

# Verfügbare Deutsche Korpora (gesprochene und geschriebene Texte)

## Chatkorpora von Michael Beißwenger

[http://www.linse.uni-due.de/tl\\_files/PDFs/Publikationen-Rezensionen/Chatkorpora\\_Beisswenger\\_2013.pdf](http://www.linse.uni-due.de/tl_files/PDFs/Publikationen-Rezensionen/Chatkorpora_Beisswenger_2013.pdf)

## Hamburg Mehrsprachigkeit

[http://www.corpora.uni-hamburg.de/sfb538/de\\_overview.html](http://www.corpora.uni-hamburg.de/sfb538/de_overview.html)

[http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/links/korpora\\_links](http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/links/korpora_links)

<http://www.ids-mannheim.de/service/quellen/texte.html#TXT>

GeWiss – Gesprochene Wissenschaftssprache kontrastiv <https://gewiss.uni-leipzig.de/de/> Prüfungsgespräche

- GeWiss = in FOLK IDS <http://dgd.ids-mannheim.de> und <http://agd.ids-mannheim.de/folk.shtml>
- IDS <http://agd.ids-mannheim.de/datenbanken.shtml>

größeres Text- und Sprachkorpora. Die Textkorpora enthalten Belletristik (u.a. Andersch, Böll, Frisch, Goethe, Grimm), Lehrbücher, Schriften von Marx und Engels, sowie Zeitungsausschnitte (Neues Deutschland, Die Welt, Mannheimer Morgen, DieZeit, stern) u.a.; außerdem das "Wendekorpora" des Projekts "Gesamtdeutsche Korpusinitiative" in Zusammenarbeit mit dem Zentralinstitut für Sprachwissenschaft (ehem. DDR) mit Artikeln, Flugblättern, Protokollen und Reden zwischen 1989 und 1990, die die "Wende" dokumentieren. Die Sprachkorpora (12 000 Tonaufnahmen) sind vor allem Tondokumente des gesprochenen Deutsch der 60er und 70er Jahre unseres Jhds, die u.a. Dialogstrukturen, überregionale Umgangssprache (von der Stanford University erarbeitet) und deutsche Mundarten enthalten. Ein aktuelles Projekt des IDS, Wandel im gesprochenen Deutsch, untersucht Veränderungen der Spracheinstellungen und des Sprachverhaltens einzelner Sprecher im Zeitraum von 35-40 Jahren. Viele der IDS-Korpora fließen als Beispielmateriale in die neue "Grammatik der deutschen Sprache" des IDS ein (G. Zifonoun et al., Berlin 1997). Kopien von Korpora, Transkripten und Begleitmaterialien können beim IDS erworben werden.

- Jugendsprachkorpora <http://www.mediensprache.net/de/medienanalyse/corpora/>

- Uni-Sprechstunden (Hoffmann-Korpus, Meer)
  - Boettcher, Wolfgang/Limburg, Anika/Meer, Dorothee/Zegers, Vera (2005)(Hg.): „ich komm (0) weil ich wohl etwas das thema meiner hausarbeit etwas verfehlt habe,“ – Sprechstundengespräche an der Hochschule. Ein Transkriptband. (Download unter: [www.verlag-gespraechsforschung.de/boettcher.htm](http://www.verlag-gespraechsforschung.de/boettcher.htm)).
  - Boettcher, Wolfgang/Meer, Dorothee (Hg.) (in Vorb.): „wer hat denn den text zu heute gelesen?“ – Kommunikation an der Hochschule. Transkripte und Analyseperspektiven. Verlag für Gesprächsforschung: Radolfzell.

## Korpora zur deutschen Sprache und Literatur, Textsammlungen

### [Textkorpora des Instituts für Deutsche Sprache, Mannheim](#)

Größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten weltweit; über das Korpus-Recherche- und Korpus-Analyse-System [COSMAS II](#) teilweise öffentlich und gebührenfrei online recherchierbar

### [Deutsches Spracharchiv am IDS Mannheim](#)

Sammel- und Dokumentationsstelle des IDS für Korpora des gesprochenen Deutsch; Serviceangebot für Materialien aus 17 Korpora

### [Projekt Gutenberg, Das deutsche Literaturzentrum im Internet](#)

Texte von deutschsprachigen Autoren, deren Urheberrechte abgelaufen sind, in nicht-wissenschaftlichen Ausgaben

### [Erlanger Liste: Digitale Texte](#)

nach Gattungen geordneter Index zu online verfügbaren deutschsprachigen und fremdsprachigen literarischen Texten

### [Erlanger Liste: Literarische Epochen](#)

Epochenbeschreibungen; Zusammenstellung von online-verfügbaren literarischen Texten nach Epochen geordnet

### [German Collection Electronic Texts](#)

Wolfram von Eschenbach, Gottfried von Strassburg, Minnesangs Frühling, Wieland, Wittgenstein, Kafka u.a.

### [Oxford Text Archive](#)

verwaltet über 2.500 Quellen in 25 verschiedenen Sprachen, über Suchformular erschließbar, Texte teilweise online verfügbar

### [Luther-Bibel](#)

Die Bibel nach der Übersetzung Martin Luthers in der revidierten Fassung von 1984.

## Deutsch Synchron

- **DWDS-Kernkorpus (schriftliche Texte)**  
<http://www.dwds.de/resource/kerncorpus/>

Korpus der Berlin-Brandenburgischen Akademie der Wissenschaften, auf dessen Grundlage das [Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts \(DWDS\)](#) erstellt wurde.

- **Projekt Deutscher Wortschatz**  
<http://wortschatz.uni-leipzig.de/>

Deutscher Wortschatz Online. Enthält 35 Millionen Sätze mit 500 Millionen Wörtern.

- **IDS-Korpora (schriftliche Texte und Gesprochene Sprache)**  
<http://www.ids-mannheim.de/kt/corpora.html>

Korpora des Instituts für Deutsche Sprache. Es handelt sich um die weltweit größte Sammlung von deutschsprachigen Textkorpora als empirische Basis für die linguistische Forschung. Online-Recherche ist mit [COSMAS II](#) möglich.

- **LIMAS-Korpus (schriftliche Texte)**  
<http://www.korpora.org/Limas/>

Repräsentatives Zeitschnittkorpus der deutschen Gegenwartssprache (Schriftsprache) von 1970: Es handelt sich um eine Auswahl von 500 Texten, respektive Textfragmenten, verschiedener Textsorten mit insgesamt 1 Million Wortformen. Das Korpus kann in seiner Gesamtheit im WWW recherchiert werden.

- **Korpus Südtirol**  
[http://www.korpus-suedtirol.it/index\\_DE](http://www.korpus-suedtirol.it/index_DE)

Eine Initiative zur Sammlung, Archivierung und korpuslinguistischen Erschließung von südtiroler deutschsprachigen Texten.

## *Deutsch Diachron*

- **Bibliotheca Augustana**  
[www.fh-augsburg.de/~harsch/augustana.html](http://www.fh-augsburg.de/~harsch/augustana.html)

litteraturae et artis collectio

- **Kali Korpus**  
[www.kali.uni-hannover.de](http://www.kali.uni-hannover.de)

Das Kali-Korpus (Kali: "Korpusarbeit Linguistik") ist ein teilannotiertes diachrones Korpus des Deutschen für Forschung und Lehre, das seit Ende 2003 am Deutschen Seminar der Universität Hannover unter der Leitung von Prof. Dr. Gabriele Diewald aufgebaut wird.

- **Textkorpus von Thomas Gloning**  
<http://www.uni-giessen.de/gloning/etexte.htm>

frei zugänglich

- **Bochumer Mittelhochdeutsch Korpus**  
[http://www.ruhr-uni-bochum.de/wegera/archiv\\_1.htm](http://www.ruhr-uni-bochum.de/wegera/archiv_1.htm)
- **Mittelhochdeutsche Begriffsdatenbank (MHDBDB)**  
<http://mhdbdb.sbg.ac.at>

Enthält 4.7 Millionen Wörter

- **CEEC (Codices Electronici Ecclesiae Coloniensis)**  
<http://www.ceec.uni-koeln.de>

Im Rahmen des Projektes CEEC werden die mittelalterlichen Kodizes der Erzbischöflichen Diözesan- und Dombibliothek Köln (DDB) digitalisiert.

- **TITUS**

<http://titus.uni-frankfurt.de/indexd.htm>

Thesaurus Indogermanischer Text- und Sprachmaterialien

- **mediavum**

<http://www.mediaevum.de>

viele Links zu historischen Texten