
Beteiligte Projekte

- Ausbau und Pflege der Korpora geschriebener Sprache
- Methoden der Korpusanalyse und -erschließung
- Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen

Beteiligte Mitarbeiter

- Dipl.-Ing. Cyril Belica
- Dr. Marc Kupietz
- Dr. Harald Längen
- Dipl.-Inf. Rainer Perkuhn
- Heike Stadler
- Dr. Andreas Witt

Kontakt

Dipl.-Inf. Rainer Perkuhn
Programmbereich Korpuslinguistik
Institut für Deutsche Sprache
Postfach 10 16 21
D-68016 Mannheim

Telefon: 0621/1581-0
Fax: 0621/1581-200
E-Mail: korpuslinguistik@ids-mannheim.de



 INSTITUT FÜR
DEUTSCHE SPRACHE

R 5, 6-13

D-68161 Mannheim

www.ids-mannheim.de

Mitglied der

Leibniz-Gemeinschaft

Das Institut für Deutsche Sprache (IDS) ist die zentrale Einrichtung zur Erforschung und Dokumentation der deutschen Sprache in ihrem Gebrauch und in ihrer neueren Geschichte. Es gehört zu den 86 außeruniversitären Forschungseinrichtungen und Serviceeinrichtungen für die Forschung der Leibniz-Gemeinschaft.

 INSTITUT FÜR
DEUTSCHE SPRACHE

Programmbereich Korpuslinguistik

DEREWO
*Korpusbasierte Wortlisten
zu DEREKO, dem
DEUTSCHEN REFERENZKORPUS*

Korpusbasierte Wortlisten zum DEUTSCHEN REFERENZKORPUS DEREKO

Muttersprachler sind in der Lage, den Stellenwert eines Wortes intuitiv einzuschätzen, wenn dieses in einer bestimmten Situation gebraucht wird. Sie stützen sich dabei auf die kontextabhängige „gefühlte“ Geläufigkeit des Wortes in ihrem mentalen Lexikon. Experten bemühen sich, solche Einschätzungen möglichst zu objektivieren, z.B. in Form von Aufnahmekriterien für ein Nachschlagewerk. Warum jeweils ein Wort als „wichtiger“ angesehen wird als ein anderes, entzieht sich jedoch auch bei ihrer Argumentation einer vollständigen, formal schlüssigen Erklärung. Allerdings trägt auch aus der fachlichen Sicht die Gebrauchshäufigkeit eines Wortes meist wesentlich zu dessen Stellenwert in einem bestimmten Verwendungszusammenhang bei.

Ausgehend von Wortformfrequenzen des DEUTSCHEN REFERENZKORPUS, und teils automatischer, teils semi-automatischer Grundformzuordnung (Lemmatisierung) ist das Ziel eines Arbeitsschwerpunkts, sich mit diesen intuitiven Vorstellungen der „Relevanz“ auseinander zu setzen. Dabei zeigt sich, dass viele Entscheidungen unterschiedlich weit reichende Auswirkungen haben. Gesondert zu beachten sind z.B. die durch die Unzulänglichkeiten der Datengrundlage beeinträchtigte Authentizität der Korpusfrequenzen, diskontinuierliche Konstituenten (z.B. präfigierte Verben), Eigennamen (insbesondere im Konflikt mit Appellativa), alte vs. neue Rechtschreibung, Varietäten – generell die paradigmatische Zuordnung (insbesondere auch bei Neologismen, sowie bei adjektivisch gebrauchten Partizipien und beliebigen Wortbildungsprodukten).

Unter dem Schlagwort DEREW0 bietet der Programmbereich Korpuslinguistik eine Plattform an, über die Erkenntnisse und Ergebnisse zu diesem Themenkomplex veröffentlicht werden. Abgesehen von der nur für den internen Gebrauch erstellten *elexiko*-Stichwortkandidatenliste mit 350.000 Einträgen finden Sie eine aktuelle Übersicht der jeweils öffentlich zur Verfügung gestellten Wortlisten unter <http://www.ids-mannheim.de/kl/derewo/>

Stand: 04/12

