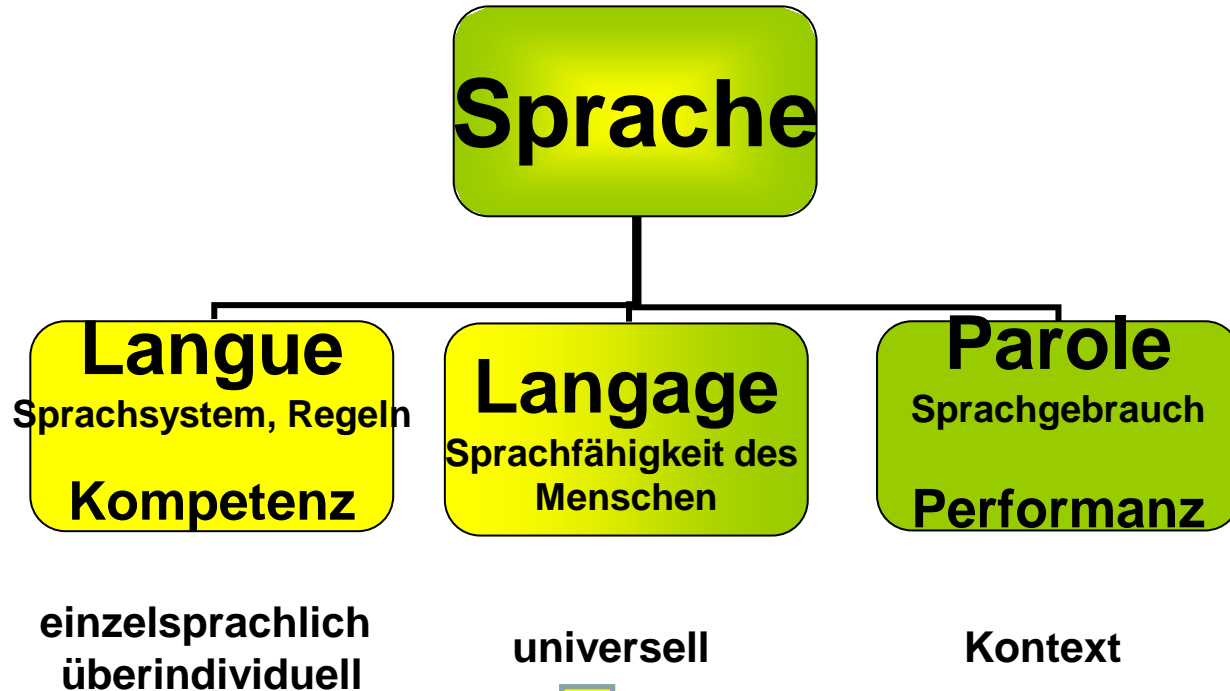


Interdisziplinäre
fachdidaktische Übung:
Sprachmodelle –
Formale Sprachen

SS 2013: Grossmann, Jenko

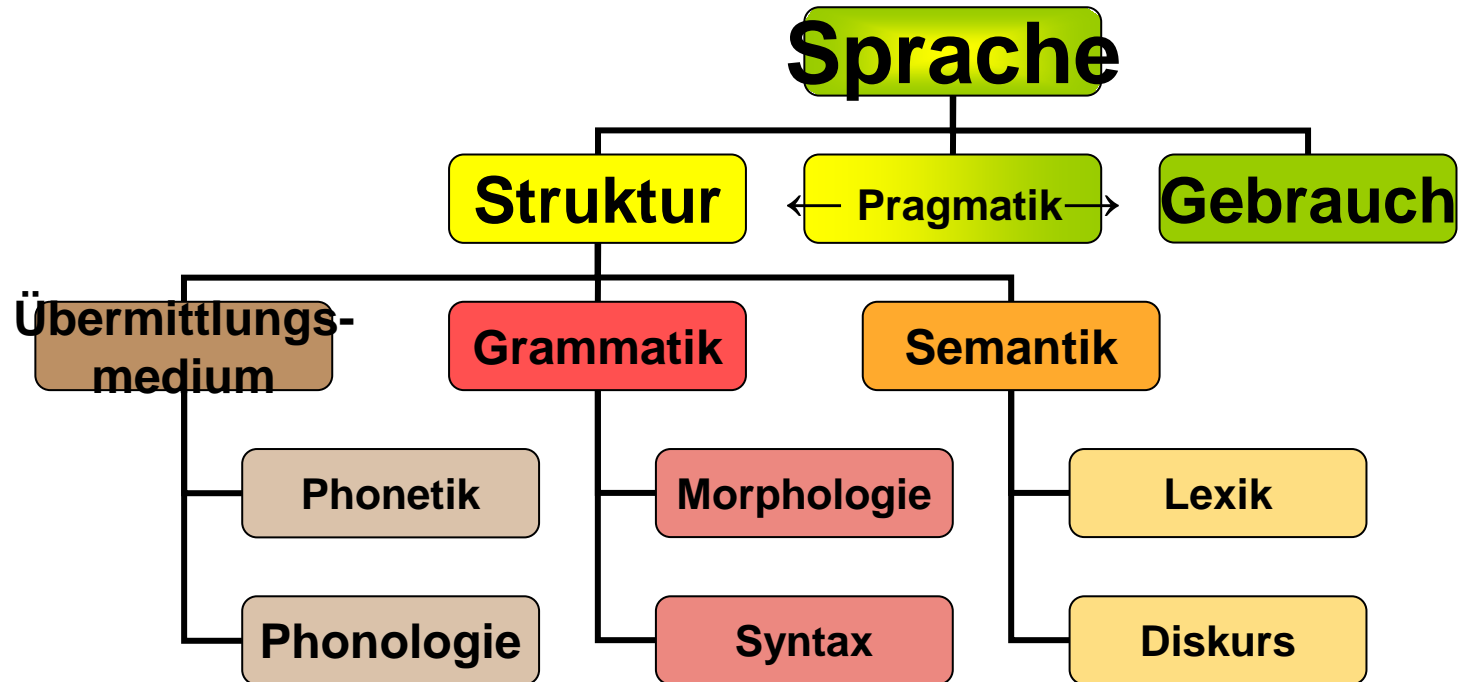
Modell Natürliche Sprachen



Hat eine Maschine Sprachfähigkeit ?

Sprachmodelle-Formale Sprachen

Modell Natürliche Sprachen



Vergleich natürliche Sprache – Formale Sprache

Wir verwenden im Folgenden für den
Vergleich den Begriff formale Sprache

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

Welche Konzepte natürlicher Sprachen
lassen sich bei formalen Sprachen
anwenden?

Struktur - Formale Sprache

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

Struktur - Formale Sprache

Beispiel:

Natürliche Sprache:

Frage: „Wird am 15.3.2013 um 13:00 in
Wien 9, Währinger Strasse 29 die Sonne
scheinen?“

Antwort: „nein“

Wie kann eine Maschine diesen Dialog
abbilden?

Struktur - Formale Sprache

Maschine erhält Instruktionen = Programm:

Eingabe:

Datum:15.3.2013; ***Uhrzeit*** 13:00,

Ort: Wien 9, Währinger Strasse 29,

Bewölkung: bedeckt; ***Sonnenaufgang*** 6:07

Instruktion: *Wenn Wetter wolkenlos und Uhrzeit nach Sonnenaufgang dann antworte mit „ja“ ansonsten mit „nein“*

Struktur - Formale Sprache

Die Maschine entscheidet also nur ob sie diese Kombination von Eingabe und Instruktion als eine für sie korrekte Zeichenkette erkennt

Nachdem sie das erkannt hat reagiert sie nach einer formalen Regel (Instruktion), im Prinzip nur mit „korrekt“ oder „nicht korrekt“

Struktur - Übermittlungsmedium

Das Übermittlungsmedium ist heute für formale Sprachen sehr vielfältig (Sprache, Zeichen, Bilder, Gesten,...)

Im Prinzip wird aber jede Eingabe in eine Zeichenkette umgewandelt

Fragen des Übermittlungsmediums stellen sich also nur bei der Umwandlung der Eingabe in Maschinensprache und der Rückübersetzung in das Ausgabemedium

Struktur - Übermittlungsmedium

Fragen der Phonetik und der Phonologie sind also nicht direkt mit einer formalen Sprache verbunden und behandeln nur einen möglichen Teilaspekt der Kommunikation und des Selbstausdruckes

Andere Disziplinen wie Graphik (statisch oder dynamisch), oder Haptik spielen eine zentrale Rolle beim Übermittlungsmedium

Multimedialität

Struktur - Grammatik

Grammatik konzentriert sich im wesentlichen auf die Syntax und ist präskriptiv (Regeln) und deskriptiv (Beschreibung aller Konstruktionen)

Morphologie als Untersuchung der Struktur von Wörtern spielt praktisch keine Rolle

Statt dessen betrachten wir ein sehr einfaches und allgemeines Syntaxmodell

Struktur - Grammatik

Wortsyntax: Wir kennen in formalen Sprachen grammatische Kategorien, die zur Beschreibung von Kommunikationsabläufen wesentlich sind

Diese Kategorien orientieren sich an logischen und mathematischen Begriffen, die für den algorithmischen Ablauf und für die Berechnung notwendig sind

Welche Kategorien kennen Sie?

Struktur - Grammatik

Einige Grammatische Kategorien formaler Sprachen:

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

Struktur - Grammatik

- Reservierte Worte für die Beschreibung von Abläufen
 - Bedingungen: **wenn** *bedingung* **dann** *aktivität*
 - Alternativen: **wenn** *bedingung* **dann** *aktivität1*
sonst *aktivität 2*
 - Wiederholtes Ausführen:
wiederhole *aktivität* **n-mal**
- Aus diesen Konstruktionen können durch Schachtelung komplexe Ausdrücke gebildet werden

Struktur - Grammatik

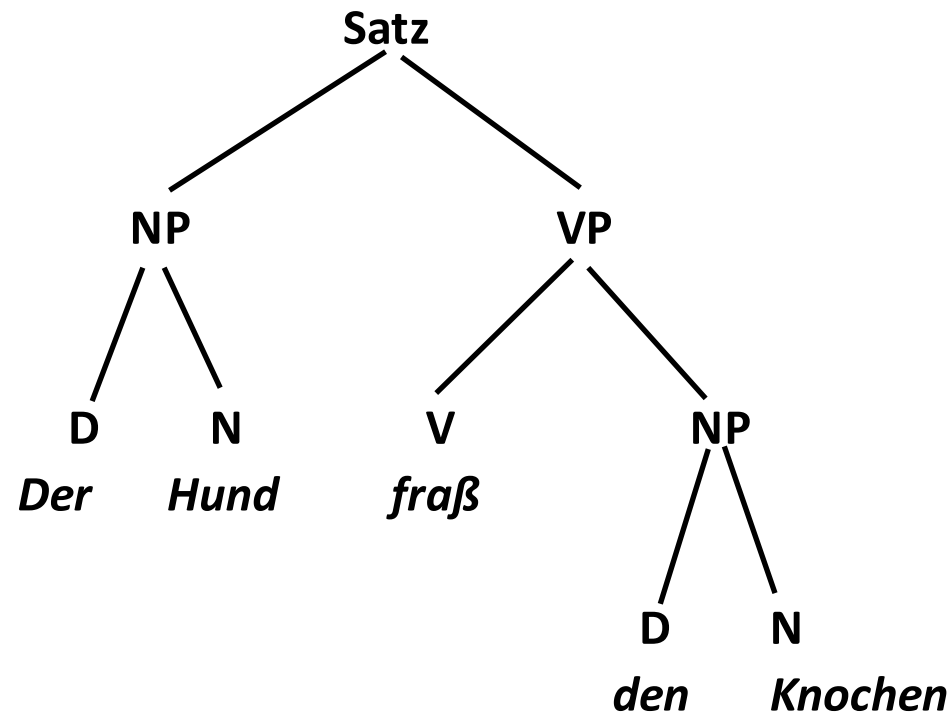
Satzsyntax: Wir konstruieren einen komplexen Ablauf oder eine Berechnung, ähnlich wie einen Satz in einer natürlichen Sprache in einer generativen Grammatik

Ein Satz wird aus einzelnen Phrasen aufgebaut

Dieser Aufbau kann durch einen Baum dargestellt werden (Ableitungsbaum)

Struktur - Grammatik

Beispiel für einen Ableitungsbaum



Struktur - Grammatik

Ähnlich definieren wir eine formale Sprache

- Eine endliche Menge von Symbolen, die man Alphabet bezeichnet. Sie bilden die Terminalen Symbole
- Eine endliche Menge von Nicht-terminalen Symbolen
- Ein Startsymbol
- Produktionsregeln: Ausdruck \rightarrow Ausdruck

Ausdrücke werden als Terme bezeichnet

Struktur - Grammatik

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: $x|y$

Wesentlich für die Charakterisierung von formalen Sprachen sind Produktionsregeln
Jede Produktion wird **Wort** genannt

Struktur - Grammatik

- 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 ausführen

Struktur - Grammatikhierarchien

Man kann sich nun die Frage stellen, welche Arten von Grammatiken von einer geeignet konstruierten Maschine ausgeführt werden können

Diese Frage führt zur Definition von Grammatikhierarchien

Struktur – Grammatikhierarchien

Allgemeinste Form sind Typ 0 Grammatiken

- Keine Einschränkung an Produktionsregeln
- Sie heißen rekursiv aufzählbare Sprachen
- Man kann ein Modell einer Maschine entwickeln, die bei jedem in der Sprache gebildeten Wort mit „richtig“ antwortet
- Allerdings kann vorkommen, dass bei nicht in der Sprache gebildet Ausdrücken keine Entscheidung getroffen wird

Struktur – Grammatikhierarchien

Typ 1 Grammatiken heißen kontextsensitive Grammatiken

- Produktionsregeln ersetzen ein nicht terminales Symbol der linken Seite durch neue Symbol, eventuell in Abhängigkeit der Umgebung des nicht terminalen Symbols
- Bei diesen Sprachen kann man entscheiden ob ein Wort zu der Sprache gehört oder nicht

Struktur – Grammatikhierarchien

Typ 2 Grammatiken heißen kontextfreie Grammatiken

- Produktionsregeln enthalten auf der linken Seite nur nicht terminale Symbole, Ersetzung hängt also nicht vom Kontext ab
- Grundlage der meisten Programmiersprachen
- Man kann entscheiden ob ein Wort zur Sprache gehört und ob es nur endlich viele Worte in der Sprache gibt

Struktur – Grammatikhierarchien

Typ 3 Grammatiken heißen reguläre Grammatiken

- Produktionsregeln enthalten auf der rechten Seite maximal ein terminales Symbol und ein nicht terminales Symbol, das immer auf der linken Seite steht
- Man kann entscheiden ob zwei Sprachen äquivalent sind
- Realisierung durch endlichen Automaten

Struktur – Beschreibung von Grammatiken

Grammatiken werden meist durch eine der folgenden beiden Methoden beschrieben

- Angabe der Produktionsregeln (Backus – Naur Form)
- Angabe der Struktur regulärer Ausdrücke

Struktur – Semantik

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

Man unterscheidet statische und dynamische Semantik

Beschreibung der Semantik oft in natürlicher Sprache

Struktur – 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+Wilfried = ?

Struktur – Semantik

Lösung dieses Problems durch

Typisierung:

Addition von ganzen Zahlen ist anders
zu verstehen als Addition von
Dezimalzahlen (Rechengenauigkeit)

Nichtbeachtung kann zu schweren
Fehlern führen

Struktur – 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 der Antwort „richtig“ oder „falsch“ beendet

Struktur – Semantik

Beispiel für ein nicht korrektes Programm:

$x = 1$

solange $x > 0$ **führe durch**

addiere zu x den Wert 1

Diese Anweisungen sind syntaktisch korrekt, aber im Sinne der dynamischen Semantik nicht korrekt

Struktur – Semantik Web

Zur Darstellung der Semantik wird im ***Semantic Web*** das Modell des ***Resource Description Framework*** verwendet (RDF):

Jede Semantik ist von der Form:
(Subjekt, Prädikat, Objekt)

Sprache – Gebrauch

Wir können im Gebrauch dieses einfachen Modells unterschiedliche Szenarien verwenden

Sprachparadigmen

Zwei wichtige Paradigmen:

- Imperative Sprachen: Handlungsanweisung, Folge von Befehlen
- Deklarative Sprachen: Spezifikation was berechnet werden soll