

# Einführung in die Computerlinguistik

## Semantik

SS 2009

Sebastian Pado



## Semantik für Dialogverarbeitung

*U: Ist der Flug um 2 nach London pünktlich?*

*S: Ihr Flug nach London Heathrow geht pünktlich um 14:05 ab und steht zum Einsteigen bereit.*

*U: Und der nächste?*

*S: Der nächste Flug nach London Gatwick geht um 15.20.*

*U: Gibt es auch einen nach Heathrow?*



2

## Semantik für Dialogverarbeitung

U: Ist der Flug um 2 nach London pünktlich?

S: Ihr Flug nach London **Heathrow geht** pünktlich um **14:05 ab**  
**und steht zum Einsteigen bereit.**

U: Und der **nächste**?

S: Der nächste Flug nach London **Gatwick geht um 15.20.**

U: Gibt es auch einen nach Heathrow?

▶ 3

## Semantik für Maschinelle Übersetzung

D: Geht es um 2 bei Ihnen?

E1: Is 2 p.m. ok for you?

E2: Can we meet at 2 in your office?

▶ 4

## Semantik für den Informationszugriff

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

*Wieviel kosten die Flugzeuge, die Lockheed an Großbritannien verkauft?*

▶ 5

## Ebenen der Semantik

- ▶ **Wortsemantik:** Was ist Wortbedeutung? Wie wird Wortbedeutung repräsentiert? Wie wird Wortbedeutung im Lexikon organisiert?
- ▶ **Satzsemantik:** Wie wird Satzbedeutung repräsentiert? Wie ergibt sich die Satzbedeutung aus Wortbedeutung und syntaktischer Struktur?
- ▶ **Diskurssemantik:** Wie ergibt sich die spezifische Bedeutung einer Äußerung im Zusammenhang von Text und Dialog? Wie ergibt sich kohärente Textinformation aus der Verknüpfung von Satzbedeutungen?

▶ 6

## Lexikalische Semantik ...

---

... ist ein hoch komplexer Gegenstand:

- ▶ Vielschichtigkeit
- ▶ Mehrdeutigkeit
- ▶ Vagheit und Ambivalenz semantischer Information
- ▶ fließender Übergang zum Weltwissen

---

▶ 7

## Lexikalische Semantik ...

---

... ist äußerst aufwändig: Der Wortschatz ist

- ▶ sehr groß
- ▶ theoretisch und praktisch unbegrenzt
- ▶ variiert extrem nach Anwendungen  
(Fachwortschätze)

---

▶ 8

## Lexikalische Semantik ...

---

- ... ist ein hoch komplexer Gegenstand,
- ... ist äußerst aufwändig,
- ... ist bisher linguistisch und computerlinguistisch nur in bestimmten Ausschnitten erschlossen.

Ist lexikalische Semantik eine hoffnungslose Angelegenheit?

Nicht, wenn man seine Ziele bescheiden formuliert – und bescheidene Ziele können große praktische Wirkung haben.

---

▶ 9

## Lexikalische Semantik und Informationszugriff

---

*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?*

---

▶ 10

## Lexikalische Semantik und Informationszugriff

---

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

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

---

▶ 11

## Lexikalische Semantik und Informationszugriff

---

Elementare lexikalisch-semantische Relationen, z.B.

- Synonymie
- Hyponymie (Unterbegriffsrelation)
- Hyperonymie (Oberbegriffsrelation)

erschließen große Mengen zusätzlicher semantischer Information.

---

▶ 12

## WordNet

- ▶ eine große lexikalische Datenbank für das Englische (z.Zt. > 150.000 Wörter; dt.Version: GermaNet, > 100.000 Wörter)
- ▶ Hierarchische Strukturierung des Lexikons in Ober- und Unterbegriffe (Hyponyme/Hyperonyme)
- ▶ Einheiten von WordNet sind „synsets“: Mengen von synonymen Ausdrücken, die Lesarten von Wörtern charakterisieren...

▶ 13

## Beispiel: engl. *case*

- |  |                         |
|--|-------------------------|
| ▶ { <i>case, carton</i> }  | ■ Kasten/Karton         |
| ▶ { <i>case, bag, suitcase</i> }   | ■ Koffer                |
| ▶ { <i>case, pillowcase, slip</i> }  | ■ Kissenbezug           |
| ▶ { <i>case, cabinet, console</i> }  | ■ Schrank               |
| ▶ { <i>case, casing</i> (the enclosing frame around a door or window opening)} | ■ Rahmen                |
| ▶ { <i>case</i> (a small portable metal container)}                            | ■ Kasten/Etui/<br>Hülle |

▶ 14

## Weitere Relationen

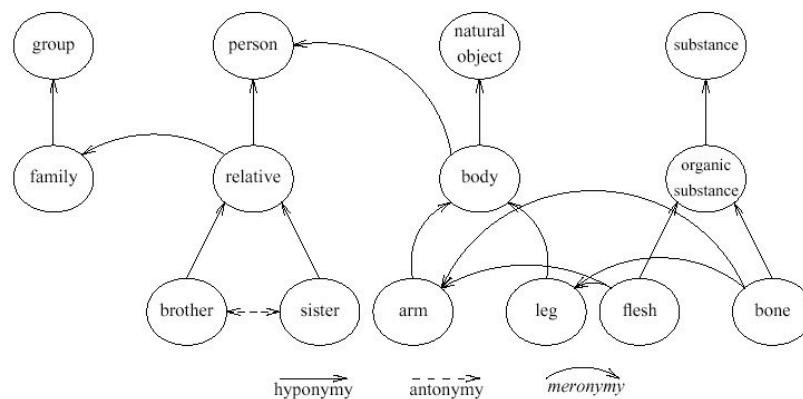
- ▶ **Meronymie**
  - ▶ Teil-Ganzes (*Ast - Baum*)
  - ▶ Gruppe - Element (*Baum - Wald*)
  - ▶ Stoff - Objekt (*Holz - Baum*)
- ▶ **Antonymie**
  - ▶ *Mann - Frau*
  - ▶ *Vater - Mutter*
- ▶ **Funktion**
  - ▶ *Messer - schneiden*
  - ▶ *Wolle – stricken*

Mit geringerer Abdeckung in WordNet kodiert.

▶ 15

## WordNet-Teilhierarchie, Beispiel

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



▶ 16

## WordNet

---

/

- ▶ WordNet-Versionen gibt es für etwa 30 Sprachen
- ▶ "GermaNet": eine deutsche WordNet-Version mit etwa 90.000 lexikalischen Einträgen
- ▶ Englisch WordNet
  - ▶ Web Interface: <http://wordnet.princeton.edu/perl/webwn>
  - ▶ Allgemeine Information: <http://wordnet.princeton.edu>

---

▶ 17



---

▶ 18

## Ebenen der Semantik

---

- ▶ **Wortsemantik:** Was ist Wortbedeutung? Wie wird Wortbedeutung repräsentiert? Wie wird Wortbedeutung im Lexikon organisiert?
- ▶ **Satzsemantik:** Wie wird Satzbedeutung repräsentiert? Wie ergibt sich die Satzbedeutung aus Wortbedeutung und syntaktischer Struktur?
- ▶ **Diskurssemantik:** Wie ergibt sich die spezifische Bedeutung einer Äußerung im Zusammenhang von Text und Dialog? Wie ergibt sich kohärente Textinformation aus der Verknüpfung von Satzbedeutungen?

---

▶ 19

## Prädikatenlogik als Repräsentationsformalismus

---

- ▶ Der Standardformalismus für die Repräsentation satzsemantischer Information ist die **Prädikatenlogik**.
- ▶ **Prädikat-Argument-Strukturen** als elementare Satzbausteine der logischen Repräsentation
- ▶ Kombination zu komplexen Repräsentationen durch **logische Operatoren** (Junktoren und Quantoren)

---

▶ 20

## Prädikat-Argument-Strukturen

---

- ▶ **Prädikat-Argument-Strukturen** als elementare Satzbausteine der logischen Repräsentation
  - ▶ arbeiten(peter)
  - ▶ kennen(hans, maria)
  - ▶ bruder-von(peter, maria)
  - ▶ geben(maria, peter, das-buch)

---

▶ 21

## Prädikatenlogik als Repräsentationsformalismus

---

- ▶ Der Standardformalismus für die Repräsentation satzsemantischer Information ist die **Prädikatenlogik**.
- ▶ **Prädikat-Argument-Strukturen** als elementare Satzbausteine der logischen Repräsentation
- ▶ Kombination zu komplexen Repräsentationen durch **logische Operatoren** (Junktoren und Quantoren)

---

▶ 22

## Von der Oberfläche zur Prädikat-Argument-Struktur

---

- ▶ *Maria mag den Film*
- ▶ *Der Film gefällt Maria*
  
- ▶ *Maria hat Peter das Buch gegeben*
- ▶ *Peter hat das Buch von Maria erhalten*
  
- ▶ *Peter hat Maria Blumen geschenkt*
- ▶ *Maria hat Peter mit Blumen beschenkt*

---

▶ 23

## Thematische Rollen

---

- ▶ **Thematische Rollen** beschreiben (Kategorien von) Mitspieler(n) einer Situation, unabhängig von ihrer grammatischen Realisierung
- ▶ **Beispiele**
  - ▶ Agens
  - ▶ Thema/ Patient/ Object
  - ▶ Rezipient
  - ▶ Instrument
  - ▶ Ausgangspunkt (source)
  - ▶ Ziel (goal)
  - ▶ Benefizient
  - ▶ Experiencer

---

▶ 24

## Satzanalysen

---

[Peter ]<sub>ag</sub> hat [Maria ]<sub>rec</sub> [das Buch ]<sub>pat</sub> gegeben

Agens = Peter

Rezipient = Maria

Patient = Buch

[Maria ]<sub>rec</sub> hat [das Buch ]<sub>pat</sub> [von Peter ]<sub>ag</sub> erhalten

Agens = Peter

Rezipient = Maria

Patient = Buch

---

▶ 25

## Syntax-Semantik-Schnittstelle

---

- ▶ Zuordnung von semantischen Argumentpositionen und syntaktischen Komplementen im Lexikon:

*geben*: Subj:        Agens  
          Obj (akk): Thema  
          Obj (dat): Rezipient

*erhalten*: Subj:        Rezipient  
              Obj (akk): Thema  
              PObj (von): Agens

---

▶ 26

## Thematische Rollen

- ✓ Erlauben eine abstraktere, bedeutungsnähere Darstellung von Äußerungs- und Dokumentinformation
  - ✓ gezielte Suche erleichtert
  - ✓ Inferenzen
  
- ✗ Es gibt kein kleines, abgeschlossenes, universell anwendbares Inventar von Rollen, mit denen sich der gesamte Wortschatz beschreiben lässt.

▶ 27

## Komplexe prädikatenlog. Repräsentationen

- ▶ Kombination zu komplexen Repräsentationen durch logische Operatoren (Junktoren und Quantoren)

The big red block is in the box.

$\text{big}(b) \wedge \text{red}(b) \wedge \text{block}(b) \wedge \text{box}(x) \wedge \text{in}(b,x)$

Delfine sind Säugetiere.

$\forall d (\text{delfin}(d) \rightarrow \text{säugetier}(d))$

Delfine leben im Meer.

$\forall d (\text{delfin}(d) \rightarrow \exists m (\text{meer}(m) \wedge \text{lebt-in}(d,m)))$

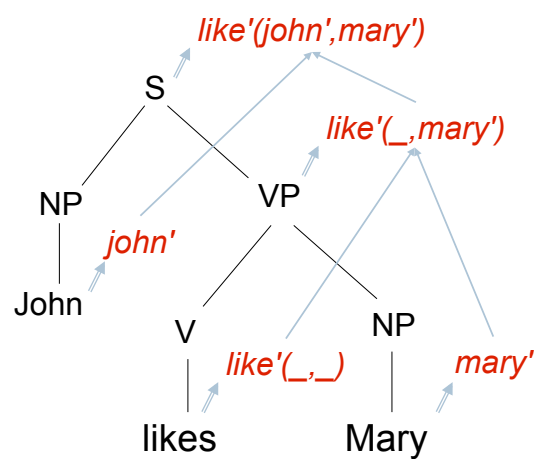
▶ 28

## Satzsemantik: Komposition

- ▶ Die **Bedeutung eines Satzes** berechnet sich aus
  - ▶ den **Bedeutungen der Wörter**, aus denen er sich zusammensetzt, und
  - ▶ seiner **syntaktischen Struktur**

▶ 29

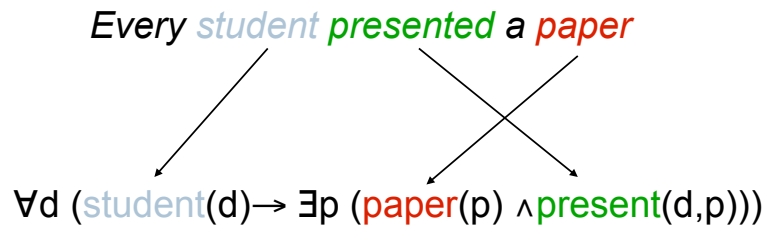
## Bedeutungskomposition: Ein Beispiel



▶ 30

## Bedeutungskomposition: Eine Herausforderung

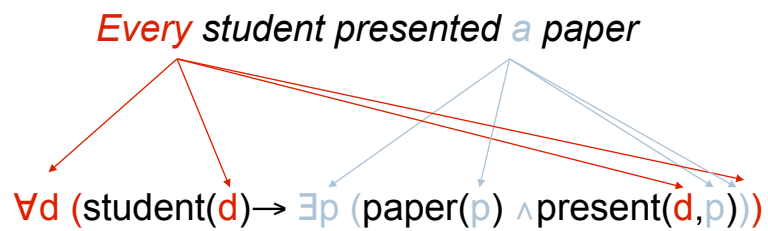
---



▶ 31

## A Challenge for Semantic Composition

---



▶ 32

## Ambiguitäten

---

- (1) *Zwei Fremdsprachen spricht jeder Linguist.*
- (2) *Alle sitzen in einem Boot*
- (3) *Jeder sitzt in einem Boot*
- (4) *Ein Teilnehmer von jedem Seminar hält zwei Referate.*
- (5) *800 000 Besucher tranken 2 Millionen Tassen Kaffee*  
(kollektive und kumulative Lesarten)

---

▶ 33

## Ambiguitäten im Diskurs

---

- ▶ **Manche Ambiguitäten werden im Diskurs aufgelöst.**
  - ▶ *Jeder Linguist spricht zwei Fremdsprachen. Es sind Englisch und Japanisch.*
- ▶ **Andere Ambiguitäten entstehen erst im Diskurs: Anaphern**
  - ▶ *Hans hat Peter nicht begrüßt. Er ist beleidigt.*
  - ▶ *Hans mag seinen Hund, obwohl er ihn manchmal beißt.*

---

▶ 34