
Semantik (Bedeutungslehre)

Sebastian Padó

Semantik für Dialogverarbeitung (1)

- Wann fängt heute die CL-Vorlesung an?
- Sie beginnt um 11.

Wann = Frage nach Zeitpunkt/raum eines Ereignisses

anfangen = *erster* Zeitpunkt des Ereignisses

heute = Eingrenzung des Zeitraums (deiktisch!)

die CL-Vorlesung = Beschreibung des Ereignisses

Semantik für Dialogverarbeitung (2)

- *Ist der Flug um 2 nach London pünktlich?*
 - *Ihr Flug nach London Heathrow geht pünktlich um 14:05 ab und steht zum Einsteigen bereit.*
 - *Und der nächste?*
 - *Der nächste Flug nach London Gatwick geht um 15.20.*
 - *Gibt es auch einen nach Heathrow?*
-

Semantik für maschinelle Übersetzung (deutsch-englisch)

Geht es um 2 bei Ihnen?

Goes it about 2 with you?

Is 2 p.m. ok for you?

Can we meet at 2 in your office?

- Semantische Analyse des deutschen Satzes
 - Erstellung einer sprachunabhängigen Repräsentation
 - Generierung eines englischen Satzes
-

Semantik für Textverstehen und Fragebeantwortung

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

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

2.5 Mrd DM.

Semantik: Grundlegende Fragestellungen

- Wie kann Bedeutungsinformation grundsätzlich dargestellt werden? (**Semantische Repräsentation**)
 - Insbesondere: Wie werden Wortbedeutungen erfasst? (**lexikalische Semantik**)
 - Wie wird die Bedeutung von komplexen Ausdrücken (Sätzen) auf der Grundlage von Wortbedeutungen und der syntaktischen Struktur berechnet? (**Semantikkonstruktion/Satzsemantik**)
 - Insbesondere: Wie werden die bei Mehrdeutigkeiten (**Ambiguitäten**) die richtigen Lesarten konstruiert und aufgelöst?
-

Semantik: Grundlegende Fragestellungen

- Wie ergibt sich die spezifische Bedeutung einer Äußerung aus dem Kontext? Wie verändert eine Äußerung den Kontext?
(Diskurssemantik - siehe Beispiele auf letzten Folien)
 - Wie kommt man von der spezifischen Äußerungsbedeutung zur relevanten Information?
(Inferenz, semantische Auswertung)
 - Großbritannien hat Flugzeuge gekauft
→ Großbritannien besitzt Flugzeuge
 - Peter vergaß zwei Jahre lang, sein Alpenveilchen zu gießen
→ Das Alpenveilchen ist vertrocknet
→ Das Alpenveilchen ist tot
-

Wortbedeutung

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

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

Synonymie

WordNet

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

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 frame around a door or window)} | ■ Rahmen |
| □ { <i>case</i> (a small portable metal container)} | ■ Kasten/Etui/Hülle |
-

WordNet: Oberbegriffe im Nominalbereich

Table 1
List of 25 unique beginners for WordNet nouns

<i>{act, action, activity}</i>	<i>{natural object}</i>
<i>{animal, fauna}</i>	<i>{natural phenomenon}</i>
<i>{artifact}</i>	<i>{person, human being}</i>
<i>{attribute, property}</i>	<i>{plant, flora}</i>
<i>{body, corpus}</i>	<i>{possession}</i>
<i>{cognition, knowledge}</i>	<i>{process}</i>
<i>{communication}</i>	<i>{quantity, amount}</i>
<i>{event, happening}</i>	<i>{relation}</i>
<i>{feeling, emotion}</i>	<i>{shape}</i>
<i>{food}</i>	<i>{state, condition}</i>
<i>{group, collection}</i>	<i>{substance}</i>
<i>{location, place}</i>	<i>{time}</i>
<i>{motive}</i>	

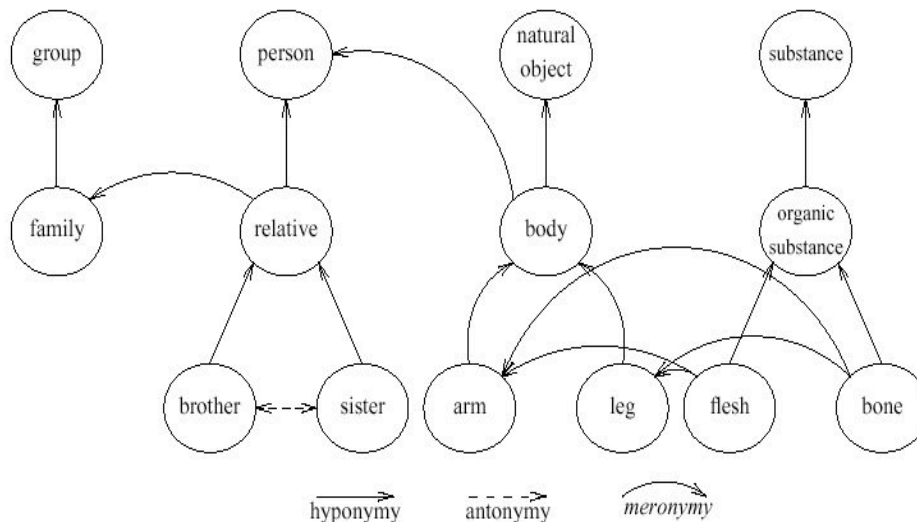
Weitere lexikalische Relationen

- 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)

WordNet-Teilhierarchie, Beispiel

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



Semantische Argumente

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

- Wieviel **kosten** die Flugzeuge, die Großbritannien bei Lockheed **bestellt** hat?

„Synonymie auf Satzebene“

- *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*

Nötig: Kategorien von semantischen Argumenten
(Mitspielern) einer Situation / eines Ereignisses

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
-

Satzanalysen

[Peter]_{ag} hat [Maria]_{rec} [das Buch]_{pat} gegeben

Agens = Peter

Rezipient = Maria

Patient = Buch

[Maria]_{rec} hat [das Buch]_{pat} [von Peter]_{ag} erhalten

Agens = Peter

Rezipient = Maria

Patient = Buch

Syntax-Semantik-Schnittstelle

- Zuordnung von semantischen Argumentpositionen und syntaktischen Komplementen im Lexikon:

<i>geben:</i>	Subj:	Agens
	Obj (akk):	Thema
	Obj (dat):	Rezipient

<i>erhalten:</i>	Subj:	Rezipient
	Obj (akk):	Thema
	PObj (von):	Agens

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.
-

Probleme mit thematischen Rollen (1)

Viele Kategorien von Mitspielern schließen sich nicht gegenseitig aus

[Peter]_{ag} hat [Maria]_{rec} [das Buch]_{pat} gegeben
[Maria]_{rec} hat [das Buch]_{pat} [von Peter]_{ag} erhalten
(Peter führt eine Handlung aus, durch die Maria etwas erhält)

[Maria]_{go} hat [das Buch]_{pat} [von Peter]_{so} erhalten
[Maria]_{go} hat [Peter]_{so} [das Buch]_{pat} gegeben
(Das Buch bewegt sich von A nach B)

Probleme mit thematischen Rollen (2)

Symmetrische Prädikate:

[Peter]_{rolle1} ähnelt [Hans]_{rolle1/2?}

[Hans]_{rolle1} ähnelt [Peter]_{rolle1/2?}

- Gleiche Rolle?
 - Gleiche Repräsentation
 - Entspricht nicht der Eindeutigkeit von thematischen Rollen
 - Probleme mit der Syntax-Semantik-Schnittstelle
 - Verschiedene Rollen?
 - Unterschiedliche Repräsentationen
-

FrameNet

- Es gibt keine generellen Rollen mehr
 - Einheiten für die Beschreibung der Wortsemantik sind **Frames** (Situationstypen) mit spezifischen Mitspielern
 - Lemmata in einem Frame (ca. 5-15 „targets“) beschreiben eine synonyme Situation
 - „WordNet auf Satzebene“
 - Pro Frame typischerweise frame-spezifische 3-5 Mitspieler („Frame-Elemente“)
 - Es gibt frame-übergreifende Generalisierungen von Rollen; sind aber nicht zentral bzw. obligatorisch
-

Beispiel: Frame Commercial_Transaction

- Targets:
 - buy, sell, pay, spend, cost, charge,
 - price, change, debt, credit, merchant, broker
 - tip, fee, honorarium, tuition

 - Frame-spezifische Rollen:
 - Money, Goods, Buyer, Seller, Price
-

Frame Elements von Commercial_Transaction

<u>FE</u>	<u>Example (in italics)</u>
Buyer	<i>Pat</i> bought a new guitar.
Seller	Pat bought a guitar <i>from Kim</i> .
Price	Kim sold the guitar <i>for \$250</i> .
Goods	Kim sold <i>the guitar</i> for \$250.
Rate	The plumber charges <i>\$20 an hour</i> .
Unit	The plumber charges <i>by the hour</i> .

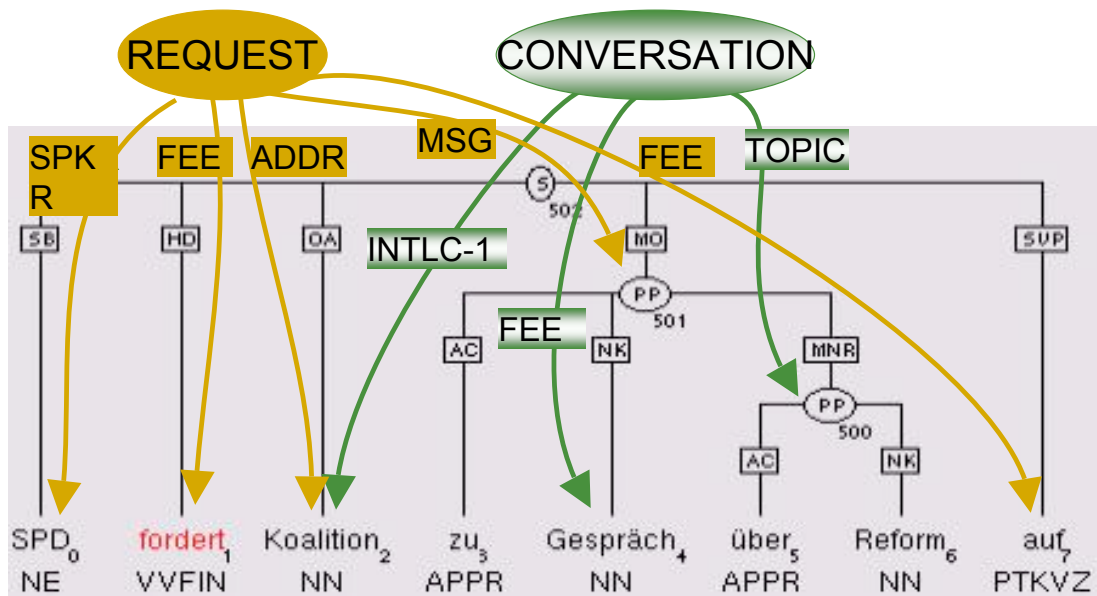
Frame Elements von Commitment

FE

Example (in italics)

Speaker	<i>Kim promised</i> to be on time.
Addressee	Kim promised <i>Pat</i> to be on time.
Message	Kim promised <i>to be on time</i> .
Topic	The government broke its promise <i>about taxes</i> .
Medium	Kim promised <i>in writing</i> to sell Pat the house.

Beispiel für deutschen Satz



Das FrameNet-Projekt in Berkeley

- FrameNet-Datenbank mit ca. 500 Frames (unvollständig)
 - Lexikon mit zur Zeit 6000 Einträgen
 - Angestrebt: 10-15000 Einträge
 - Frames bilden Verwandtschafts-Hierarchie (siehe WordNet!)
 - Verwandte Frames überlappen oft in Frame-Elementen
 - Beispiel: Alle Kommunikationsframes, z.B. COMMITMENT, CONVERSATION; REQUEST; STATEMENT teilen sich die Frame-Elemente: Speaker, Addressee, Message, Topic, Medium
-

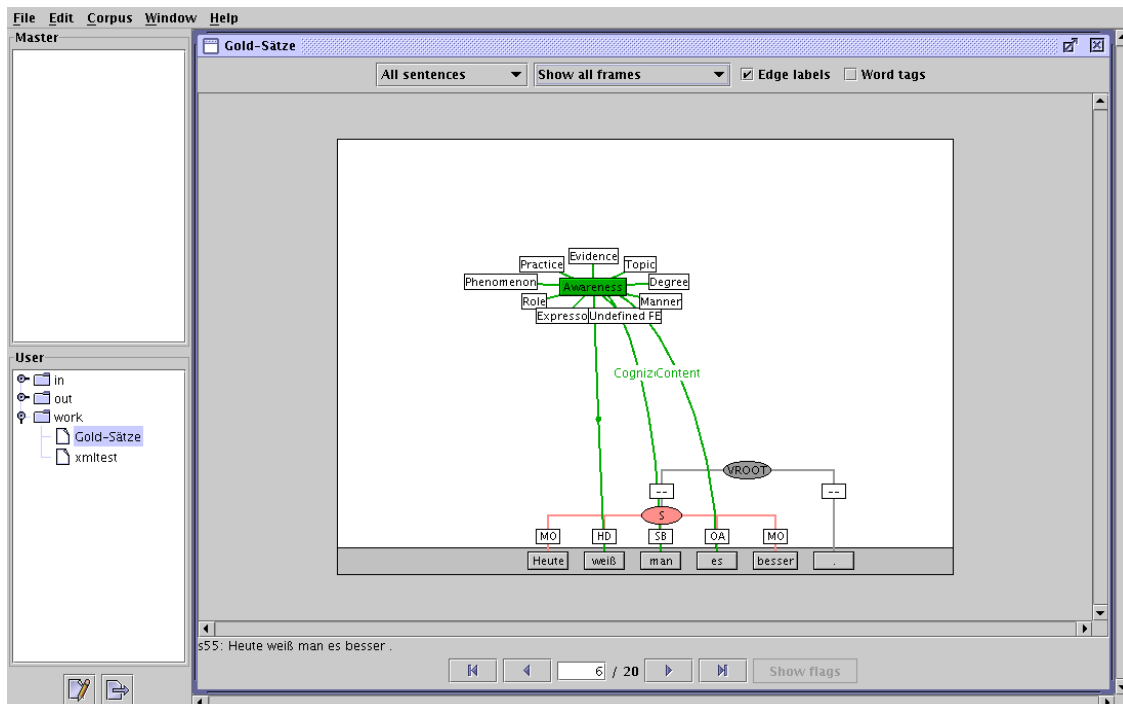
SALSA

- The Saarbrücken Lexical Semantics Annotation and Analysis Project
 - Manuelle Annotation eines deutschsprachigen Korpus: FR/TIGER Korpus mit 1.5 Mio. Textwörtern
 - Erstellung eines deutschen Frame-Lexikons
 - Entwicklung von Lernverfahren zur automatischen Annotation und zur Extraktion lexikalisch-semantischer Information
-

Schamlose Werbefolie

- Das SALSA-Projekt sucht Annotatoren
 - Markieren von Frames und Frame-Elemente in Zeitungstext
 - Interesse an linguistischer Arbeit (Unterscheidung von Wortbedeutungen)
 - Deutsche Muttersprachler
 - Gute Betreuung (wöchentliche Treffen)
- Bei Interesse mailt mir:
`pado@coli.uni-sb.de`

Das Annotationstool



Über semantische Argumente hinaus: Satzsemantische Ambiguitäten (1)

- Quantoren und Skopus
 - *Jeder Linguist spricht zwei Fremdsprachen*
 - *Ein Teilnehmer von jedem Kurs hält zwei Referate.*
 - Anaphern
 - *Hans hat Peter nicht begrüßt. Er ist beleidigt.*
 - *Hans mag seinen Hund, obwohl er ihn manchmal beißt.*
-

Satzsemantische Ambiguitäten (2)

- Ellipsen
 - *Hans verwöhnt seinen Hund, und Peter seine Katze.*
 - *Hans verwöhnt seinen Hund, und Peter auch.*

Nötig zur Ableitung von Satzambiguitäten

- reiches Lexikon
- ausdrucksstarker Formalismus zur Kombination von Wortbedeutungen

Nutzung von Logikformalismen

Kreative Verwendung von Sprache

Beispiel: Metapher und Metonymie

Börsenkurse auf Talfahrt
Der Betzenberg tobt
Ich stehe im Halteverbot

Fahren wir nach Stuttgart!
Treffen wir uns nach Ostern!
Treffen wir uns nach Stuttgart!

Oft nehmen Worte erst im Kontext besondere Bedeutungen an: fließender Übergang zur Semantikkonstruktion

Der “lexikalische Flaschenhals”

- Satzsemantik setzt Wortsemantik voraus
 - Großes Lexikon nötig
- Lexika werden i.A. per Hand erstellt
 - Beispiel: WordNet
 - Einzige Möglichkeit für logikbasierte Verfahren
- Nachteile
 - Riesiger Zeitaufwand
 - Neue Sprachen?
 - Konsistenz ?
 - Abdeckung?
 - Angemessenheit?

Automatisches Verfahren
wäre schön!

Distributionale Semantik

- “You shall know a word by the company it keeps” -- John Firth
- Ähnliche Worte kommen in ähnlichen Kontexten vor
 - Spezifischer: **Ähnliche** Worte kommen **ähnlich oft** mit **denselben (Kontext-)Worten** vor

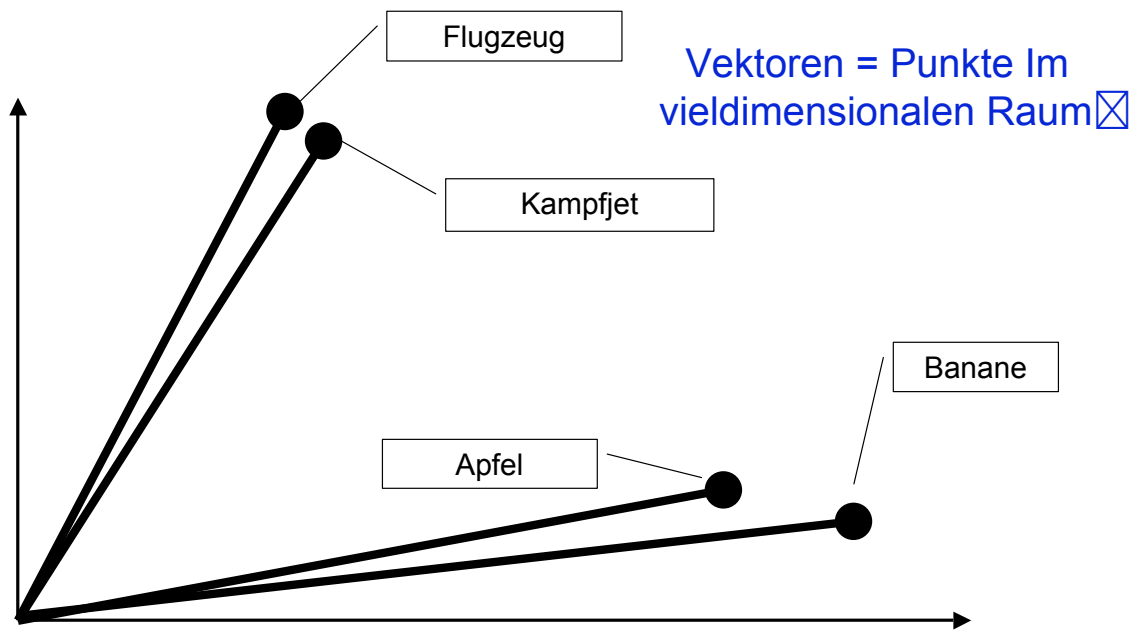
Strategie: Kontextwörter in Korpus auszählen
Bsp.: zwei Wörter links, zwei rechts

Wörter als Vektoren



	Flugzeug	Kampffjet	Apfel	Banane
Flughafen	100	80	4	2
Preis	50	55	8	10
Regierung	20	30	15	15
Luftwaffe	40	50	3	4
Supermarkt	2	1	30	35
frisch	0	0	20	25
krumm	0	0	3	50

Semantischer Raum (Projektion)



Distributionale Semantik

■ Vorteile

- Automatisch, keine tiefen Verfahren notwendig
 - Funktioniert für jede Sprache

■ Nachteile

- Keine symbolische Darstellung
- Ähnlichkeit von Vektoren schwer zu interpretieren
 - Synonymie: Flugzeug, Kampfjet
 - Hyponymie: Obst, Apfel
 - Antonymie: schwarz, weiss