

Interdisziplinäre
fachdidaktische Übung:
Modelle für
Sprachen in der Informatik

SS 2016: Grossmann, Jenko

Modelle für Sprachen in der Informatik

Die Beschreibung orientiert sich am Begriffssystem der Beschreibung natürlicher Sprachen

„*Sprache in der Informatik*“ steht dabei stellvertretend für alle drei Begriffe

- Formale Sprache
- Programmiersprache
- Computersprache

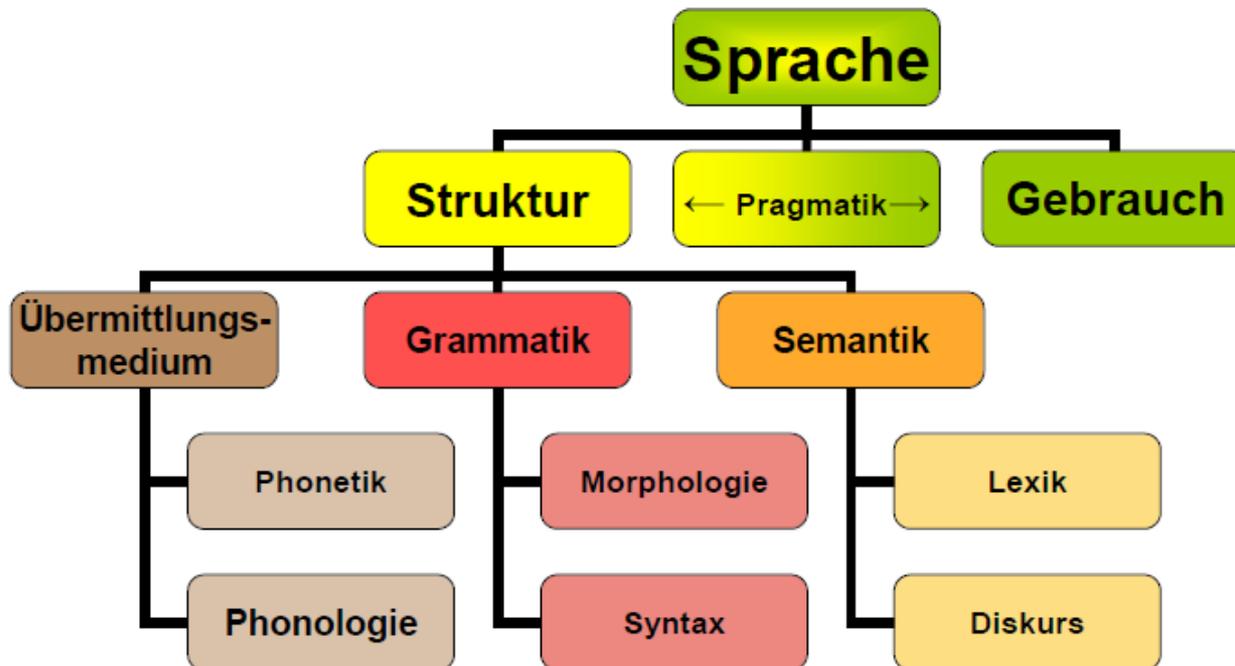
Modelle für Sprachen in der Informatik

Da eine Maschine keine Sprachfähigkeit hat muss der Begriff Sprache von einem abstrakten Modell her entwickelt werden

Die Strukturbeschreibung formaler Sprachen wird wesentlich durch die Aufgabe „Kommunikation mit einer Maschine“ bestimmt

Modelle für Sprachen in der Informatik

Übertragung der Konzepte von natürlichen Sprachen (vgl. Referat 1)



Modelle für Sprachen in der Informatik

Diese Konzepte werden

- unterschiedlich interpretiert
- haben bei den drei Begriffen Formale Sprachen – Programmiersprache – Computersprache unterschiedliches Gewicht

Übermittlungsmedium

Das Übermittlungsmedium ist sehr vielfältig
(Sprache, Zeichen, Bilder, Gesten,...)

Multimedialität

Für die Kommunikation ist dies ein
wesentlicher Aspekt für Computersprachen

- Ikonen, Graphische Eingabesprachen,
Bildbeschreibungssprachen
- Human Computer Interfaces
- Spracheingabe

Übermittlungsmedium

In der Kommunikation mit einer Maschine wird alles in eine Zeichenkette umgewandelt

Für die verschiedene Übermittlungsmedien gibt es eigene Programmiersprachen für die Umwandlung der Eingabe in eine formale Maschinensprache und der Rückübersetzung in das Ausgabemedium

Übermittlungsmedium

Dabei wird eine Folge von Transformationen angewendet, die unterschiedliche Sprachen verwenden



Übermittlungsmedium

Jedes Eingabemedium nutzt Wissen von unterschiedlichen Disziplinen

- Fragen der Phonetik und der Phonologie spielen bei der Spracheingabe eine Rolle
- Andere Disziplinen wie Graphik (statisch oder dynamisch), oder Haptik spielen eine zentrale Rolle beim Übermittlungsmedium

Grammatik, Morphologie

Struktur von Wörtern oder Objekten

Einfachste Definition von „Wort“: eine Zeichenkette

Alle Programmiersprachen haben eine Reihe von reservierten Wörtern

Es gibt auch Regeln zur Konstruktion von neuen Wörtern

Komplexer: Struktur von Objekten

Grammatik, Morphologie

Fragen:

- Beispiele von reservierten Wörtern in der Kommunikation mit dem Computer
- Beispiele von Regeln zur Konstruktion von neuen Wörtern in der Kommunikation mit einem Computer
- Beispiele von grammatischen Kategorien in Programmiersprachen

Grammatik, Syntax

- In Sprachen in der Informatik spielt eine präskriptive Grammatik eine zentrale Rolle
- Regeln zur Bildung von korrekten Ausdrücken
 - Regeln welche die Überprüfung der Korrektheit eines Ausdruckes im Sinne der Regeln erlauben

Grammatik, Syntax

Einige Grammatische Kategorien formaler Sprachen:

- Zeichen (Alphabet, Symbole)
- Zeichenketten (Strings) „Wörter“
- Zahlen
 - Natürliche Zahlen,
 - Dezimalzahlen mit vorgegebenen Kommastellen

Grammatik, Syntax

- Reservierte Worte für die Beschreibung von Abläufen (grundlegende Kontrollstrukturen)
 - Bedingungen: **wenn** *bedingung* **dann** *aktivität*
 - Alternativen: **wenn** *bedingung* **dann** *aktivität*1
sonst *aktivität* 2
 - Wiederholtes Ausführen:
wiederhole *aktivität* **n-mal**
- Aus diesen Konstruktionen können durch Schachtelung komplexe Ausdrücke gebildet werden

Grammatik, Syntax

Satzsyntax: Wir konstruieren einen komplexen Ablauf oder eine Berechnung, ähnlich wie einen Satz in einer natürlichen Sprache in einer generativen Grammatik

Ein Satz wird aus einzelnen Phrasen aufgebaut

Dieser Aufbau kann ausgehend von einem Startsymbol durch einen Baum dargestellt werden (Ableitungsbaum)

Grammatik, Syntax

Andere Darstellungen: Backus – Naur Form

Definition regulärer Ausdrücke

Einfachster Fall: Im Prinzip gibt es nur drei Regeln um Terme zu bilden:

- Alle Elemente Alphabets sind Terme
- Verkettung von Termen: $(x, y) \rightarrow xy$
- Alternative Auswahl von Termen: $x \mid y$

Grammatik, Syntax

- Die Verwendung dieser Syntax ist zweifach
- Konstruktion von syntaktisch richtigen Sätzen durch Menschen oder eine Maschine, die ein ausführbares Programm definieren
 - Analysieren von Sätzen, d.h. die Maschine kann nach diesen Regeln das eingegebene Programm „verstehen“ und die Handlungsanweisungen in eindeutiger Weise ausführen

Grammatik, Syntax

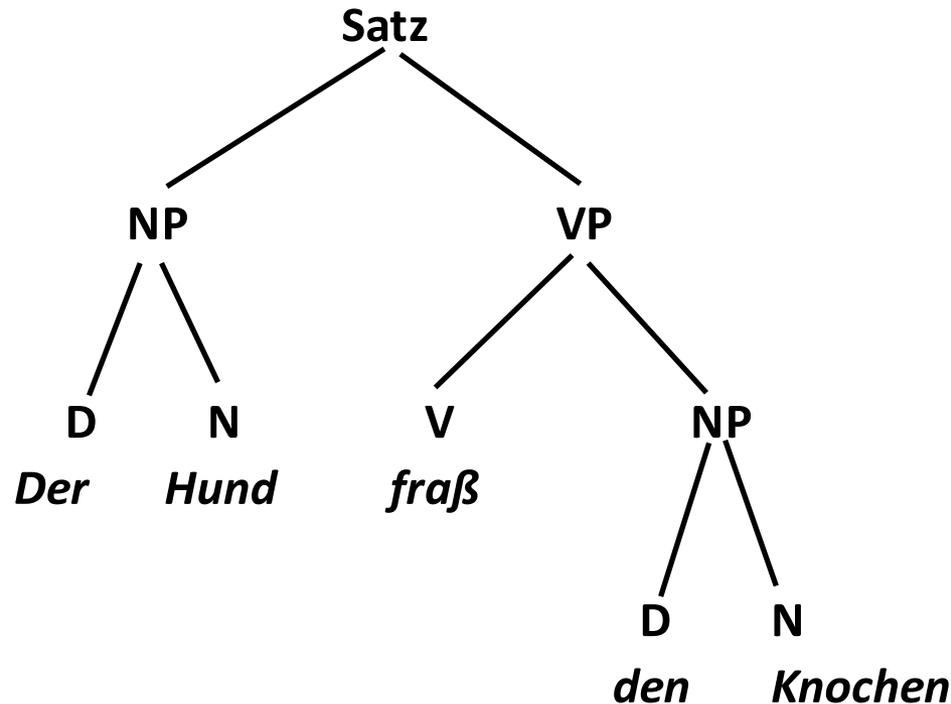
Fragen:

Kann man mit solchen Regel Sätze einer natürlichen Sprache erzeugen?

Kann man mit solchen Regeln ein Programm schreiben, das natürliche Sprachen versteht

Grammatik, Syntax

Beispiel für einen Ableitungsbaum eines Satzes in natürlicher Sprache



Grammatik, Syntax

Syntaktisch korrekte Sätze einer natürlichen Sprachen erlauben im allgemeinen keine eindeutige Rekonstruktion eines Satzes

Beispiele:

Ich sah den Mann mit dem Fernglas

Der Chef lachte über die Forderung der Angestellten nach einer Gehaltserhöhung im Mai

Aufgabe:

Bilde mögliche Ableitungsbäume für diese Sätze

Grammatik, Syntax

Die Rekonstruktion eines Satzes hinsichtlich seiner syntaktischen Struktur (inklusive der Morphologie) ist eine herausfordernde Aufgabe (PART OF SPEECH Tagging)

Stanford Parser online:

<http://nlp.stanford.edu:8080/corenlp/process>

Vollständige Version des Programms:

<http://nlp.stanford.edu/software/tagger.shtml>

Semantik

Was bedeutet Semantik in
Sprachen der Informatik?
Natürlichen Sprachen?

Informatik: Semantik beschäftigt sich, wie
bei natürlichen Sprachen mit dem
Aufbau und der Bedeutung der durch die
Grammatik gebildeten Ausdrücke

Beschreibung oft in natürlicher Sprache

Semantik

Statische Semantik

Regeln, die in der Sprache erfüllt sein müssen und formal überprüft werden können

Regeln entsprechen unserer Vorstellung von „semantisch sinnvollen“ Ausdrücken

Beispiel: Addition nur für Zahlen sinnvoll

Mojca \oplus Wilfried = ?

Semantik

Lösung dieses Problems durch
Typisierung:

Addition von ganzen Zahlen

Addition von Dezimalzahlen
(Rechengenauigkeit)

Addition von Vektoren

Semantik

Dynamische Semantik

Semantik, die sich aus der Abarbeitung der Befehle ergibt

Meist in einem eingeschränkten Kontext behandelt: Ist Programm **korrekt**?

Korrekt bedeutet, dass das Programm seine Analyse in endlicher Zeit mit einer Antwort beendet

Semantik

Die Semantik einer natürlichen Sprache ist
aber weit komplexer

Welcher Aufwand ist notwendig um diese
Semantik nachzubilden?

Beispiel: ConceptNet zum Verstehen von
„Common Sense“

<http://conceptnet5.media.mit.edu/>

Spezialbereich: Sentiment Analyse