

KORPUSLINGUISTIK

MIT ONLINE -

RESSOURCEN

EINE INTERAKTIVE EINFÜHRUNG FÜR LINGUISTEN

STTS

Stefan Evert
Stefanie Dipper
Heike Zinsmeister

Berlin, 23.2.2010

Das Stuttgart-Tübingen Tagset

- Anne Schiller, Simone Teufel, Christine Thielen, Christine Stöckert (1999)
- Wortarten-Analyse für die manuelle (und automatische) Annotation von deutschen Textkorpora
- Vereinfachung: Leerzeichen = Wortgrenze
 - es werden zunächst keine Mehrwortlexeme berücksichtigt
- Analog zum Standard der Text Encoding Initiative (TEI)
- 11 Hauptwortarten, spezifiziert in 54 Wortarten-Tags
 - 48 Wortarten- & 6 Zusatztags:
 - Interpunktion, Fremdsprachliches, Nicht-Wörter, Truncs)

Wortartenklassen

- **Morphologische Kriterien**
 - flektierbar oder nicht, Art der Flektion
 - wortartenspezifische Affigierung in der Wortbildung
- **Semantische Kriterien**
 - teilweise Bildung von Unterklassen durch Bedeutungsunterschiede
- **Distributionelle Kriterien**
 - Wortartenkontexte / Positionen im Satz, in denen ein Element typischerweise auftritt
- **Keines der Kriterien gilt absolut. Es gibt Interaktionen und Ausnahmen.**

STTS-Hauptwortarten

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

STTS-Detailwortarten

- 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