

Kurs: 1699 – Automatische Sprachverarbeitung  
Datum: 2.3.2005  
Prüfer: Prof. Helbig  
Note: 1  
Dauer: ca. 25 Minuten

Welche Schichten der Verarbeitung natürlicher Sprache gibt es?

*akustische Erkennung, morphologische Analyse*

Aha, was passiert da?

*Zerlegung in Morpheme*

Was erhält man als Ergebnis?

*Wortstamm (daraus Grundwort), Informationen über Flexion, ...*

konkret am Beispiel un-er-freu-lich-em

*un --> Negation, em --> Dativ, ...*

Was kommt nach der morphologischen Analyse?

*lexikalische Analyse*

Was passiert da?

*Grundwort nachschlagen im Lexikon bringt Informationen zu Wortart, Flexionstypen, Kombinierbarkeit*

Was für ein Problem gibt es dabei?

*lexikalische Mehrdeutigkeiten, z.B. Homographen (sein): gleiche Schreibweise, unterschiedliches*

*Verhalten, auch unterschiedliche Wortarten*

Was gibt es neben der Homographie?

*Polysemie, Beispiel Pferd, gleiche Schreibweise, gleiches Verhalten*

Was gibt es als drittes (nicht schlimm, wenn Sie das nicht wissen)?

*Bedeutungsmoleküle, z.B. Schule, Bedeutungswechsel innerhalb des Satzes möglich*

weitere Beispiele?

*Kirche, Parlament, Oper*

nächster Schritt nach der lexikalischen Analyse?

*syntaktische Analyse: Satzstruktur*

Wie erfolgt das? Was gibt es da für Möglichkeiten?

*äh, z.B. Top-Down / Bottom-Up?*

Ja gut, was braucht man dafür?

*Grammatik*

Was ist eine Grammatik nach Chomsky?

*Definition  $G = \langle V_N, V_T, P, S \rangle$  mit  $V_N, V_T$  disjunkt, Produktionsregeln Abbildung Zeichenkette auf*

*Zeichenkette, Startsymbol aus  $V_N$  ... Typ-Hierarchie mit allgemeinstem Typ 0*

Worin unterscheiden sich die Typen?

*Form der Regeln*

Wie sehen die Produktionsregeln für uneingeschränkte und kontextfreie Grammatiken aus?

*uneingeschränkt: beliebige Zeichenketten aus  $V^*$ , keine Ableitung nach  $\epsilon$*

*kontextfrei: genau ein nichtterminales Symbol abgeleitet nach beliebiger Zeichenkette aus  $V^*$  (nicht  $\epsilon$ )*

Welcher Typ ist denn für die Beschreibung natürlicher Sprache geeignet?

*eigentlich CSG, aber*

Aber wo ist das Problem?

*keine effizienten Algorithmen*

Welchen Typ nimmt man stattdessen?

*CFG*

Schreiben Sie eine einfache CFG auf, mit der sich der Satz "Der kleine Hund schläft in der Hütte." ableiten lässt!

*$S \rightarrow NP VP PP$ ;  $NP \rightarrow det NP_0$  usw.*

Wie würde damit das Ergebnis der Analyse des Satzes aussehen?

*Konstituentenbaum aufgezeichnet*

Welche Probleme treten bei CFG auf?

*Abbildung von Kongruenzen, z.B. zwischen Artikel und Nomen*

Worin kongruieren die?

*Genus, Numerus, Kasus*

Könnte man das Problem der Kongruenz in CFG lösen?

*Ja, mit Unterkategorien für jede Kombination von Kongruenzen*

Was lässt sich in CFG aber nicht lösen?

*Verben wie abschreiben: "Er schreibt von ihr ab." in CFG nicht darstellbar.*

Was treten auf syntaktischer Ebene für Mehrdeutigkeiten auf?

*strukturelle Mehrdeutigkeiten, z.B. Anbindung der PP (Grammatikregeln von oben um VP --> verb PP erweitert)*

Wo sind solche Mehrdeutigkeiten typisch?

*Abhängigkeit der PP von Verb oder NP*

Beispiel?

*Polizei erschoss Mann mit Schwert*

Wie lassen sich diese Mehrdeutigkeiten auflösen?

*verschiedene Prinzipien, am sinnvollsten nach lexikalischer Präferenz: Valenzen von Verb und NP auswerten*

Welche Mehrdeutigkeiten gibt es neben lexikalischen und strukturellen noch?

*Konstituenten- vs. Satznegation*

Wie werden semantische Informationen festgehalten?

*Feature-Strukturen*

Welche Grammatiken verwenden die?

*LFG, HPSG*

Wie geschieht das bei der LFG?

*über Annotationen an den CFG-Regeln --> beispielhaft eingezeichnet für erste Regel*

Was bedeutet das?

*↑F-Struktur der übergeordneten Konstituente, ↓ der Konstituente selbst; für NP: F-Struktur entspricht dem Subjekt-Teil der F-Struktur des Satzes*

Wie kann man semantische Zusammenhänge noch darstellen?

*semantische Netze*

Zeichnen Sie ein semantisches Netz für den Satz "Der Junge schlägt das Pferd mit der Peitsche."

*aufgezeichnet, Pferd mit OBJ oder AFF an Sachverhalt, wichtig auseinander zu halten generische und spezielle Objekte für Junge, Pferd, Peitsche (jeweils mit SUB-Kante verbunden)*

Wie müsste das Netz erweitert werden für den Satz "Der Junge schlägt sein Pferd mit der Peitsche."

*POSS-Kante vom speziellen Junge-Knoten zum speziellen Peitsche-Knoten eingezeichnet*

Nanu?

*ups*

Was würde das heißen?

*"Der Junge schlägt das Pferd mit seiner Peitsche."*

Unterschied zwischen den beiden "Pferd-Knoten"?

*speziell (genau das Pferd, das der Junge schlägt) / generisch (Klasse der Pferde)*

Wie kann man Semantik noch formal repräsentieren?

*nach einigem hin und her: Prädikatenlogik*

Welche Beschränkungen hat die Prädikatenlogik bei der Semantikdarstellung?

*keine Quantoren für unscharfe Mengenabgrenzungen, z.B. fast alle, wenige*

Wer hat sich mit diesen Quantoren beschäftigt?

*Generalisierte Quantoren-Theorie*

Nach einer kurzen Begrüßung und Scheinchenübergabe für das Finanzamt kam auch schon unmittelbar die erste Frage. Die Prüfungsatmosphäre war sehr angenehm. Prof. Helbig hat ungefähr genauso viel geredet wie ich, sehr ausführliche Fragen gestellt und zu einigen Themen noch zusätzliche Bemerkungen angebracht. Entgegen anderer Protokolle hatte ich auch keine Probleme mit der Art der Fragestellung. Meist war klar, worauf er hinaus wollte, ansonsten hat eine kurze Nachfrage geholfen.