

58. Jahrestagung des Leibniz-Instituts für Deutsche Sprache vom 15. bis 17. März 2022
(als Online-Konferenz)

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

METHODENMESSE

Mittwoch, 16. März 2022, 15:45 Uhr bis 17:45 Uhr

Forschungsdatenbank Lernertexte (FD-LEX)

Sabine Hachmeister, Sandra Tietjens, Rebekka Wanka, Charlotte Stehr, Michael Becker-Mrotzek

Die webbasierte Forschungsdatenbank Lernertexte (FD-LEX) ist eine stetig wachsende Datenbank, die Texte von Lernerinnen und Lernern samt ausgewählter Meta- und Testdaten der Probandinnen und Probanden aus verschiedenen Forschungsprojekten zur wissenschaftlichen Nachnutzung bereitstellt. Die Korpora stammen aus unterschiedlichen Forschungsprojekten und umfassen mehrere tausend Texte auf verschiedenen Sprachen, darunter Deutsch, Türkisch oder Russisch. Entsprechend der unterschiedlichen Forschungsdesigns und Fragestellungen der Projekte, liegen Texte verschiedener Textsorten von Schülerinnen und Schülern unterschiedlicher Altersgruppen und Schulformen vor. Alle Texte werden in transkribierter Form im .pdf-Format und zum Teil auch als Scans der handschriftlichen Originale bereitgestellt. Die Daten der einzelnen Korpora können anhand verschiedener Kriterien nach den eigenen Interessen gefiltert werden. Zu jeder Filtervariable sind kurze Erläuterungen hinterlegt. Die Rechercheergebnisse können als Excel- und Datentabelle ausgegeben werden, die transkribierten und – falls vorhanden – die handschriftlichen Lernertexte als PDF-Dateien. Außerdem liegen Informationen zu den jeweiligen Forschungsprojekten vor, beispielsweise die dort verwendeten Schreibimpulse als PDF-Dateien. Hier beispielhaft eine Übersicht über zwei Korpora, die derzeit zur Verfügung stehen:

Das Scriptoria-Textkorpus umfasst 5.628 Texte von 938 Schülerinnen und Schülern auf Deutsch. Es handelt sich hierbei um Berichts- und Argumentationstexte, die in fünften und neunten Klassen an Gesamtschulen und Gymnasien in Hannover und Köln zwischen den Jahren 2013 und 2015 erhoben wurden. Von den Schülerinnen und Schülern wurden zu drei Messzeitpunkten je zwei Texte geschrieben. Dies geschah im Rahmen des Forschungsprojekts *Unterrichtliche Förderung von Teilkomponenten der Schreibkompetenz*, einem Verbundvorhaben der Professur für Pädagogische Psychologie an der Leibniz Universität Hannover (Prof. Grabowski) und der Professur für Deutsche Sprache und ihre Didaktik an der Universität zu Köln (Prof. Becker-Mrotzek). Es handelt sich um die Weiterführung des Projektes *Diagnose und Förderung von Teilkomponenten der Schreibkompetenz*.

Das TSK-Korpus umfasst 585 Texte von 39 Schülerinnen und Schülern auf Deutsch. Die Texte stammen aus einer Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter, welche an der Universität Siegen durchgeführt wurde. Es wurde anhand von zwei Klassen ermittelt, wie sich die Fähigkeit von Kindern in der 2., 3. und 4. Klasse entwickelt, Texte im Rahmen verschiedener kommunikativer Anforderungen (Textsorten) zu schreiben: Erzählung,

Bericht, Instruktion, Beschreibung, Argumentation. Der Erhebungszeitraum lag zwischen den Jahren 2001 und 2003.

Die Datenbank ist als ein gemeinsames Projekt des Mercator-Instituts für Sprachförderung und Deutsch als Zweitsprache der Universität zu Köln und der Professur für Pädagogische Psychologie der Leibniz-Universität Hannover entstanden und wurde von der Stiftung Mercator gefördert. Nach der Entwicklungsphase im Jahr 2017, konnte sie 2018 online geschaltet werden, zunächst nur mit den Textdaten des umfangreichen Scriptoria-Korpus. Ende 2019 startete das Projekt zum Ausbau der Datenbank in Kooperation mit dem Verbund Forschungsdaten Bildung und im Dezember 2021 fand der Relaunch der Datenbank-Website statt. Danach werden stetig neue Textkorpora aufgenommen.

Derzeit werden die frequentesten Wörter aus dem Scriptoria-Korpus in einem weiteren Projekt des Mercator-Instituts verwendet. Das Projekt „Entwicklung eines Referenzwortschatz für die Schule“ beschäftigt sich dabei mit der Entwicklung und Bereitstellung eines annotierten, sprachwissenschaftlich und sprachdidaktisch begründeten Wortschatzes für die Klassen 1 bis 10 in Form einer webbasierten Anwendung. Die häufigsten Wörter aus dem Scriptoria-Korpus werden im Teilkorpus „Schreibwortschatz“ des Referenzwortschatzes in die webbasierte Anwendung annotiert eingespeist.

In unserem Beitrag wird der Aufbau der Forschungsdatenbank samt bereits vorhandener Korpora und Recherchemöglichkeiten vorgestellt. Dabei wird beispielhaft eine Recherche im Bestand der Datenbank durchgeführt, um die verschiedenen Filtermöglichkeiten und die Datenausgabe vorzuführen. Zudem wird das Projekt „Entwicklung eines Referenzwortschatzes für die Schule“ vorgestellt und dabei auf die Verknüpfungspunkte zur Datenbank FD-LEX eingegangen.

Links:

www.fd-lex.de

<https://www.mercator-institut-sprachfoerderung.de/de/forschung-entwicklung/aktuelle-projekte/forschungsdatenbank-lernertexte-fd-lex/>

<https://www.mercator-institut-sprachfoerderung.de/de/forschung-entwicklung/aktuelle-projekte/referenzwortschatz-fuer-die-schule/>