

Deutsche Lexikologie

Frühlingssemester 2019

Korpuslinguistik

Einleitendes:

- Eine relativ neue linguistische Disziplin (seit etwa 1967)
- Begründer der KL: Henry Kucera und Nelson Francis
- Beliebtheit der KL in vielen europäischen Ländern

Definition:

- KL ist eine korpusbasierte Untersuchung des schriftlichen und mündlichen Gebrauchs der Sprache, ihrer Elemente und ihrer Struktur und die darauf aufbauende Theoriebildung

Korpus, Korpora:

- Korpora bilden die Grundlage der KL
- Korpora sind Sammlungen von nach bestimmten Kriterien ausgewählten und geordneten Ausschnitten (Texte, Textteile, Vorträge, Radiosendungen) einer natürlichen Sprache
- Korpora können auch metalinguistische Daten (soziale, strukturelle Informationen) enthalten
- Ein Korpus besteht aus sprachlichen Einheiten *tokens* (Textwörter, Wortformen) und *types* (abstrakte sprachliche Einheiten)
- Die Größe der Korpora wird in Textwörtern gemessen (1 Mio.-500.000 Mio. Wörter)
- **Klassifikationsmöglichkeiten** der Korpora:
 - nach Größe, Inhalt, Beständigkeit und Repräsentativität
 - nach formalen und inhaltlichen Kriterien
 - nach dem Speichermedium
 - nach dem Sprachmedium
 - nach dem zeitlichen Bezugsrahmen
 - nach dem Umfang
 - nach dem Fachgebiet
- **Auswertung** der Korpora:
 - qualitativ
 - quantitativ
- **Arbeitsschritte** bei korpusgestützten Analysen:
 - Formulierung der Fragestellung
 - Auswahl eines Korpus
 - Aufbereitung des Korpus
 - Erhebung der relevanten Daten
 - Aufbereitung der Daten
 - Auswertung und Interpretation der Daten

Fragestellungen in der KL:

- Kann das Vorfeld eines deutschen Satzes mehrfach besetzt sein? Wenn ja, mit welchen Satzgliedern?
- Kommt es in Texten von E-Mails öfter zu Schreibfehlern als in traditionellen Briefen? Welche Typen von Fehlern sind charakteristisch für E-Mails?

- Welche Fehler machen Lerner des Deutschen auf einem bestimmten Niveau besonders häufig, werden bestimmte Wörter oder grammatische Konstruktionen von diesen Lernern vermieden?
- Mit welchen Adjektiven tritt das Nomen „Haar“ typischerweise gemeinsam auf?
- Werden Modalpartikeln in der gesprochenen Sprache häufiger, weniger häufig oder anders verwendet als in geschriebener Sprache?

Problematische Aspekte der KL:

- **Irreführende positive Evidenz:** In spontanen gesprochenen und sogar in überlegt formulierten schriftsprachlichen Äußerungen können bis zu einem gewissen Grad Abweichungen von der sprachlichen Norm auftreten. Bei der Untersuchung eines Korpus kann es im Einzelfall schwierig sein zu entscheiden, ob eine (meist kleine) Menge von Belegen eines bestimmten sprachlichen Phänomens Ausdruck eines tatsächlich existierenden systematischen Sprachgebrauchs ist
- **Negative Evidenz:** Viele Aussagen zu sprachlichen Phänomenen lassen sich dann selbst in sehr großen Korpora nicht belegen, wenn der Gebrauch bestimmter sprachlicher Konstruktionen sehr selten ist.

Anwendungsbereiche der KL:

- Lexikologie
- Lexikographie
- Semantik
- Textlinguistik
- Translationswissenschaft
- Fremdsprachenunterricht

Vertreter der KL:

- L. Lemnitzer, R. Marcinkevičienė, C. Scherer, H. Zinsmeister

Korpora der deutschen und litauischen Sprache:

- <http://www.dwds.de/>
- <http://wortschatz.uni-leipzig.de/>
- <http://www.ids-mannheim.de/cosmas2/>
- <http://tekstynas.vdu.lt/>