
Semantische Sprachverarbeitung

Carola Eschenbach

Universität Hamburg, MIN-Fakultät, Dept. Informatik
AB Wissens- und Sprachverarbeitung (WSV)

Sommersemester 2011

IAF Computerlinguistik (Bachelor, Master)

CL1

- Einführung in die Linguistik

CL2

- VL Semantische Sprachverarbeitung: Eschenbach
 - VL Syntax und Parsing: Menzel
 - Übungen: Menzel & Eschenbach
-

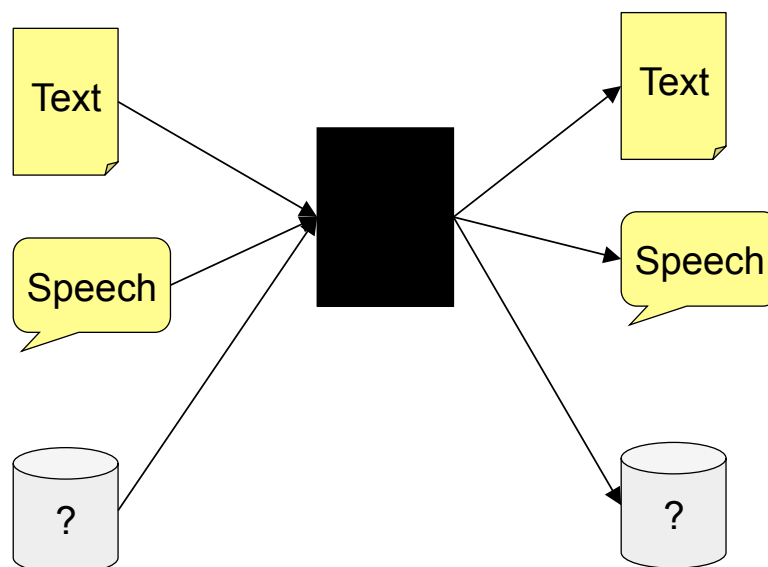
Semantische Sprachverarbeitung

Sitzung 1

- Einordnung, Motivation
- Bedeutung und Semantik
- Bedeutungsrelationen



Sprachverarbeitung



Semantische Sprachverarbeitung

Inhaltsbezogener Zugriff auf Dokumente (Information Retrieval)

- Erkennung der besten Kandidaten in einem großen Dokumentenbestand
- Internet-Suchmaschinen
- Knowledge-Management

Automatische Textzusammenfassung

- Extraktion von wesentlichen Aussagen

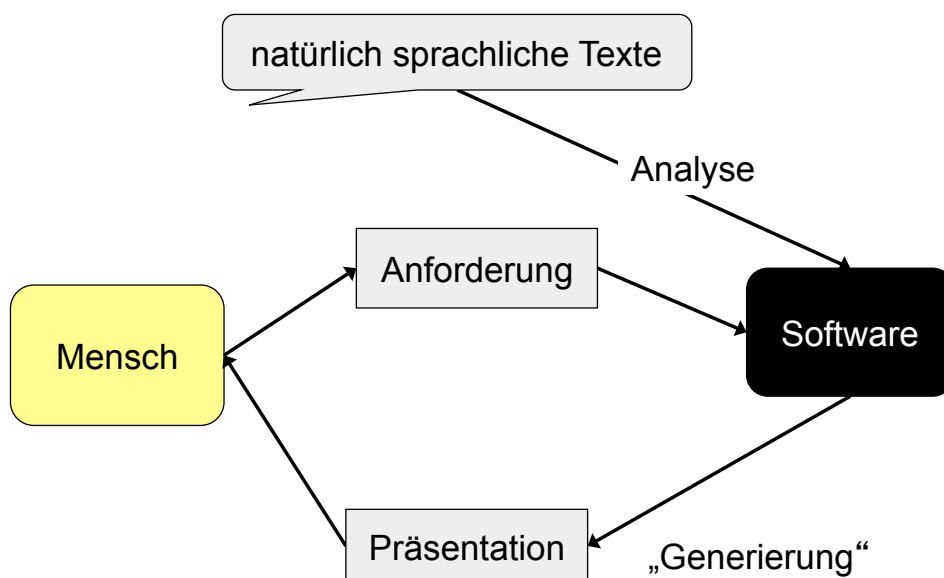
Maschinelle Übersetzung

- Erschließung von Texten unbekannter Sprachen
 - Automatische Paraphrasierung

Mensch-Maschine-Interaktion

- Anfragesysteme
 - Fragebeantwortung
- **Berücksichtigt die Semantik (Bedeutung)**

Sprachbasierte Service-Software



Beispiel: Eine Wegbeschreibung

Du verlässt Haus R über die Treppe am Hintereingang von Haus F. Du gehst an Haus F entlang Richtung Pförtner, so als ob Du das Gelände verlassen würdest. Rechts von Dir liegt erst mal Gebäude D, dann kommen drei riesige Schwarzpappeln, und erst dahinter biegst Du rechts in einen Weg ein, der Dich über einen größeren Platz weiter geradeaus an Haus B entlang führt. Du gehst dann links in das Gebäude hinein und ein Stockwerk hoch, da ist die Mensa.


[Startseite](#) › [Werkzeuge](#)

Babel Fish-Übersetzung

[Hilfe](#)

Textblock übersetzen - Maximal 150 Wörter

Englisch ins Deutsche 

Übersetzen Sie eine Webseite

http://

Englisch ins Deutsche 

Fügen Sie [Babel Fish Übersetzung](#) zu Ihrer Website hinzu.
Tipp: Seiten mit Frames können nun übersetzt werden.



Babelfish-Übersetzungstools

[E-Mails übersetzen!](#)

[Fügen Sie die Übersetzung zu Ihrer privaten oder geschäftlichen Site hinzu.](#)

[Nahtlose Übersetzungs-Plugins](#)
für MSOffice - Word, PowerPoint, Excel, Internet Explorer und Outlook!

NEW SYSTRAN 4.0
日本語 简体中文 中文
SUPPORTS ASIAN LANGUAGES



You leaves house R over the stairs at the back door of house F. Du go at house F along direction doorman, thus as whether you the area would leave. Right of you only times building D lies, then three enormous Schwarzpappeln come, and only behind it you bend right into a way, which continues to lead you across a larger place straightforward at house B along. You go then left into the building in and a floor highly, there are the refectory.

Übersetzungsprobleme

Mehrdeutigkeit / Wortwahl

- über die Treppe ⇒ **over** the stairs
- biegst Du ... ein ⇒ you **bend** ... into ...
- ein Stockwerk hoch ⇒ a floor **highly**

Wortform / Syntax

- Du verlässt ⇒ You **leaves**
- da ist die Mensa ⇒ there **are** the refectory
- das Gelände verlassen würdest ⇒ **the area would leave**

Komplexität der Ausdrücke

- so also ob ⇒ **thus as whether**
- an Haus F entlang ⇒ **at** house F **along**
- erst mal ⇒ **only times**

Unbekannte Wörter

- **Schwarzpappeln, Du** (?)

Ko-Referenz (Wortwahl)

- dahinter ⇒ behind **it**

Hintergrund der Probleme

Bedeutungsbeziehungen

- nur auf Wortebene erfasst
- keine Nutzung von Interaktionen
- keine Nutzung der sprachlichen Struktur

Funktioniert nur

- wenn der interpretierende Mensch die Zusammenhänge selbst herstellt

Keine 'einfachere Sprache'

Rückübersetzung: <http://babelfish.altavista.com/>

Sie verlassen Haus R über der Treppe an der Hintertür des Hauses F. DU gehen an Haus F entlang Richtung Doorman, so wie, ob Sie, die der Bereich gehen würde. Recht von Ihnen setzt nur Zeit Lügen des Gebäudes D fest, dann kommen drei enormes Schwarzpappeln, und nur hinter ihm verbiegen Sie nach rechts in eine Weise, die fortfährt, Sie über einem größeren Platz entlang zu führen, der an Haus B direkt ist. Sie steigen dann nach links in das Gebäude innen ein und ein Fußboden in hohem Grade, dort sind das Refektorium.

SPAM-Email (1)

Glückwunsch-Sir/Frau, Du bist als einer von zwei Siegern des BRITISCHEN NATIONALEN LOTTERIE-Computerstimmzettels zeichnest vorgewählt worden und folglich bist du eine privilegierte Empfänger des großartigen Betragpreiss von £650,412.00 (sechs hundert und fünfzig tausend vierhundert und zwölf Briten zerstoßen Sterling), im Bargeld. Du/deine Firmae-mail Adresse angebracht zur Karte Zahl: UK/000056389 mit Seriennummer 4398/05 zeichnete die glücklichen Zahlen: (12, 13, 25, 30,45and 47) und die Prämie Nr. (42). Um für deinen Anspruch einzuordnen, in Verbindung treten unser Regulierer unten mit den folgenden Informationen.

Mrs. Maria Warner-Feeko

E-mail: mariawarner_uknldept@yahoo.co.uk

Voller Name Kontakt-Adresse

Nationalität Geschlecht Alter

Telefon/Telefax-Zahl Zunächst von den Stämmen

SPAM-Email (2)

Congratulations Sir/Madam,

You have been selected as one of two winners of the UK NATIONAL LOTTERY computer ballot draws and thus you are a privileged recipient of the grand draw prize of £650,412.00 (Six Hundred and Fifty Thousand Four Hundred and Twelve British pounds sterling) in cash. You/Your company e-mail address attached to ticket number: UK/000056389 with Serial number 4398/05 drew the lucky numbers :(12, 13, 25, 30,45and 47) and the bonus number (42). To file for your claim, contact our claims agent below with the following information.

Mrs. Maria Warner-Feeko

E-mail: mariawarner_uknldept@yahoo.co.uk

Full Name..... Contact Address.....

Nationality.....

Sex..... Age..... Tel/Fax

Number..... Next of Kin.....

Zusammenfassung

Word-Zusammenfassung (50%)

- Du verlässt Haus R über die Treppe am Hintereingang von Haus F. Du gehst an Haus F entlang Richtung Pförtner, so als ob Du das Gelände verlassen würdest. Du gehst dann links in das Gebäude hinein und ein Stockwerk hoch, da ist die Mensa.

Apple-Summary-Service (~50%)

- Rechts von Dir liegt erst mal Gebäude D, dann kommen drei riesige Schwarzpappeln, und erst dahinter biegst Du rechts in einen Weg ein, der Dich über einen größeren Platz weiter geradeaus an Haus B entlang führt.

Word – Apple – Mensch

- Du verlässt Haus R über die Treppe am Hintereingang von Haus F. Du gehst an Haus F entlang Richtung Pförtner, so als ob Du das Gelände verlassen würdest. Rechts von Dir liegt erst mal Gebäude D, dann kommen drei riesige Schwarzpappeln, und erst dahinter biegst Du rechts in einen Weg ein, der Dich über einen größeren Platz weiter geradeaus an Haus B entlang führt. Du gehst dann links in das Gebäude hinein und ein Stockwerk hoch, da ist die Mensa
- Verlass Haus R über die Treppe am Hintereingang von Haus F. Geh daran entlang, an Haus D und den Bäumen vorbei, dann nach rechts. Geh über den Platz, an Haus B entlang, dann links in das Haus. Im ersten Stock ist die Mensa.

Hintergrund der Probleme

Keine Erfassung der Textbedeutung

- Auswahl beruht nicht auf Bedeutung
 - Der Zweck der Wegbeschreibung und
 - daraus erwachsende Anforderungen
 - werden nicht berücksichtigt
- Keine Schlussfolgerungen über Gehalt von Text und Text-Abschnitten
- Keine Gewichtung nach Informativität
- Keine Berücksichtigung von Kontextbezügen (Anaphorik, imaginäre Wanderung)

Die flache Verarbeitung

Modellierung thematischer Bezüge

- Wahl eines Ausdrucks mit richtigem Bezug

keine (möglichst wenig) strukturelle Beschränkungen

- Unabhängigkeit von syntaktischen Strukturen
- Funktionstüchtigkeit bei fehlender Syntax

keine (möglichst wenig) lexikalische Beschränkung

- Einsatz für große Sprach-Korpora (unrestringierte Texte)
- Anpassbarkeit auf andere Sprachen

Einordnung

Natürliche Sprache wird definiert allein

- durch den Menschen bzw.
- eine Sprachgemeinschaft.

kognitiv orientierte Modellierung

- ist für die tiefen Sprachverarbeitung erforderlich
- auch bei starker Anwendungsorientierung
 - ‚Was tun Menschen, wenn sie einen Text verstehen / interpretieren.‘
 - ‚Was kann ein Mensch an Versteheleistung bei seinem Kommunikationspartner (normalerweise) voraussetzen.‘

Vorkenntnisse

Linguistik / Sprachwissenschaft

Syntax, Parsing

Logik, Lambda-Kalkül

KI / Wissensbasierte Systeme

PROLOG / Unifikation

Organisatorisches

Vorlesung

- regulärer Termin: Fr 10:15 – 11:45

Unterlagen

- Elektronisch: www.informatik.uni-hamburg/wsv/teaching/vorlesungen/SemSprach_SoSe11
- FBI → Zentren und Arbeitsbereiche → WSV → Lehre → Semantische Sprachverarbeitung
- inkl. Literaturangaben und interessante Links

Übungen

- Regulärer Termin: Di 8:30 – 10:00, im Wechsel mit ‚Syntax und Parsing‘
- Problemerkörterungen, Diskussion
- Aufgaben / Programmier-Übungen
- <http://nats-www.informatik.uni-hamburg.de/view/CompLing11/UebungenCL>
- WSV → Lehre → Übung zu IAF Computerlinguistik II

Literatur: Sprachorientierte KI

- Allen, James (1987). *Natural Language Understanding*. Menlo Park, Ca: Benjamin/Cummings.
- Blackburn, Patrick & Johan Bos (1999). *Representation and Inference for Natural Language. A First Course in Computational Semantics*. Ms. **Auf Anfragen**
- Blackburn, Patrick & Johan Bos (1999). *Working with Discourse Representation Theory. An Advanced Course in Computational Semantics*. Ms. **Online**
- Jurafsky, Daniel & James H. Martin (2000). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Literatur: Linguistik

- Chierchia, Gennaro & Sally McConnell-Ginet (1990). *Meaning and Grammar. An Introduction to Semantics*. Cambridge, MA: MIT Press.
- Cruse, D.A. (1986). *Lexical Semantics*. Cambridge, NY: Cambridge University Press.
- Gamut, L.T.F. (1991). *Logic, Language, and Meaning 1 & 2*. Chicago & London: University of Chicago Press.
- Löbner, Sebastian (2002). *Understanding Semantics*. London: Arnold.
- Löbner, Sebastian (2003). *Semantik. Eine Einführung*. Berlin: Walter de Gruyter.
- Lyons, John (1980 & 1983). *Semantik 1 & 2*. München: Beck.

Sprache und Bedeutung

Sprachliche Ausdrücke

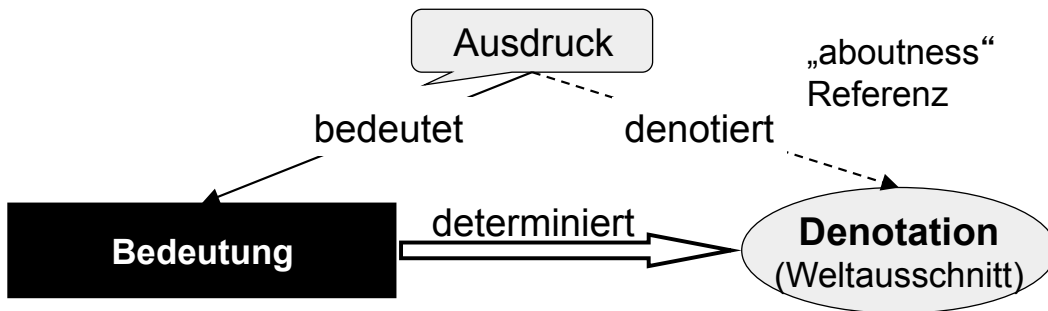
- werden vom Menschen produziert und interpretiert
- beschreiben Situationen in der Welt

Die Bedeutung

- determiniert, welche Situationen beschrieben werden
- korrespondiert zu mentalen Repräsentationen

Bedeutungen

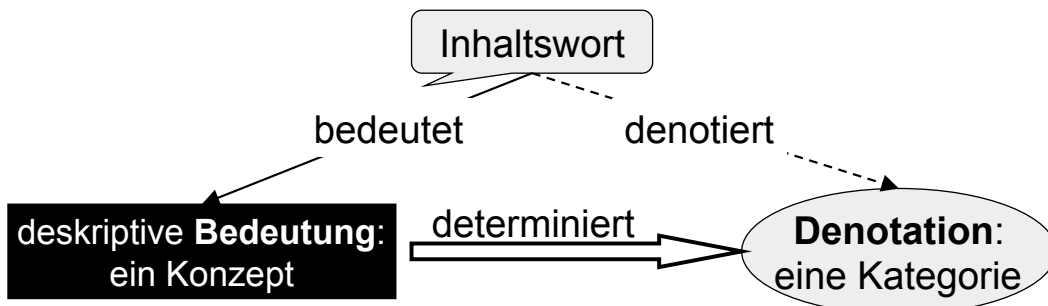
- grundlegender als Denotation
- abstrakt



Löbner (2003)

Bedeutungen

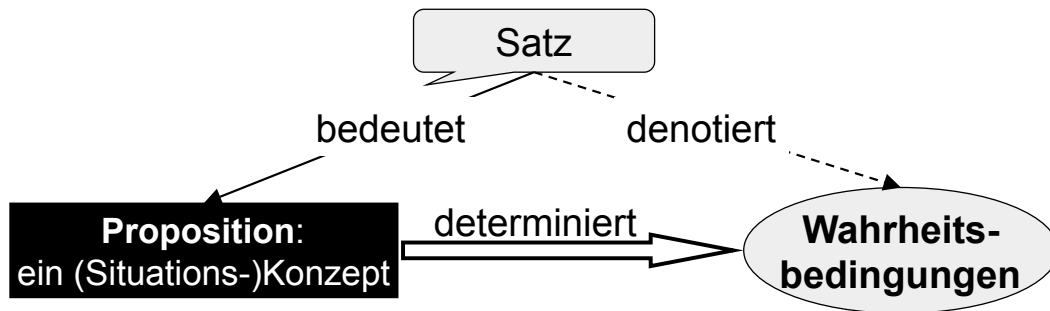
- grundlegender als Denotation
- abstrakt



Löbner (2003), Abb. 2.2 (S. 32)

Bedeutungen

- grundlegender als Denotation
- abstrakt



Löbner (2003), Abb. 2.3 (S. 34)

Bedeutung

- grundlegender als Denotation
- abstrakt
- von verschiedenen Menschen geteilt
 - Übereinkunft einer Sprachgemeinschaft
- mental repräsentiert

Mehrdeutigkeit (Ambiguität), Kontextabhängigkeit

Mehrdeutigkeit

- Lexikalisch
 - Jeden Morgen geht Quasimodo zu seiner Bank.
- Syntaktisch, strukturell
 - Kompetente Frauen und Männer haben in dieser Firma Einfluss.
- Skopus
 - Jede Studentin nimmt an drei Vorlesungen teil.
- Pragmatisch
 - Kannst Du lauter sprechen?

Vagheit, Unterspezifität

- Viele Studenten müssen Geld verdienen.
- Rüdiger sitzt im Hörsaal.

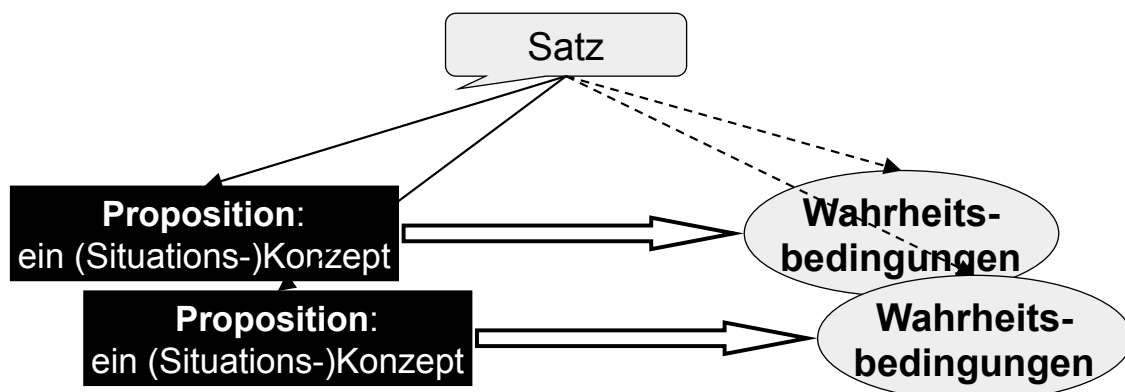
Deixis, Indexikalität

- Ich, wir, du, ihr, hier, jetzt, ...

Mehrdeutigkeit (Arbeitsdefinition)

Ein Satz ist mehrdeutig

- wenn er bezogen auf dieselbe Situation sowohl als falsch als auch als wahr angesehen werden kann.



Bedeutungsrelationen

- Empirisch feststellbare Relationen zwischen sprachlichen Ausdrücken, die ihren Bedeutungen zuzuschreiben sind.
- Theorien über die Bedeutung sprachlicher Ausdrücke sollten die Bedeutungsrelationen rekonstruieren.

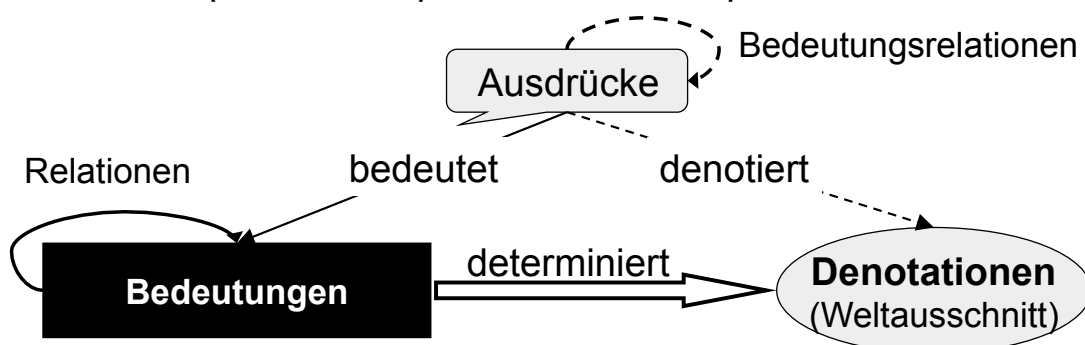
Beispiel: Folgerung (Entailment)

- 1) Joachim küsst Kordula.
- 2) Kordula wird geküsst.
- 3) Joachim berührt Kordula.

Die Information, die in 2) und 3) enthalten ist, ist bereits in 1) enthalten

Bedeutungen

- von verschiedenen Menschen geteilt
 - identifizierbar über Inferenzpotential
- mental repräsentiert
 - Adäquatheit entsprechend Inferenzpotential



in Anlehnung an Löbner (2003)

Bedeutungsrelationen: B folgt genau dann aus A,

- wenn die Information, die B trägt, in der Information, die A trägt, enthalten ist
- wenn immer dann, wenn A wahr ist, auch B wahr ist
- wenn jede Situation, die durch A beschreibbar ist, auch durch B beschreibbar ist
- wenn "A, aber nicht B" nicht wahr sein kann / keine Situation beschreiben kann

- 1) Joachim küsst Kordula.
 - 2) Kordula wird geküsst.
 - 3) Joachim berührt Kordula.
- 2) und 3) folgen aus 1)

Folgerung

A

- | | |
|---------------------------|---------------------------------|
| 1) Dies ist ein A. | 1) Dies ist ein Elefant. |
| 2) Dies ist blau. | 2) Dies ist groß. |
| 3) Dies ist ein blaues A. | 3) Dies ist ein großer Elefant. |
| 3) folgt aus 1) und 2) | 3) folgt nicht aus 1) und 2) |

- Folgerung ist eine Beziehung zwischen Sätzen, die von der Bedeutung der Sätze und nicht allein der Form der Sätze abhängt.

Bedeutungserhaltung (Arbeitsdef.)

Ein Verarbeitungsprozess für natürliche Sprache

- ist (exakt) bedeutungserhaltend, wenn alle Folgerungen aus dem Eingabetext auch Folgerungen aus dem Ausgabertext sind.
 - das ist bei der Übersetzung nicht immer möglich
 - [Deutsch] Sie ist groß. (eine weibliche Person ...)
 - [Finnisch] Hän on iso. (eine Person ...)
- ist (semantisch) korrekt, wenn der Ausgabertext (und damit alle seine Folgerungen) aus dem Eingabetext folgt.
 - semantisch korrekte Verarbeitung könnte aber pragmatisch ungenügend sein.

Präsuppositionen: Beispiele

- 1.a) Petras Freund ist blond.
- 1.b) Petras Freund ist nicht blond.
- 1.c) Wenn Petras Freund blond ist, dann ist er langhaarig.
- 1.d) Ist Petras Freund blond?

- 1.e) Es gibt (gegenwärtig genau) einen Freund von Petra.

- 2.a) Die größte Primzahl ist gerade.
- 2.b) Die größte Primzahl ist nicht gerade.
- 2.c) Wenn die größte Primzahl gerade ist, dann endet sie auf 2.
- 2.d) Ist die größte Primzahl gerade?

- 2.e) Es gibt eine größte Primzahl.

Präsuppositionen: Beispiele

- 3.a) Nils hat aufgehört, mit dem Auto zur Arbeit zu fahren.
- 3.b) Nils hat nicht aufgehört, mit dem Auto zur Arbeit zu fahren.
- 3.c) Wenn Nils aufgehört hat, mit dem Auto zur Arbeit zu fahren, dann kann er mich nicht mitnehmen.
- 3.d) Hat Nils aufgehört, mit dem Auto zur Arbeit zu fahren?
- 3.e) Nils ist regelmäßig mit dem Auto zur Arbeit gefahren.

A präsupponiert B / B ist eine Präsupposition von A genau dann, wenn

- die Beurteilbarkeit von A die Wahrheit von B voraussetzt und entsprechendes für die korrespondierenden Sätze des Schemas a)-d) gilt.

Aufhebung von Präsuppositionen

Keine implizite Aufhebung

- *Petras Freund ist nicht blond, aber Petra hat gar keinen Freund.
- *Die größte Primzahl ist ungerade und es gibt gar keine größte Primzahl.
- *Nils hat aufgehört, mit dem Auto zur Arbeit zu fahren, aber er hat auch gar kein Auto.

Explizite Zurückweisung

- Petras Freund ist *nicht* blond, *denn* Petra hat gar keinen Freund.
- Die größte Primzahl ist *nicht* gerade, *denn* es gibt gar keine größte Primzahl.
- Nils hat *nicht* aufgehört, mit dem Auto zur Arbeit zu fahren, *denn* er ist nie mit dem Auto zur Arbeit gefahren.

Kontradiktion, Widerspruch (1)

Beispiele

- Als Stephan in Paris war, war er nicht in Frankreich.
- Dieses ist ein farbloser blauer Buchstabe.

A ist genau dann widersprüchlich, wenn

- A niemals wahr sein kann,
- keine Situation möglich ist, die durch A beschreibbar wäre.

Kontradiktion, Widerspruch (2)

Beispiele

- Thorsten liest die "Wahlverwandtschaften". Thorsten liest niemals Goethe.
- Ulrich ist Junggeselle. Ulrich ist verheiratet.

A und B sind genau dann widersprüchlich, wenn

- A und B nicht beide wahr sein können,
- keine Situation, die durch A beschreibbar ist, auch durch B beschreibbar ist.

Anomalien

Beispiele

- Die Quadratwurzel von Veronikas Schreibtisch trinkt Menschlichkeit.
- Colorless green ideas sleep furiously.
- Lachen ist sehr regnerisch.
- Die Tatsache, dass Käse grün ist, verschwand unbemerkt.

Eine Aussage A ist genau dann eine Anomalie, wenn

- A weder wahr noch falsch sein kann,
- nicht feststellbar ist, ob A eine Situation beschreibt.

Koreferenz und Koindizierung

(Bestimmte) Nominalphrasen referieren

- Wenn [Christian]₁ sieht, dass [der Himmel]₂ klar ist, kommt [er]₁ mit [dem Fahrrad]₃.
- Wenn [sie]₄ anruft, sag [Steffi]₄, dass [ich]₅ schon zu[m Tennisspielen]₆ gegangen bin.
- [She]₇ is proud of [herself]₇.
- [Er]₈ wäscht [sich]₈ täglich.
- [Er]₉ wäscht [ihn]₁₀ täglich.

Koindizierung

- ist ein übliches Mittel, Koreferenz (= Referenz auf dieselbe Entität) anzuzeigen.

Koreferenz und Koindizierung

Anaphorik

- Wird die Referenzbestimmung einer Nominalphrase allein durch eine andere Nominalphrase vermittelt, dann heißt die erste "Anapher" und die zweite "Antezedent".

Koindizierung ohne Referenz

- [Jede Frau]₁ glaubt, dass [sie]₁ [Kinder]₂ besser erziehen kann als [[ihre]₁ Mutter]₃.
- [Kein Mann]₄ sollte [sich]₄ [der Leistungen [[seiner]₄ Kinder]₅]₆ brüsten.
- [Welcher Kandidat]₇ wird [sich selbst]₇ wählen?

Sprache und Bedeutung

Sprachliche Ausdrücke

- werden vom Menschen produziert und interpretiert
- beschreiben Situationen in der Welt

Die Bedeutung

- determiniert, welche Situationen beschrieben werden
- korrespondiert zu mentalen Repräsentationen
- kann (nur) über Bedeutungsrelationen erfasst werden

Tiefe semantische Sprachverarbeitung

- unabdingbare Orientierung an der menschlichen Sprachfähigkeit und -verarbeitung