

Einführung in die Computerlinguistik

Semantik

WS 2006/2007

Manfred Pinkal

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?

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?*

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?

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?

Was ist Wortbedeutung?



Was ist Wortbedeutung?



Lexikalische Semantik ...

... ist ein hoch komplexer Gegenstand:

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

Lexikalische Semantik ...

... ist äußerst aufwändig: Der Wortschatz ist

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

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.

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?*

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?*

Lexikalische Semantik und Informationszugriff

Elementare lexikalisch-semantische Relationen, z.B.

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

erschließen große Mengen zusätzlicher semantischer Information.

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...

Beispiel: engl. case

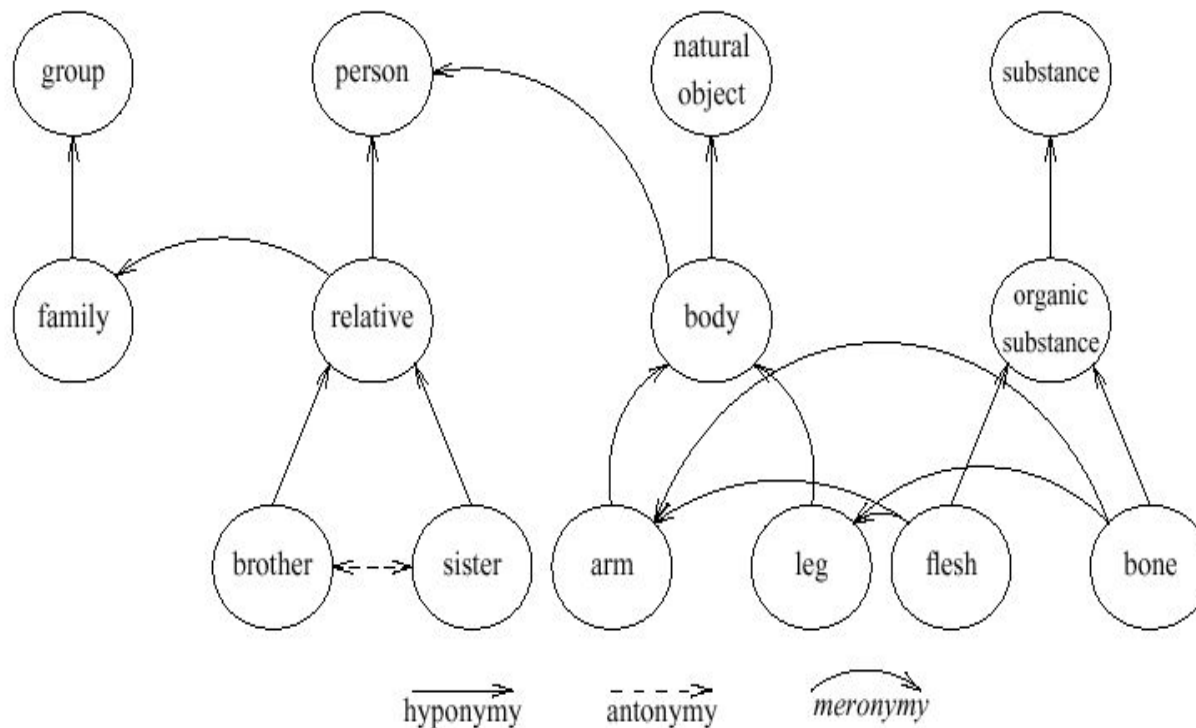
- {*case*, *carton*}
- {*case*, *bag*, *suitcase*}
- {*case*, *pillowcase*, *slip*}
- {*case*, *cabinet*, *console*}
- {*case*, *casing* (the enclosing frame around a door or window opening)}
- {*case* (a small portable metal container)}

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

WordNet-Teilhierarchie, Beispiel

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



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>

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?

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)

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)

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)

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.

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

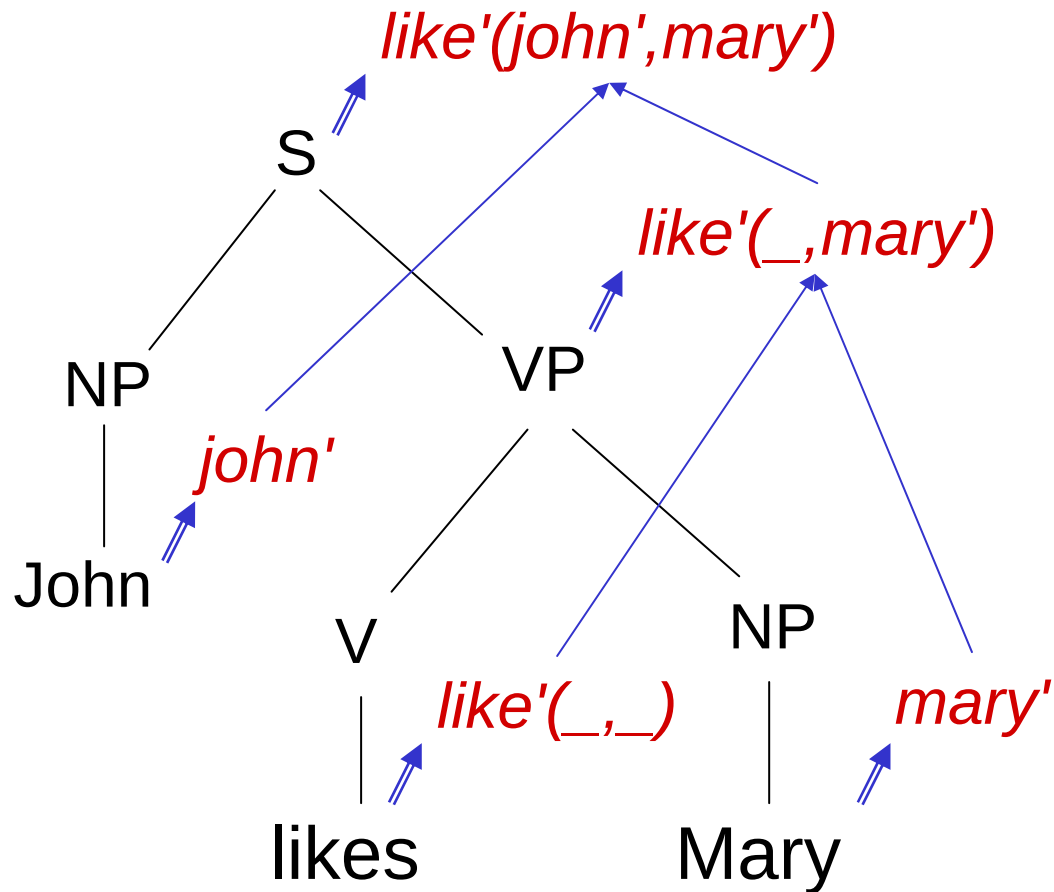
Delfine leben im Meer.

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

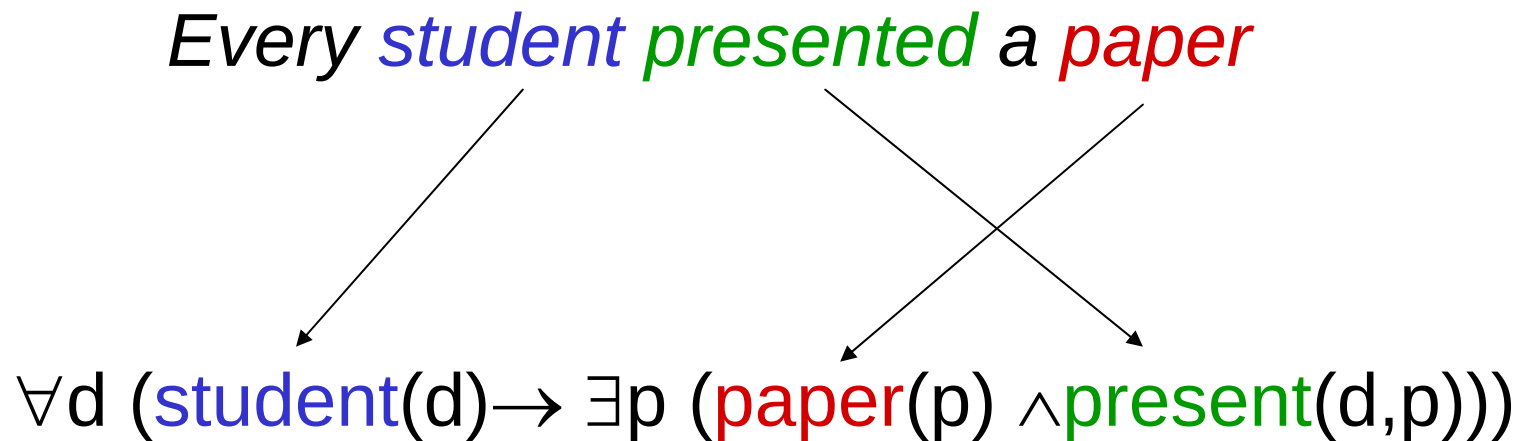
Satzsemantik: Komposition

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

Bedeutungskomposition: Ein Beispiel

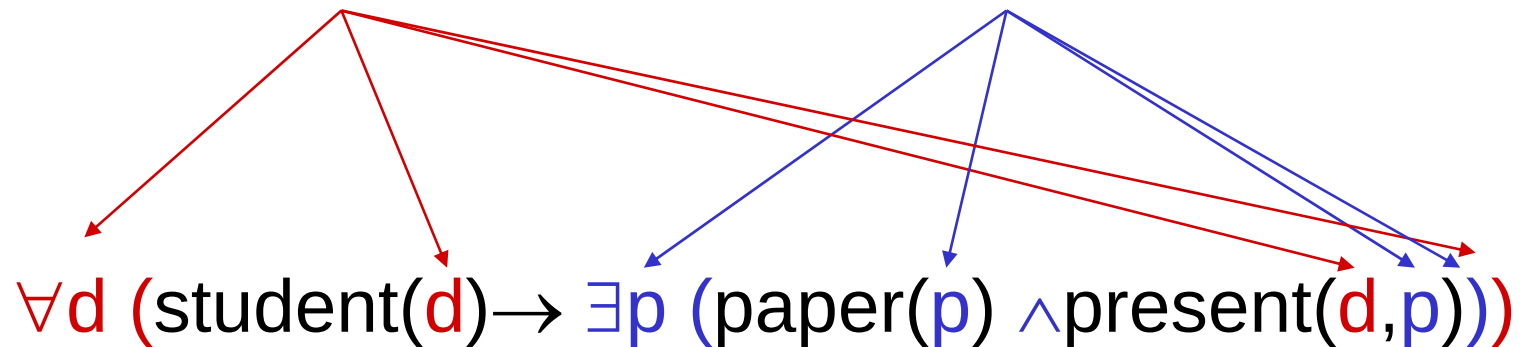


Bedeutungskomposition: Eine Herausforderung



A Challenge for Semantic Composition

Every student presented a paper



Quantoren und Skopus-Ambiguität

- (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)*