Computerlinguistik

E06: Satzebene

IDH INSTITUT FÜR DIGITAL HUMANITIES UNIVERSITÄT ZU KÖLN

Syntax

- Begriffe: Grammatik – Satz – Phrase
- Konstituenten – Ermittlung und Klassifikation
- Phrasenstrukturen
- Dependenzstrukturen
- Referat: Grammatikprüfung
- Referat: Textgenerierung

Syntax: Grundbegriffe

- **Prinzip der doppelten Artikulation:**
  - Inventar sprachlicher Einheiten (phonologische Ebene)
  - Verknüpfungsregeln (syntaktische Ebene)

Ebene	Laut	Wort	Satz
Ausdruck	**Phonetik / Phonologie**	**Morphologie**	**Syntax**
Inhalt	**ø**	**Wortsemantik**	**Satzsemantik**

- **Satzhaftigkeit: Vollständigkeit + Grammatikalität**
- **Strukturabhängige syntaktische Regularitäten**

(6) a.

b.

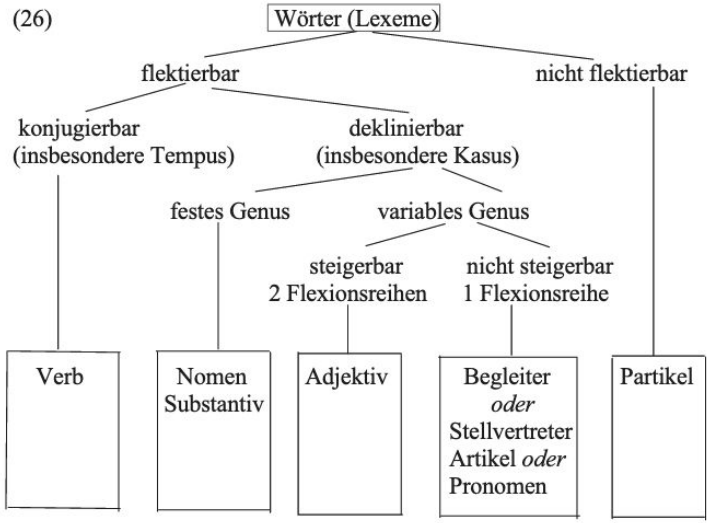
Konstituenten: Ermittlung

- Ersetzungsprobe
- Pronominalisierungstest
- Fragetest
- Permutationstest (Verschiebeprobe)
- Tilgungstest

Konstituenten: Klassifikation

- **Wortartenklassifikation**
  - morphosyntaktische Eigenschaften (Merkmale)
  - syntaktische Eigenschaften (Distribution)

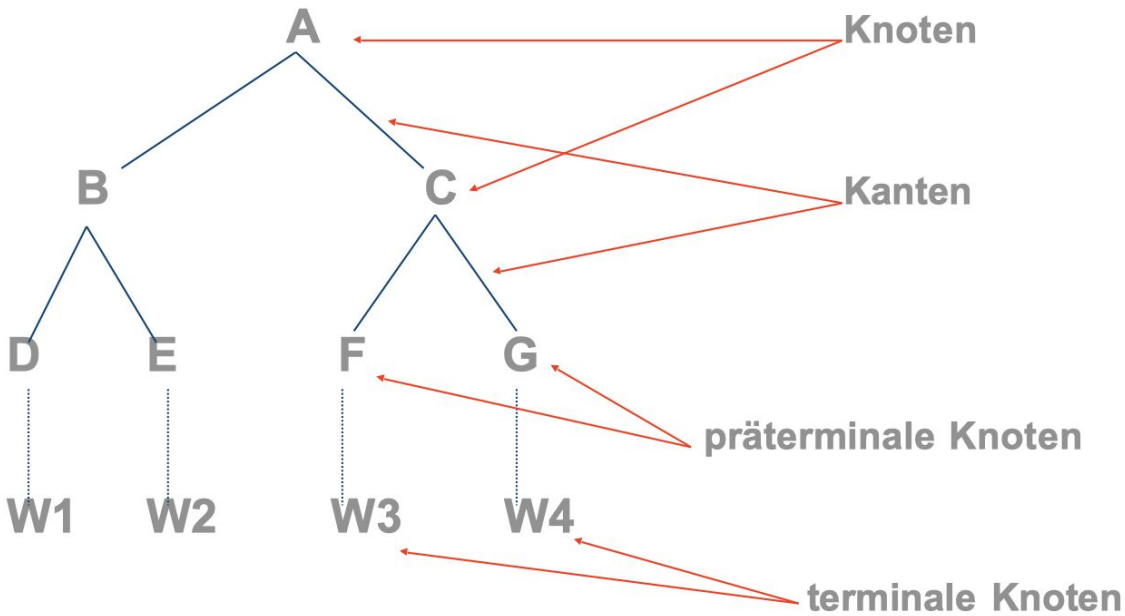
- **Klassifikation von Phrasen**
  - Nominalphrasen
  - Verbalphrasen
  - Adjektivphrasen
  - Präpositionalphrasen



aus Ramers (2007:24)

Phrasenstrukturbäume

Baumdarstellung:



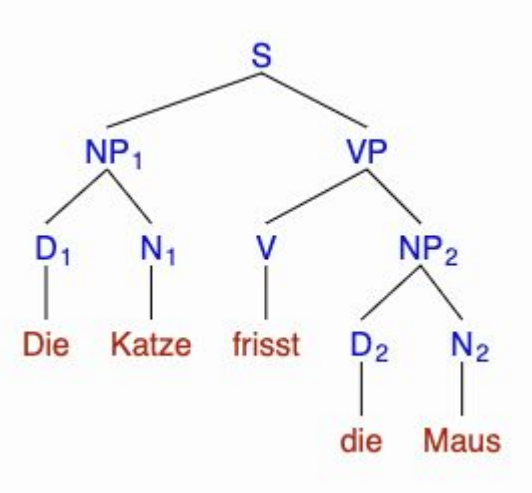
Klammerstrukturdarstellung: [A[B[D[W1] E[W2]] C[F[W3] G[W4]]]]

Phrasenstrukturregeln - Beispiele

- **Syntaktische Regeln:**
  - S → NP VP
  - NP → D N
  - VP → V NP
- **Lexikalische Regeln:**
  - D → die
  - N → katze
  - N → maus
  - V → frisst

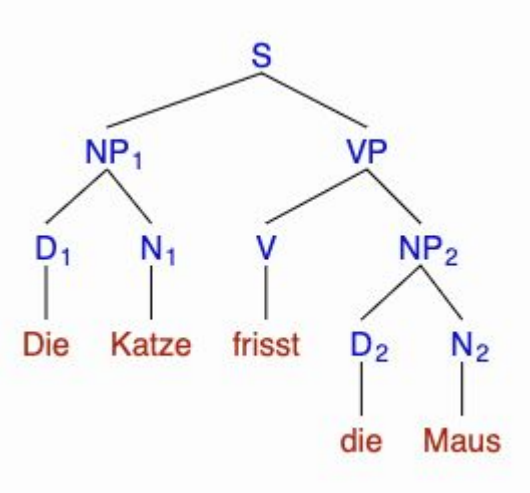
Phrasenstrukturregeln - Beispiele

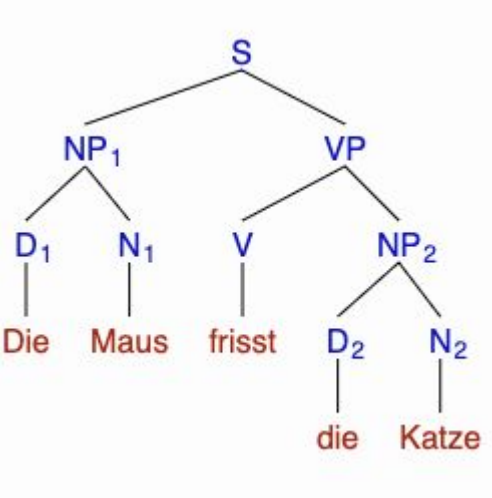
- **Syntaktische Regeln:**
  - S → NP VP
  - NP → D N
  - VP → V NP
- **Lexikalische Regeln:**
  - D → die
  - N → katze
  - N → maus
  - V → frisst



Phrasenstrukturregeln - Beispiele

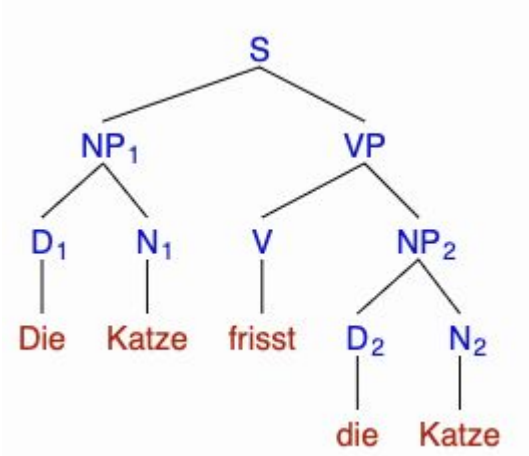
- **Syntaktische Regeln:**
  - S → NP VP
  - NP → D N
  - VP → V NP
- **Lexikalische Regeln:**
  - D → die
  - N → katze
  - N → maus
  - V → frisst

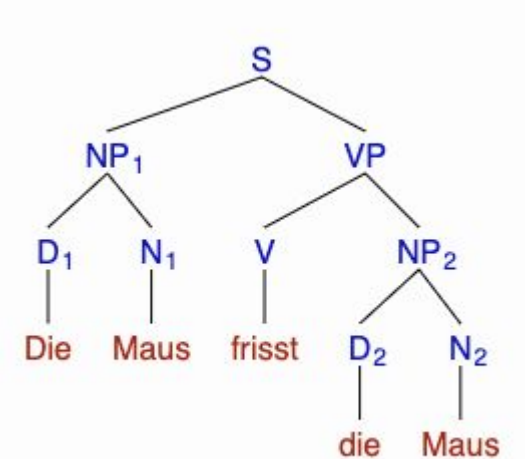




Phrasenstrukturregeln - Beispiele

- **Syntaktische Regeln:**
  - S → NP VP
  - NP → D N
  - VP → V NP
- **Lexikalische Regeln:**
  - D → die
  - N → Katze
  - N → Maus
  - V → frisst





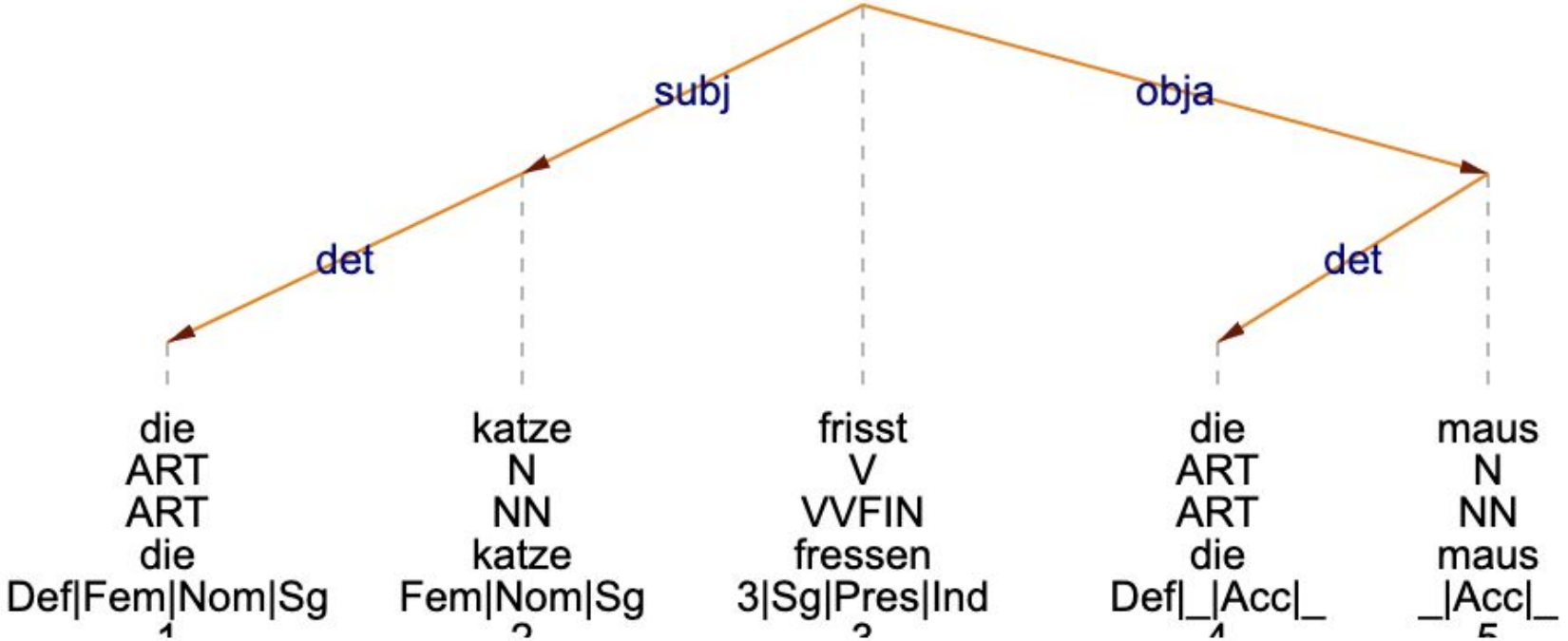
Phrasenstrukturregeln - Schema

- Kontextfreie Regeln
  - Links muss genau ein Kategorialsymbol stehen
  - Rechte muss mindestens ein Kategorial- oder ein lexikalisches Symbol stehen
- Rekursive Strukturen sind zulässig (z.B. NP → D N NP)
- Subkategorisierungsrahmen für Verben muss eingehalten werden
  - wohnen[PP] → “Das Kind wohnt in Düsseldorf.” vs. ”*Das Kind wohnt.”
  - schenken[NP, NP] → “Die Mutter schenkt dem Kind den Ball.” vs. “*Die Mutter schenkt dem Kind.”
  - rufen[NP] → “Der Mann ruft die Frau.” vs. “*Der Mann ruft die Frau den Hund.”

Dependenzstrukturen - Schema

- Valenzgrammatik (Tesniere 1959)
  - Dependenzrelationen (Stemmata): Regens - Dependens (1:n)
  - Verb als Zentralnexus des Satzes mit
    - Aktanten (Akteure des Geschehens) und
    - Angaben (Umstände des Geschehens)
- Typisierte Dependenzen
  - Relationen zwischen Knoten werden durch Kanten ausgedrückt
  - Typ der Relation ist Label der Kante

Dependenzstrukturen - Beispiel



Das CONLL-U Datenformat

- CONLL (Conference on Computational Natural Language Learning) ist ein Standard-Format für die Annotation von natürlicher Sprache.
- Tabellarisches Format mit Zeilen für einzelne Tokens und Spalten mit zusätzlichen Informationen:

ID	Form	Lemma	POS-Tag	Features	Head	Relation
1	Peter	Peter	NE	Masc|Nom|Sg	2	subj
2	sitzt	sitzen	VVFIN	3|Sg|Pres|Ind	0	root
3	.	.	$.	_	0	root

Literatur / Hausaufgabe

- **Zur Nachbereitung**:
  - Lesen Sie: Ramers (2007): Kapitel 1, 2 und 4 (S. 11-55 und 77-88)
  - Bearbeiten Sie die schriftlichen Aufgaben in ILIAS
  - Testen Sie den ParZu-Parser https://pub.cl.uzh.ch/demo/parzu/
- Die Texte (bzw. Links) finden Sie im Ilias-Seminarordner.