

# Einführung in die Computerlinguistik

Semantik

WS 2013/14

Manfred Pinkal

Vorlesung am 17.12.:  
Gebäude B 3.1, HSII (Raum 0.13)

(wie Einführung Allgemeine Sprachwissenschaft)

Schallsignal



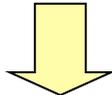
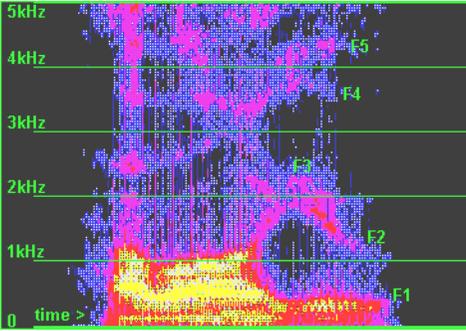
Wortkette



Information

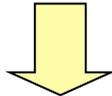
Spracherkennung

Sprachverstehen



Laura

schläft



# 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?

# Ein Beispiel

*Sie beginnt um 10.*

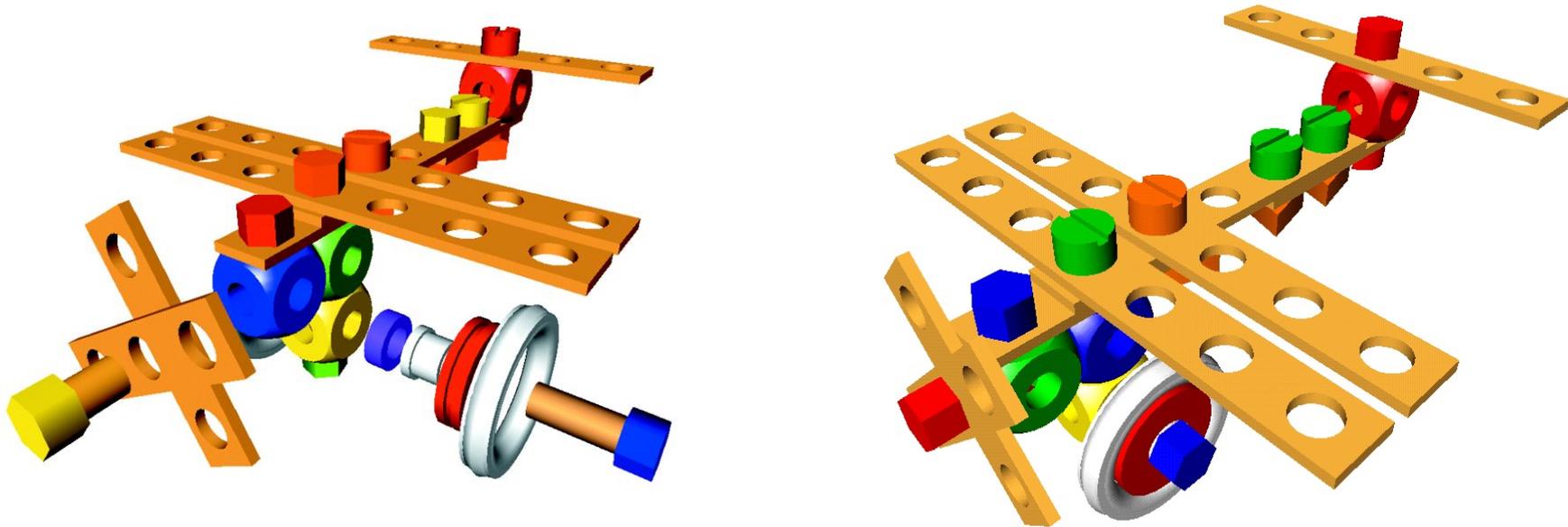
# Ebenen der Bedeutung

- 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

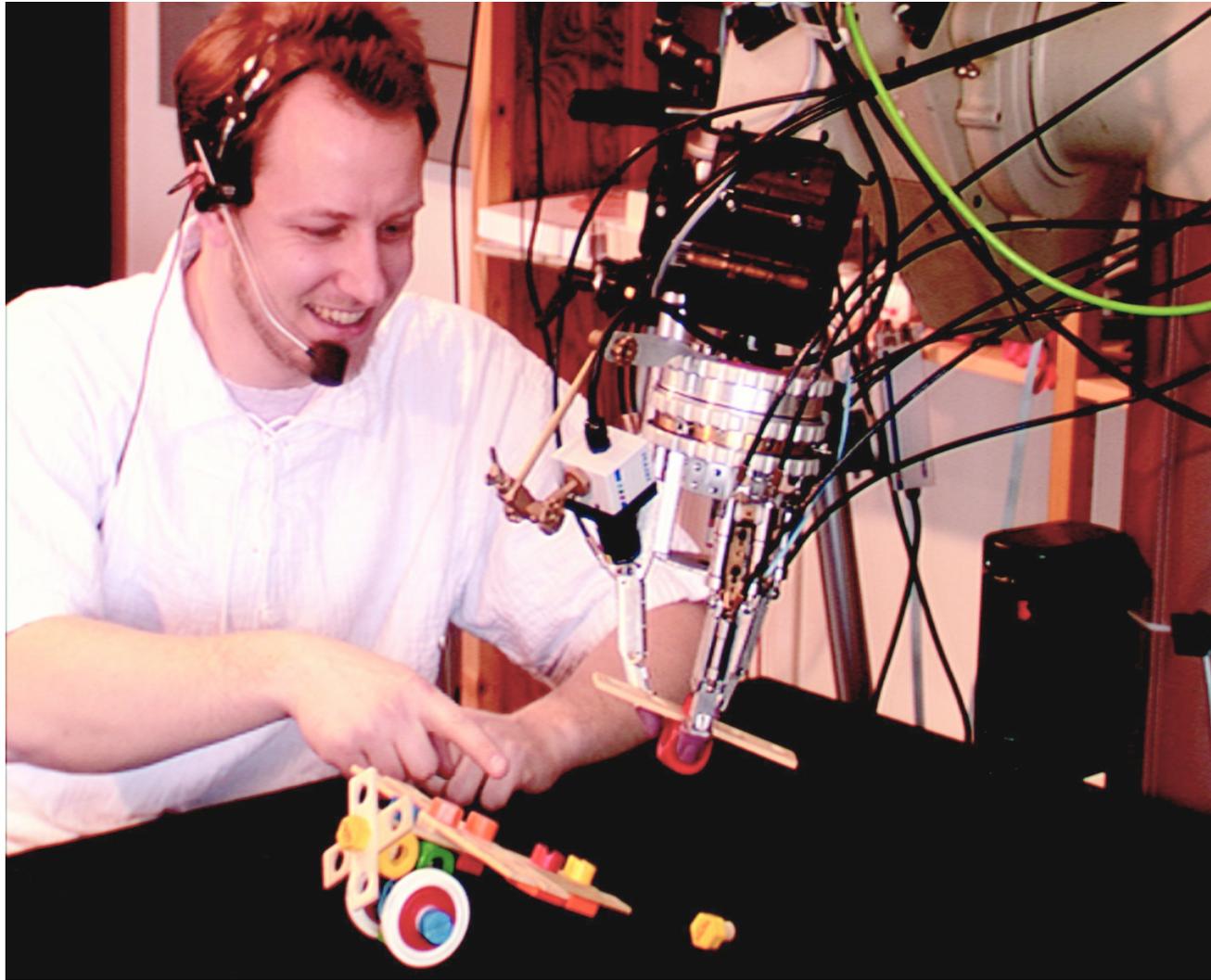
## Was ist Wortbedeutung eigentlich?



## Eine Robotik-Anwendung



Sonderforschungsbereich  
„Situerte Künstliche Kommunikatoren“  
Bielefeld



Vorlesung "Einführung in die CL" 2013/2014 © M. Pinkal UdS Computerlinguistik

# Stereotypische Information



Tisch



# Sprachliche Bedeutung ist vielschichtig und heterogen

Zur Bedeutung können gehören:

- **Propositionale/ konzeptuelle Information** – begrifflich erfassbare, z.B. definitorische Information
- Visuelle (und andere sensorische) **prototypische Information**
- Handlungs-/ funktionsbezogene Information
- Sterotypische Information – nur im Regelfall gültig (Default-Information)

# 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

## Wie repräsentieren wir Satzbedeutungen?

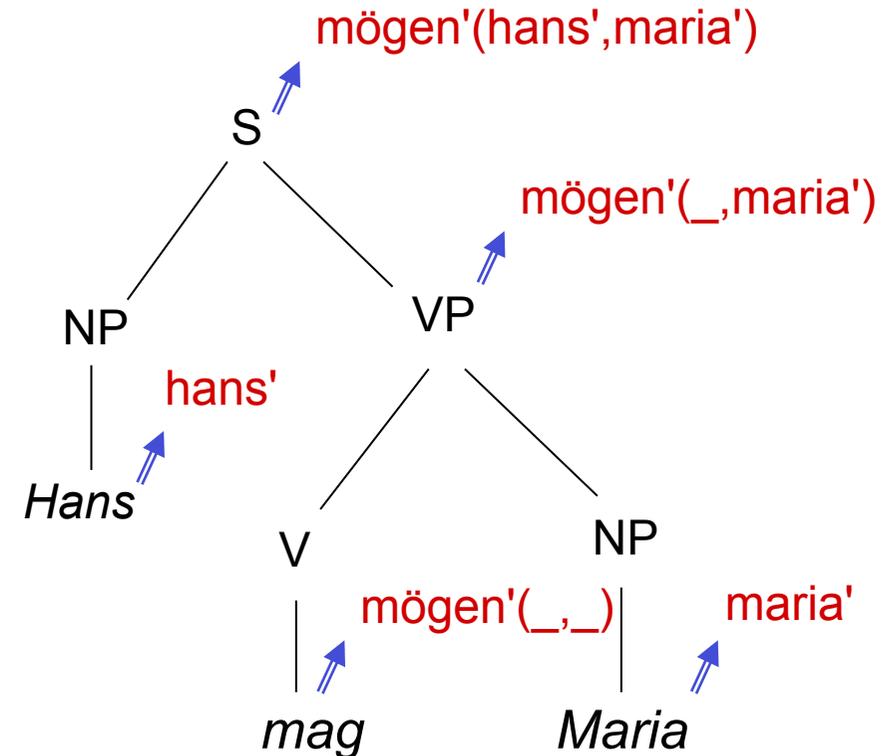
- Der Standardformalismus für die Repräsentation satzsemantischer Information ist die **Prädikatenlogik**.
- Inhaltswörter werden als Konstanten repräsentiert.
- Beispiele:
  - **laura'**, **hans'**, **maria'** sind "Individuenkonstanten"
  - **schlafen'** ist einstelliges, **mögen'** zweistelliges Prädikat
- Sätze werden als einfache Prädikat-Argument-Strukturen oder als komplexe logische Formeln präsentiert
- Beispiele
  - Laura schläft*  $\Rightarrow$  **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*  
 $\Rightarrow$   **$\forall t (\text{teilnehmerin}'(t) \rightarrow \exists p (\text{papier}'(p) \wedge \text{präsentieren}'(t,p)))$**

# 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

# Semantik-Konstruktion: Der einfache Fall

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



# Semantikkonstruktion: Eine Herausforderung

*Jede Teilnehmerin präsentierte ein Papier*

$\forall t (\text{teilnehmerin}'(t) \rightarrow \exists p (\text{papier}'(p) \wedge \text{präsentieren}'(t,p)))$

# Semantikkonstruktion: Eine Herausforderung

*Jede Teilnehmerin präsentierte ein Papier*

$\forall t (\text{teilnehmerin}'(t) \rightarrow \exists p (\text{papier}'(p) \wedge \text{präsentieren}'(t,p)))$

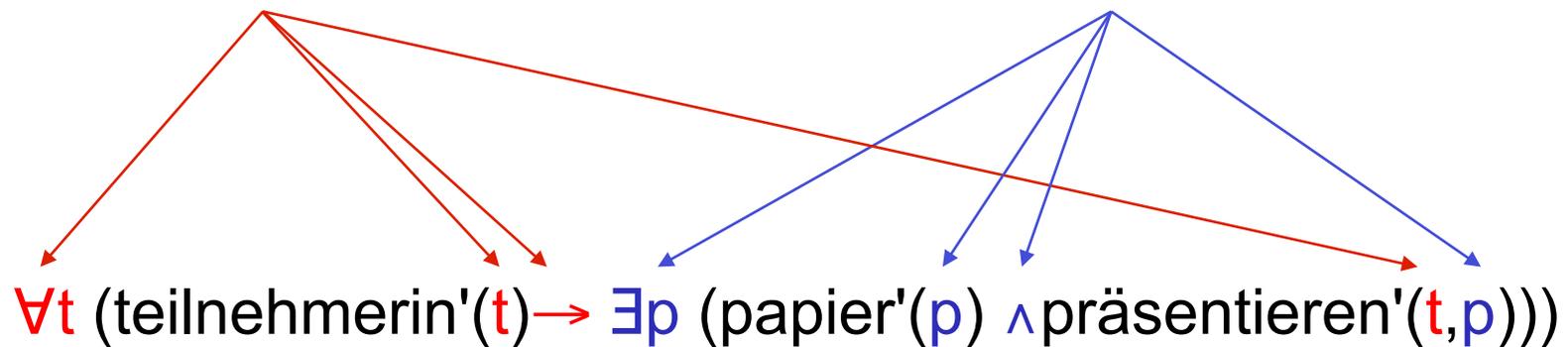
# Semantikkonstruktion: Eine Herausforderung

*Jede Teilnehmerin präsentierte ein Papier*

$\forall t (\text{teilnehmerin}'(t) \rightarrow \exists p (\text{papier}'(p) \wedge \text{präsentieren}'(t,p)))$

## Semantikkonstruktion: Eine Herausforderung

*Jede Teilnehmerin präsentierte ein Papier*



# 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**

# 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.

# Inferenz

- Um Inferenzen auf sprachlichen Ausdrücken durchzuführen, benötigen wir nicht unbedingt die Repräsentation der vollen Wortbedeutung – nicht einmal die volle "propositionale" Information
- 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.

# Inferenz

- Das Standardwerkzeug für die Modellierung für Inferenzen ist die Prädikatenlogik (Deduktion)
- Beispiel:  
Maria besitzt ein Cabrio.  
Cabrios sind Wagen.  
Also besitzt Maria einen Wagen.

$\exists y(\text{cabrio}(x) \wedge \text{besitzen}(\text{maria}, x))$

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

$\exists x(\text{wagen}(x) \wedge \text{besitzen}(\text{maria}, x))$

# 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*

# Die Wort-Bedeutungs-Relation

- Der Wortschatz natürlicher Sprachen ist extrem mehrdeutig. Ein Wort ist in der Regel mit verschiedenen Konzepten assoziiert :  
**Lexikalische Ambiguität**



# 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**

# 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.

# WordNet-Synsets, Beispiel

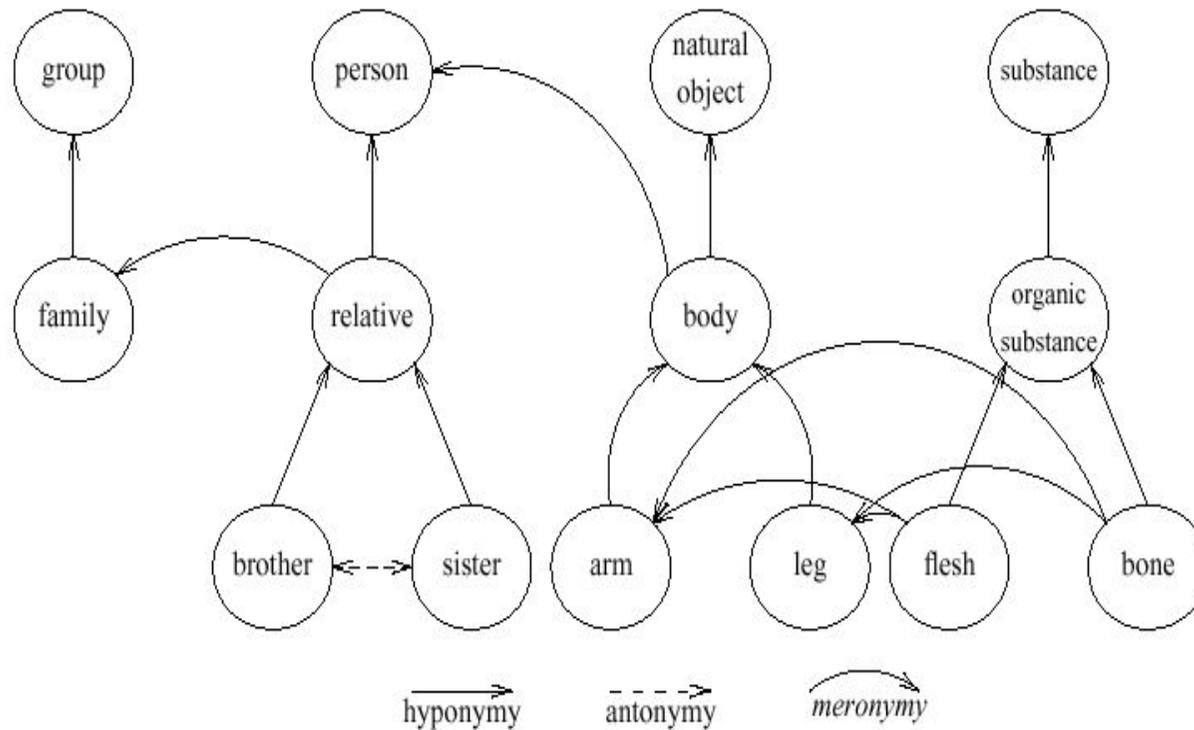
- **Absatz**

- { Absatz, Abschnitt, Paragraph }
- { Absatz, Treppenabsatz }
- { Absatz, Verkauf, Verkaufsziffern }
- ...
  
- <http://wordnetweb.princeton.edu/perl/webwn>

Englisches WordNet

## 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
  - 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/>
  - Allgemeine Information: <http://wordnet.princeton.edu>

## 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.

# 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**

# 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)

## Automatische Disambiguierung

*Sie beginnt um 10.*

*Ich gehe jetzt zur Bank, denn ich muss nachher einkaufen.*

*The box was in the pen – The pen was in the box*

- Disambiguierung durch Inferenz über Alltagswissen.
- Das Inferenzproblem ist grundsätzlich lösbar (Logik, Deduktion, Theorembeweiser)
- Das unlösbare Problem ist die Kodierung von Alltagswissen in ausreichendem Umfang und in der angemessenen Form. („Knowledge Bottleneck“)
- Alternative: Statistische Modellierung