

Computergestützte Untersuchung von Textkorpora zur Erhöhung der fachlichen Medienkompetenz

Frank Richter

Goethe Universität, Frankfurt a.M.

8. Oktober 2015

Hintergrund: IEAS, Abteilung Linguistik

- Das Fachgebiet: Teilbereiche der Grammtiktheorie
 - ▶ Phonetik/Phonologie
 - ▶ Morphologie
 - ▶ Syntax
 - ▶ Semantik
 - ▶ Pragmatik
- Paradigmenwechsel in der Linguistik
 - ▶ Experimentelle Methoden
 - ▶ Statistik auf großen Datenmengen
 - ▶ Computerlinguistik
 - ▶ Korpuslinguistik

www.english-linguistics.de

Hintergrund: IEAS, Abteilung Linguistik

- Das Fachgebiet: Teilbereiche der Grammtiktheorie
 - ▶ Phonetik/Phonologie
 - ▶ Morphologie
 - ▶ Syntax
 - ▶ Semantik
 - ▶ Pragmatik
- Paradigmenwechsel in der Linguistik
 - ▶ Experimentelle Methoden
 - ▶ Statistik auf großen Datenmengen
 - ▶ Computerlinguistik
 - ▶ Korpuslinguistik

www.english-linguistics.de

Was sind Korpora?

- Sprachdatensammlungen

- ▶ Texte
- ▶ Audio
- ▶ Video

- mit bestimmten Eigenschaften

- ▶ maschinenlesbar
- ▶ erzeugt in einer natürlichen Kommunikationsumgebung
- ▶ repräsentativ hinsichtlich der Sprachvarietät
- ▶ ausgewogen hinsichtlich der Varietät

Was sind Korpora?

- Sprachdatensammlungen
 - ▶ Texte
 - ▶ Audio
 - ▶ Video
- mit bestimmten Eigenschaften
 - ▶ maschinenlesbar
 - ▶ erzeugt in einer natürlichen Kommunikationsumgebung
 - ▶ repräsentativ hinsichtlich der Sprachvarietät
 - ▶ ausgewogen hinsichtlich der Varietät

Entwicklung

- Brown Corpus (Kucera and Francis, 1967)
AE, ca. 1.000.000 Wörter, 15 Textkategorien
- British National Corpus (1991-94)
BE, 100 Millionen Wörter, schriftlich und mündlich, PoS Tags
- CHILDES
Spracherwerbsdaten
- Switchboard
Telefongespräche
- Penn Treebank Project
syntaktisch annotierte Daten
- ...
- Korpora aus (teil-) automatisch gesammelten Internetdaten
mehrere Milliarden Wörter

Textkorpora im Fremdsprachenunterricht

- Methodenorientierte Seminare:
 - ▶ Statistik, Programmierung
 - ▶ Ressourcen
 - ▶ Anwendungsbeispiele
- Themenorientierte Seminare: Negative polarity items
- Zugrunde liegendes Sprach-Lehr/Lern-Konzept
 - ▶ authentischer Sprachgebrauch
 - ▶ induktives Lernen
 - ▶ Lernerautonomie
 - ▶ → korpus-basiertes, daten-gestütztes Lernen ist kompatibel mit Lernzielen in Bildungsplänen

Ressourcen: Ein Beispiel

BYU-Korpora, <http://corpus.byu.edu>

- Einheitliche Nutzeroberfläche für:
 - ▶ British National Corpus,
 - ▶ Corpus of Contemporary American English,
 - ▶ Google Books mit Google n-grams,
 - ▶ Wikipedia Corpus,
 - ▶ Corpus of Global Web-Based English,
 - ▶ ...
- Youtube-Videos, z.B.
<https://www.youtube.com/channel/UCL39Kisoscb82UsJPhWS0Yw>
- Freier Zugang, beschränkte Anzahl von Abfragen
- oder: Zugang mit Anmeldung

Danke für Ihr Interesse!