



Universität Zürich

Deutsches Seminar

Schönberggasse 9

CH-8001 Zürich

Tel. ++41 1 634 25 61

Fax ++41 1 634 49 05

spitzmueller@access.unizh.ch

www.unizh.ch/ds

Jürgen Spitzmüller

Vorlesung: Methoden der Linguistik

WS 2004/05

Korpuslinguistik (7.12.2004)

1 Überblick

1. Theorie – Grundlegende theoretische Fragestellungen:
 - a) Was sind überhaupt Korpora?
 - b) Wozu Korpora?
 - c) Was sollen Korpora leisten/was können sie *nicht* leisten?
 - d) Welche Korpusstypen gibt es?
 - e) Wie müssen linguistische Korpora aufbereitet sein?
2. Praxis – Konkrete Arbeit mit einem vorgefertigten Korpus
 - a) Vorstellung der Korpora des IDS und des COSMAS II-Systems
 - b) Versuch, einige kleine lexikologische, semantische und syntaktische Fragestellungen korpusanalytisch zu lösen

2 Wichtige Korpora des Deutschen

1. IDS-Korpora/COSMAS II-System (<http://www.ids-mannheim.de/kt/projekte/korpora/>, <http://www.ids-mannheim.de/cosmas2/>)
 - Online einsehbar (Anmeldung erforderlich, Zugriff teilweise eingeschränkt)
 - Recherche mit Hilfe eines »Client«-Programmes (momentan nur Windows, Mac- und Linuxversionen für Herbst 2004 angekündigt)
 - ca. 1,52 Milliarden Wortformen
 - Teilweise annotiert (26 Mio. Textwörter)
 - Textsorten: Zeitungstexte, Belletristik, Fachtexte, transkribierte Gespräche, auch historische Texte
 - Besonderheiten: Sehr mächtige Recherchemöglichkeiten; Recherche in Gesprächstranskripten; sehr umfangreich; modulare Zusammenstellung eines spezifischen Untersuchungskorpus aus mehreren (156) kleinen, sehr gut beschriebenen Teilkorpora möglich; Belege jederzeit (auch von Dritten) verifizierbar
 - Einschränkungen: Texte werden (aus urheberrechtlichen Gründen) nur auszugsweise angezeigt (nicht geeignet für diskursanalytische Studien)
2. Leipziger »Wortschatz-Projekt« (<http://wortschatz.uni-leipzig.de>)
 - Online einsehbar (keine Anmeldung erforderlich)



- Recherche mit Hilfe des Browsers (Eingabeformular)
 - Textsorten: v. a. Zeitungstexte, z. T. auch spez. Wortlisten und Fachtexte und Belletristik
 - ca. 518 Mio. Wortformen
 - Besonderheiten: Kollokationen (grafisch), Suche nach Anagrammen
 - Einschränkungen: Keine detaillierte Auskunft über die Korpuszusammensetzung, Ausgabe nur einiger ausgewählter Textstellen
3. TIGER-Korpus (<http://www.ims.uni-stuttgart.de/projekte/TIGER/>)
- Muss komplett heruntergeladen und mit einem speziellen Programm (<http://www.tigersearch.de>) eingesehen werden. Anmeldung (für Programm und Korpus jeweils separat) erforderlich.
 - 700.000 Wortformen
 - Textsorten: Zeitungstexte (Frankfurter Rundschau)
 - Besonderheiten: *Treebanks* (Generierung von Strukturbäumen)
 - Einschränkungen: relativ klein, relativ homogen.

3 Literatur

- [1] Biber, Douglas/Conrad, Susan/Reppen, Randi (1998): *Corpus Linguistics. Investigating Structure and Use*. Cambridge.
- [2] Carstensen, Kai-Uwe (Hg.) (2004): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. München, 2., überarb. und erweiterte Aufl.
- [3] Dittmar, Norbert (1988): Quantitative – qualitative Methoden. In: *Soziolinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, hg. v. Ulrich Ammon, Norbert Dittmar u. Klaus J. Mattheier, Berlin/New York, Bd. 2, S. 879–899 (HSK 3.2).
- [4] Jung, Matthias (1994): Zählen oder deuten? Das Methodenproblem der Diskursgeschichte am Beispiel der Atomenergiedebatte. In: *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*, hg. v. Dietrich Busse, Fritz Hermanns u. Wolfgang Teubert, Opladen, S. 60–81.
- [5] Kennedy, Grame (1998): *An Introduction to Corpus Linguistics*. London/New York.
- [6] Lenz, Susanne (2000): *Korpuslinguistik*. Tübingen (Studienbibliographien Sprachwissenschaft; 32).
- [7] McEnery, Tony/Wilson, Andrew (1998): *Corpus Linguistics*. Edinburgh.
- [8] Paprotté, Wolf (2002): Korpuslinguistik. In: *Arbeitsbuch Linguistik*, hg. v. Horst M. Müller, Paderborn u. a., S. 364–281 (UTB 2169) [als Lernmodul auch online unter <http://luna.lili.uni-bielefeld.de/babelon/>].
- [9] Teubert, Wolfgang (1998): Korpus und Neologie. In: *Neologie und Korpus*, hg. v. Wolfgang Teubert, Tübingen, Bd. 11, S. 129–170 (Studien zur deutschen Sprache).