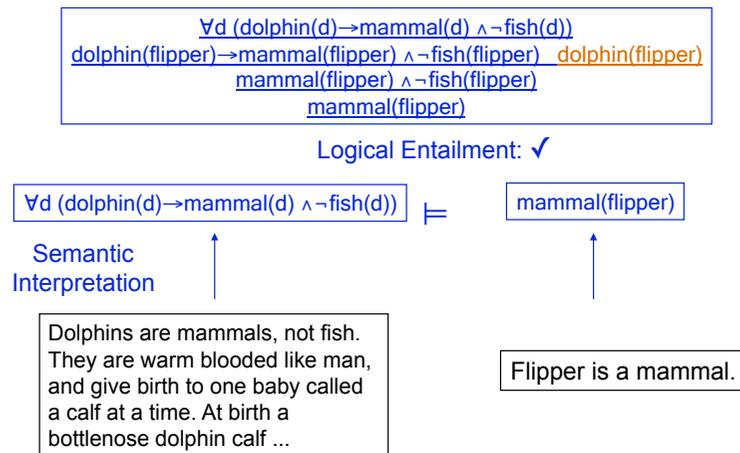
# Semantik in der Computerlinguistik

- Vollständige Beschreibung der sprachlichen Bedeutung ist nicht möglich.
- Bedeutungsbeschreibung ist notwendigerweise partiell - auf eine Schicht/ bestimmte Schichten semantischer Information eingeschränkt.
- Die Computerlinguistik hat sich bisher hauptsächlich mit propositionaler Bedeutung befasst, und hat sie mit Logik repräsentiert.

## Logik-basierter Ansatz



## Ist Flipper ein Säugetier?



## Lexikalische Semantik und Inferenz

*Einige **Autohersteller** haben sich zur wirtschaftlichen Entwicklung 2010 positiv geäußert.*

*Haben sich **Industriefirmen** zur wirtschaftlichen Entwicklung 2010 positiv geäußert?*

-

## Lexikalische Semantik und Inferenz

*Der US-Flugzeughersteller Lockheed hat von Grossbritannien den Auftrag fuer 25 **Flugzeuge** des Typs Hercules C130J erhalten. Vertreter des Verteidigungsministeriums bezifferten den Wert des Auftrags mit umgerechnet 2,5 Mrd. DM.*

- *Wieviel kosten die **Maschinen**, die Lockheed an Großbritannien verkauft hat?*

## Elementare semantische Relationen

- **Synonymie**: Flugzeug - Maschine - Flieger
- **Hypernymie / Hyponymie**, die Ober-/Unterbegriffsrelation:
  - *Flugzeug - Propellermaschine*
  - *Tier - Hund*
  - *töten - umbringen*
- **Meronymie/ Holonymie** die Teil-von-Relation:
  - Ast - Baum
  - Holz - Baum
  - Baum - Wald
- **Antonymie**: Kontrastrelation
  - gut - schlecht
  - teuer - billig
  - Freund - Feind

## Die Wort-Bedeutungs-Relation

- Genau genommen bestehen semantische Relationen nicht zwischen **Wörtern**, sondern zwischen **Konzepten** bzw. **Wortbedeutungen** ("word senses"): Die Abbildung zwischen phonologischen/ orthographischen Wörtern und Konzepten ist in beiden Richtungen nicht eindeutig.
- Ein Konzept kann in verschiedenen Wörtern kodiert sein: **Synonymie**
- Ein Wort ist in der Regel mit verschiedenen Wörtern assoziiert : **Lexikalische Ambiguität**
- Ambiguität zwischen nicht-verwandten Konzepten heißt **Homonymie**
  - *Bank: Geldinstitut /Bank: Sitzmöbel*
- Ambiguität zwischen semantisch verwandten Konzepten heißt **Polysemie**
  - *Maschine: Flugzeug / Maschine: Motorrad*
  - *Sitz im Auto / im Bundestag*
  - *einen Plan / einen Koffer aufgeben*

## 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.
- Synsets liefern direkte Information zur Synonymierelation und zur Wort-Konzept-Abbildung: Ein orthographisches Wort hat genau die Synsets als Lesarten, in denen es als Element vorkommt.
- Dies gilt im Prinzip. praktisch sind meist keine oder zu wenige Synonyme vorhanden, um Konzepte eindeutig zu charakterisieren. Deshalb enthält WordNet für alle Synsets „Glossen“ (Umschreibungen) und Anwendungsbeispiele.

## 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**

## Beispiel: Synset, Glosse, Sprachbeispiel

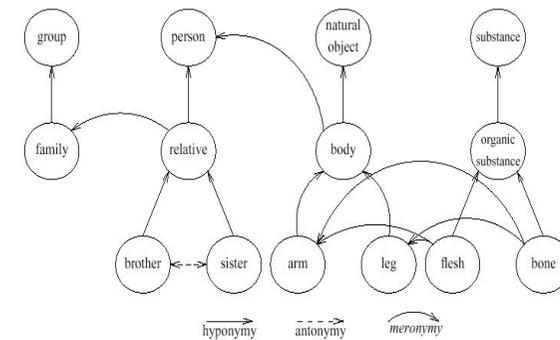
- **car**
  - { [car](#), [auto](#), [automobile](#), [machine](#), [motorcar](#) }
  - a motor vehicle with four wheels; usually propelled by an internal combustion engine
  - "he needs a car to get to work"

## 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)

## Ein kleines Fragment von WordNet

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



## WordNet-Daten

- Englisches WordNet hat zurzeit einen Umfang von ca. 200.000 Einträgen
- WordNet-Versionen gibt es für etwa 45 Sprachen
- WordNet wird in vielen sprach- und informationstechnologischen Anwendungen erfolgreich genutzt.
- "GermaNet": eine deutsche WordNet-Version mit etwa 100.000 lexikalischen Einträgen
- Englisches WordNet
  - Web Interface: <http://wordnet.princeton.edu/perl/webwn>
  - Allgemeine Information: <http://wordnet.princeton.edu>

## 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.