

DDDTS — ein POS-Tagset für historische Daten

Stefanie Dipper

Sprachwissenschaftliches Institut
Ruhr-Universität Bochum

24. September 2012

CLARIN-D-Workshop
Das STTS-Tagset für Wortartentagging:
Stand und Perspektiven

Outline

- 1 Hintergrund
- 2 Unterschiede zum STTS

Projekt-Hintergrund

- DFG-Projekte *Referenzkorpus Altdeutsch (750–1050)*, *Referenzkorpus Mittelhochdeutsch (1050–1350)*
- Ziel: ausgewogenes historisches Korpus des Deutschen, annotiert mit
 - Morphologie
 - POS-Tags
 - historische Lemma-Formen
 - (historische Normalform)

Outline

- 1 Hintergrund
- 2 Unterschiede zum STTS

POS-Annotation

- ursprünglich: STTS mit Erweiterungen
- jetzt: an STTS angelehnt, mit vielen Abweichungen

POS-Annotation

- ursprünglich: STTS mit Erweiterungen
- jetzt: an STTS angelehnt, mit vielen Abweichungen
 - (o) lemma- vs. belegspezifische Annotation
 - (i) zusätzliche Unterscheidungen
 - (ii) spezifische Anforderungen für historische Daten
 - (iii) Systematisierung ausgewählter Tagnamen

(o) Zweifache Annotation: lemma- vs. belegspezifisch

- z.B. Adjektive
 - lemmaspez: ADJ
 - belegspez: ADJA, ADJD, (ADV), ...

(o) Zweifache Annotation: lemma- vs. belegspezifisch

- z.B. Adjektive
 - lemmaspez: ADJ
 - belegspez: ADJA, ADJD, (ADV), ...
- Nomen: *Hans der **Bauer***
 - lemmaspez: NN
 - belegspez: NE

(o) Zweifache Annotation: lemma- vs. belegspezifisch

- z.B. Adjektive
 - lemmaspez: ADJ
 - belegspez: ADJA, ADJD, (ADV), ...
- Nomen: *Hans der **Bauer***
 - lemmaspez: NN
 - belegspez: NE
- Nomen: *man*
 - lemmaspez: NN
 - belegspez: PIS

(i) Zusätzliche Unterscheidungen

Z.B.

- STTS: ART für Artikel allgemein
- DDDTS: DD (definit) vs. DI (indefinit) (s.u.)

(ii) Spezifische Anforderungen für historische Daten

Z.B.

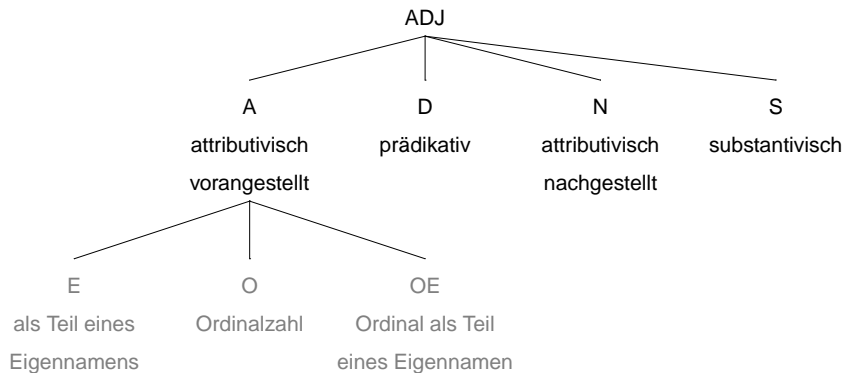
- STTS: attributive (\approx flektierte) Adjektive vs. Rest
- DDDTS: zusätzlich prädikativ (z.B. *snell*, 'schnell') vs. adverbial (*snello* (ahd), *snelle* (mhd))

(ii) Spezifische Anforderungen für historische Daten

Z.B.

- STTS: attributive (\approx flektierte) Adjektive vs. Rest
- DDDTS: zusätzlich prädikativ (z.B. *snell*, 'schnell') vs. adverbial (*snello* (ahd), *snelle* (mhd))
- STTS: Groß-/Kleinschreibung relevant: *er ist der **Größte**/NN* vs. *sie werden als **letzte**/ADJA geheuert*
- DDDTS: immer ADJS

(ii) Spezifische Anforderungen: Adjektive



(iii) Systematisierung ausgewählter Tagnamen

- STTS: eigenes Tag ART wegen besonderer Stellung des Artikel im Gegensatz zu Determinern im Allgemeinen
- DDDTS: Artikel als eine Unterklasse der Determiner-Ausdrücke
- STTS: attributive Verwendung (\approx Determiner) vs. substantivische Verwendung (\approx Pronomen)
- DDDTS: nur Personalpronomomen als Pronomen

(iii) Systematisierung: Determiner und Pronomen

