

Gesprochenes Deutsch: Struktur, Variation, Interaktion

METHODENMESSE

Mittwoch, 06. März 2024, 16:15 Uhr bis 18:00 Uhr

Die Datenbank für Gesprochenes Deutsch und das Forschungs- und Lehrkorpus für Gesprochenes Deutsch

Silke Reineke / Mark-Christoph Müller (IDS)

Die Datenbank für Gesprochenes Deutsch (DGD) ist eine Korpusplattform, über die der wissenschaftlichen Öffentlichkeit aktuell 40 mündliche Korpora (über 4.700 Stunden Audioaufnahmen, rund 260 Stunden Videoaufnahmen und 20 Millionen transkribierte Tokens) aus dem Archiv für Gesprochenes Deutsch zur Verfügung gestellt werden. Die DGD bietet Nutzer/-innen mit unterschiedlichen Forschungsinteressen und für verschiedene Fragestellungen die Möglichkeit des web-basierten explorativen Browsens, gezielter Suchanfragen auf verschiedenen Annotationsebenen, eine Wortlisten-Funktionalität sowie Möglichkeiten zum Download von kurzen Daten-Ausschnitten.

Der Bestand der in der DGD zugänglichen Daten wird laufend erweitert und beinhaltet aktuell unter anderem wichtige Gesprächskorpora, Variationskorpora und Interviewkorpora des AGD.

Zu den online in der DGD verfügbaren Korpora gehört auch das „Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)“ (Schmidt 2014; Reineke et al. 2023), das kontinuierlich erweitert wird. FOLK speist sich sowohl aus Aufnahmen des Projektes selbst als auch aus Aufnahmen, die Interaktionsforschende für ihre Projekte erheben und anschließend für das FOLK-Korpus ‚spenden‘. FOLK stellt alle Aufnahmen in einheitlich erschlossener Form der Wissenschaft zur Verfügung. Das Korpus umfasst aktuell (Version 2.20 vom Juni 2023) 414 Gesprächsaufnahmen (347h Audio-Aufnahmen, 220h Video-Aufnahmen aus vielen Bereichen des gesellschaftlichen Lebens (Arbeit, Freizeit, Bildung, öffentliches Leben, Dienstleistungen etc.). Die Daten des FOLK-Korpus werden nach zeitgemäßen Standards erschlossen (transkribiert, annotiert und dokumentiert). Durch die Annotationen der Daten auf vier Ebenen (Transkription, Normalisierung, Lemmatisierung, Part-of-Speech-Tagging) sind in der DGD gezielte Recherchen auf insgesamt 3.3 Millionen transkribierten Wörtern auf allen Annotationsspuren möglich. Die kontextsensitiven Suchergebnisse lassen sich in Form von KWIC-Listen, Transkripte und Audio-/Video-Stream in der DGD ansehen, mit umfangreichen Metadaten kombinieren und in verschiedenen Dateiformaten lokal durch Forschende weiter verarbeiten.

Auf der Projektmesse sollen die Datenbestände und Recherchemöglichkeiten der DGD – insbesondere mit Blick auf FOLK als zeitgemäß und umfassend annotiertes Korpus – dargestellt werden.

Literatur

Schmidt, Thomas (2014): Gesprächskorpora und Gesprächsdatenbanken am Beispiel von FOLK und DGD. In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion

15, S. 196-233. <http://www.gespraechsforschung-ozs.de/fileadmin/dateien/heft2014/px-schmidt.pdf>

Reineke, Silke / Deppermann, Arnulf / Schmidt, Thomas (2023): Das Forschungs- und Lehrkorpus für Gesprochenes Deutsch (FOLK). Zum Nutzen eines großen annotierten Korpus gesprochener Sprache für interaktionslinguistische Fragestellungen. In: Deppermann, Arnulf / Fandrych, Christian / Kupietz, Marc / Schmidt, Thomas (Hg.): Korpora in der germanistischen Sprachwissenschaft. Berlin/Boston: De Gruyter, S. 71–102.