

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

DGD
DATENBANK FÜR
GESPROCHENES
DEUTSCH

- ▶ Plattform für den Zugriff auf Korpora des Archiv für Gesprochenes Deutsch (zur wissenschaftlichen Nutzung)
- ▶ Aktuell 41 Korpora, darunter das Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)
- ▶ Rund 19.000 Sprechereignisse, über 4.700 h Audioaufnahmen, rund 260 h Videoaufnahmen, über 21 Mio. transkribierte Tokens
- ▶ Über 17.500 registrierte Nutzer:innen (06/2024)

FOLK

FORSCHUNGS- UND LEHRKORPUS FÜR GESPROCHENES DEUTSCH

- Großes **Gesprächskorpus des Deutschen**, kontinuierlicher Ausbau (seit 2008)
 - **Audio-/Videoaufnahmen** aus vielen Bereichen des gesellschaftlichen Lebens (Arbeit, Freizeit, Bildung, öffentliches Leben, Dienstleistungen etc.)
 - **Ziel:** Wissenschaftsöffentliche Bereitstellung einer großen, nach **aktuellen Standards erschlossenen** und breit **diversifizierten Datenbasis** für die Untersuchung **gesprochener Sprache in natürlicher Interaktion**, als **Referenzkorpus**
 - **Erhebungen im Projekt und Datenspenden im Rahmen von Kooperationen**
- Umfang** (Version 2.22, 06/2024)
- **436** Gespächereignisse, **1.381** dokumentierte Sprecher:innen
 - Rund **365 h Audio-/Videoaufnahmen** von Interaktionen, davon 192 h reine Audioaufnahmen und 172 h Videoaufnahmen, z.T. aus mehreren Perspektiven
 - Transkripte mit insgesamt **3,45 Mio. Tokens**
- Annotationen und Metadaten**
- Transkription (cGAT-Minimaltranskript)
 - Orthographische Normalisierung
 - Lemmatisierung
 - Part-of-Speech-Tagging
 - Umfangreiche Metadaten zu Sprechereignissen und Sprecher:innen
- **Größtes und korpustechnologisch am weitesten entwickeltes (Gesprächs-)Korpus des AGD**
- **Exploration und gezielte Recherchen in FOLK mit Funktionalitäten der DGD**

BROWSING

- **Korpusexploration**, qualitative Analyse
- Ansehen von Metadaten zu Korpora, Sprechereignissen und Sprecher:innen, Zusatzmaterial
- Anhören und Ansehen von Aufnahmen (**Audio und Video**) und alignierten Transkripten

RECHERCHE

- Gezielte **Token suche** über alle Annotations-ebenen
- **KWIC-Trefferanzeige** im Transkript-Kontext
- **Filtern** nach Metadaten, Kontext, Position (z.B. vor/nach Sprecherwechsel, in Überlappung etc.)
- **Audio-Wiedergabe** direkt aus aligniertem Suchergebnis
- Erstellung von **Wortlisten**
- **Metadaten suche** zur Selektion von Sprechereignissen und Sprecher:innen, Speichern als **virtuelle Korpora** (z.B. als Vorauswahl für die Tokensuche)

DATENMANAGEMENT

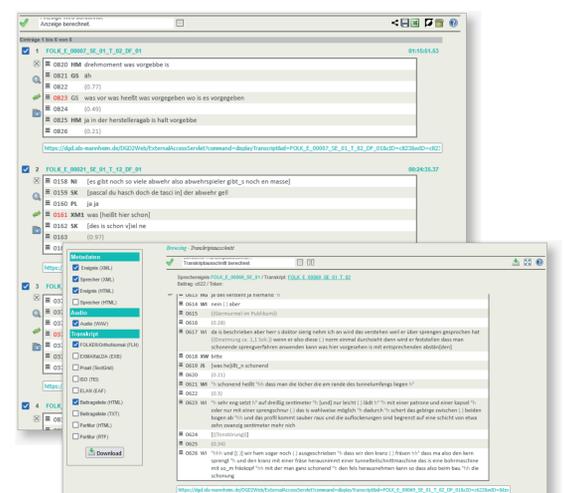
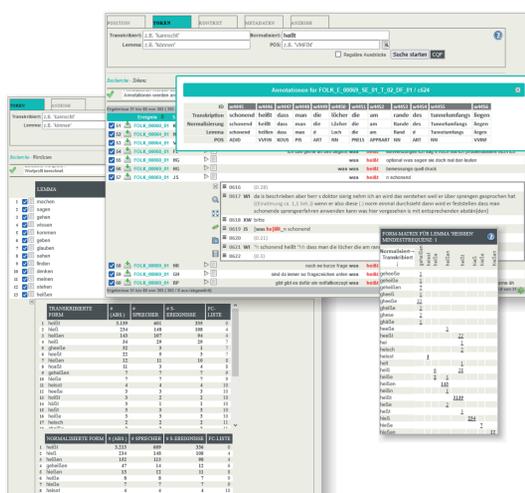
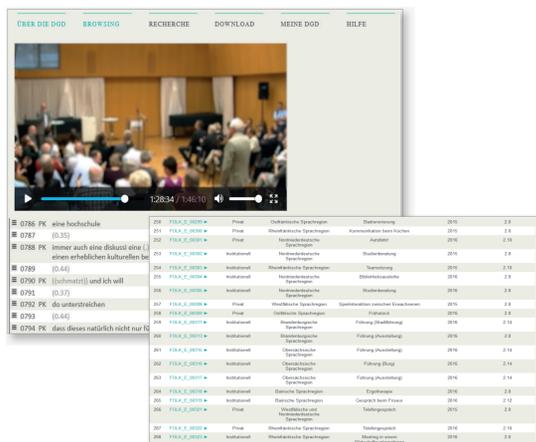
- **KWIC-Suchergebnisse, Wortlisten, Kollektionen** und **virtuelle Korpora speichern** und online mit anderen DGD-Nutzer:innen teilen
- Ausgewählte Transkriptausschnitte zu **Kollektionen** zusammenstellen
- Einzelne **Transkriptausschnitte (inkl. Audio)** in diversen Formaten herunterladen
- KWIC-Suchergebnisse, Wortlisten und Kollektionen als **TXT- bzw. XML-Datei** herunterladen
- **Sampling** und **quantitative** Analyse von Suchergebnissen

Kontakt:
Dr. Silke Reineke
Dr. Mark-Christoph Müller
Abteilung Pragmatik
Leibniz-Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim
dgd-support@ids-mannheim.de

Hausanschrift:
Leibniz-Institut für Deutsche Sprache
R 5, 6-13
68161 Mannheim

Tel.: +49 621 1581-0
Fax: +49 621 1581-200
info@ids-mannheim.de
www.ids-mannheim.de

© 2024 IDS Mannheim/ÖA



Laufende Liste von Publikationen
auf Basis des FOLK-Korpus