

Tutorium der DGfS-Sektion Computerlinguistik

KORPUSLINGUISTIK

MIT ONLINE -

RESSOURCEN

EINE INTERAKTIVE

EINFÜHRUNG

FÜR LINGUISTEN



STTS

Stefanie Dipper
Stefan Evert
Heike Zinsmeister

München, 28.1.2011

Das Stuttgart-Tübingen Tagset

- Anne Schiller, Simone Teufel, Christine Thielen, Christine Stöckert (1999)
- Wortarten-Analyse für die manuelle (und automatische) Annotation von deutschen Textkorpora
- orientiert am TEI Standard (1991)
- Vereinfachung: Leerzeichen = Wortgrenze
 - es werden zunächst keine Mehrwortlexeme berücksichtigt
- 11 Hauptwortarten, spezifiziert in 54 Unterwortarten-Tags
 - 48 Unterwortartentags (Bsp. "ADJA")
 - 6 Zusatztags: Interpunktion, Fremdsprachliches, Nicht-Wörter, Truncs
- "Großes" und "kleines STTS"

Wortartenklassen

- **Morphologische Kriterien**
 - flektierbar oder nicht; Art der Flexion
 - **Distributionelle Kriterien**
 - Wortartenkontexte / Positionen im Satz, in denen ein Element typischerweise auftritt
 - eigene Klassen für wichtige Funktionswörter
 - **Semantische Kriterien**
 - teilweise Bildung von Unterklassen durch Bedeutungsunterschiede
- Keines der Kriterien gilt absolut. Es gibt Interaktionen und Ausnahmen.

STTS-Hauptwortarten

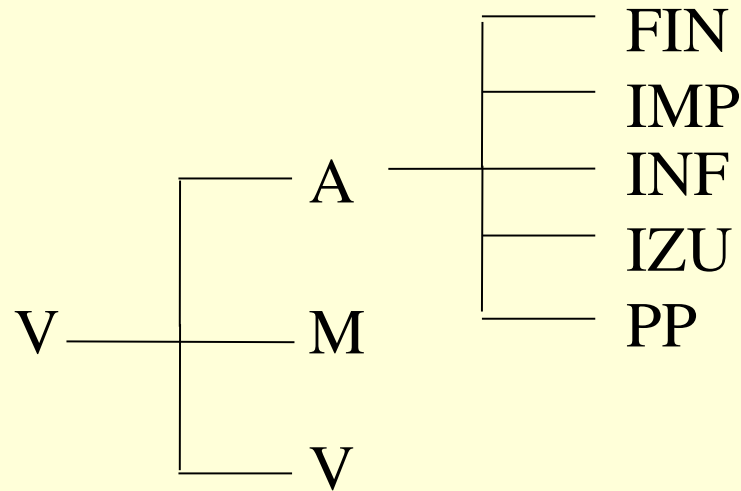
1. Nomina (N)
2. Verben (V)
3. Artikel (ART)
4. Adjektive (ADJ)
5. Pronomina (P)
6. Kardinalzahlen (CARD)
7. Adverbien (ADV)
8. Konjunktionen (KO)
9. Adpositionen (AP)
10. Interjektionen (ITJ)
11. Partikeln (PTK)

STTS-Unterwortarten

- Aufbau der Labels
 - hierarchisch von links nach rechts: vom Allgemeinen zum Spezifischeren
- Bsp.: Pronomen (P)
- (1) Diese Lösung klappt.
- (2) Diese klappt nicht.
 - ad (1): Pronomen **demonstrativ** **attribuierend**,
PDAT
 - ad (2): Pronomen **demonstrativ** **substituierend**,
PDS

Unterspezifizierte Anfragen

- Bsp.: Verben (V)



- Weitgehend: gleiche "Slot"-Längen
- Bsp-Anfragen:
 - [pos="V.FIN"]
 - [pos="VA.*"]

Weitere Unterwortarten

Nomen	NN, NE
Adjektive	ADJA, ADJD
Adpositionen	APPR, APPO, APZR, APPRART
Konjunktionen	KON, KOUS, KOUJ, KOKOM
Pronomen	PPER, PPOS(S AT), PREL.*, PW.*, PAV
Satzzeichen	\$., \$,, \$(

Referenzen

- Anne Schiller, Simone Teufel, Christine Stöckert und Christine Thielen (1999): *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. Technical report, University of Stuttgart and University of Tübingen. <http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-1999.pdf>
- TEI (1991): TEI AI1W2: *List of Common Morphological Features for Inclusion in TEI Starter Set of Grammatical Annotation Tags*.