

Korpuslinguistik und Statistik

(050370)

Fabian Barteld, M.A.

Fabian.Barteld@ruhr-uni-bochum.de

6. Sitzung

Annis – ANNotation of Information Structure

- ▶ Programm zum Durchsuchen und von annotierten Texten
- ▶ wird für Falko verwendet
 - ▶ weiterführende Informationen zu Annis
<http://www.sfb632.uni-potsdam.de/d1/annis/>

Untersuchungsablauf

1. Fragestellung
2. Korpusauswahl/ Erstellung
3. relevante Belege suchen (Suchabfrage)
4. Auswertung der Belege (Statistik)

Fragestellung auswählen

- ▶ Stellen Sie sich gegenseitig in Ihrer Gruppe die einzelnen Fragestellungen vor
(je 5 min)
- ▶ Wählen Sie eine Fragestellung aus, die Sie gemeinsam bearbeiten möchten
(5 min)

Abgrenzung: Korpus – Belegsammlung

Textsammlung
unsystematisch

Korpus
systematisch + vollständig

Belegsammlung
unvollständig

Suchanfrage

Ziel: Korpus → Belegsammlung

Von der Fragestellung zur Suchanfrage

- ▶ was möchten Sie finden? → beispielhafte Aufzählung (Extension)
- ▶ wie lassen sich diese Gegenstände beschreiben? (Intension)
- ▶ welche linguistischen Informationen (Annotation) benötigen Sie?
- ▶ welche weiteren Informationen (Metadaten) benötigen Sie?

→ wie lässt sich die Beschreibung mithilfe der Annotation und der Suchsyntax realisieren?

Gruppenarbeit

Erstellen Sie eine Suchanfrage zur Ihrer Fragestellung!

Versuchen Sie die Ergebnisse zu speichern, so dass mit ihnen weitergearbeitet werden kann.

Dokumentieren Sie Ihre Fragestellung und die Suchanfrage.

(Zeit: 30 min)

Literatur

Abschnitte zur Suche bei Perkuhn, Keibel und Kupietz (2012, S. 22–38)

Vertiefung:

- ▶ Präzision, Recall (Perkuhn u. a., 2012, 39ff)
- ▶ reguläre Ausdrücke:
 - ▶ erster Einstieg: Bubenhofer, 2006–2011
<http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=regexp.html>
 - ▶ weiterführend:
Friedl, J. E. F. (2003). *Reguläre Ausdrücke* (2. Auflage). Beijing u. a.: O'Reilly

Ausblick

- ▶ Wiederholung: Werkzeugkasten Korpuslinguistik
- ▶ Statistik-Einführung
- ▶ Aufbereiten von Belegensammlungen für statistische Auswertungen