

# DATENBANK FÜR GESPROCHENES DEUTSCH

Browsen und Recherchieren in AGD-Korpora

Die DGD2 bietet derzeit (März 2014) Zugriff auf 19 Korpora aus den Beständen des Archivs für Gesprochenes Deutsch (AGD) mit einem Gesamtumfang von etwa 9000 Interaktionen, 2500h Aufnahmen und 8 Millionen transkribierter Tokens. Die Datenbasis wird kontinuierlich ausgebaut.

## KORPORA

- BR** Biographische und Reiseerzählungen
- DS** Dialogstrukturen
- EK** Elizitierte Konfliktgespräche
- FOLK** Forschungs- u. Lehrkorpus für gesprochenes Deutsch
- FR** Grundstrukturen: Freiburger Korpus
- HL** Deutsche Hochlautung
- IS** Emigrantendeutsch in Israel
- ISW** Emigrantendeutsch in Israel: Wiener in Jerusalem
- KN** Deutsche Standardsprache: König-Korpus
- OS** Deutsche Mundarten: ehemalige deutsche Ostgebiete
- PF** Deutsche Umgangssprachen: Pfeffer-Korpus
- ZW** Zwirner-Korpus

## Browsen, Recherche, Download

Die DGD2 ermöglicht das **Browsen** in Metadaten (zu Korpora, Sprechereignissen und Sprechern), Tonaufnahmen, Transkripten und Zusatzmaterialien.

### Ereignis ZW--\_E\_00001

Basisdaten	
Beschreibung	Gedante Aufnahmeaktion
Sonstige Bezeichnungen	ZW001 ; V1
Datum	1955-01-01 (Monat und Tag nicht dokumentiert)
Ort	Kreis: Freiburg im Breisgau Ortsname: Burkheim Planquadrat: 4807
Sprechereignisse und Sprecher	
1 Sprechereignis	ZW--_E_00001_SE_01 (Erzählung)
Themen	Weinbau ; Ausbau ; Absatz

Metadaten zu einer Aufnahme aus dem Zwirner-Korpus

Die einzelnen Datentypen sind dabei vielfach untereinander verknüpft, so dass zum Beispiel beim Lesen eines Transkripts auch die zugehörigen Stellen in der Aufnahme angehört und Informationen zu den Sprechern abgerufen werden können.

KORPUSBESCHREIBUNGEN ERGEBNISDOKUMENTATIONEN SPRECHERDOKUMENTATIONEN TRANSKRIPTE

FOLK\_E\_00001\_SE\_01\_T\_02 | FOLK\_E\_00007\_SE\_01\_T\_02

Doppelklick auf eine Zeile im Transkript zum Hören der originalen Aufnahme (12 Sekunden maximal). Klick auf die **Beleg**-Buttons zum Anfordern der originalen Aufnahme

0001	[[Winnengewinn; 4.1.1]]
0002	GS [heute noch mal]
0003	XM [[kurzert sich]]
0004	[[0:24]]
0005	GS heute noch mal ( ) zwei stunden mit aufzeichnung ( ) ja

Ausschnitt aus einem FOLK-Transkript

Für die **Recherche** in Metadaten und Transkripten stellt die DGD2 eine Volltextsuche sowie verschiedene Formen strukturierter Suchen zur Verfügung.

Dabei lassen sich Abfragen über Metadaten mit Abfragen über das sprachliche Material in vielfältiger Weise verknüpfen, Suchergebnisse schrittweise automatisch oder manuell verfeinern, und nutzerdefinierte virtuelle Korpora, Beleg- oder Beispielsammlungen dauerhaft in der Datenbank speichern.

Ereignis	Sprecher	Treffer
1	FOLK_00001 DM	oder was ich doch sehr interessant <b>find</b> zum rhythmus wenn der rhythmus immer
2	FOLK_00004 JK	an da waren nie <b>gewacht</b> dort
3	FOLK_00007 GS	s ihm oder sie lassen selbst <b>achte</b> so lang bis er s hat oder auch till
4	FOLK_00009 LB	<b>finden</b> so an modernen fektischen so gut wie gar nich
5	FOLK_00011 BK	noch mal so s gilt gelegentl <b>finden</b> einen loszuwerden
6	FOLK_00011 VK	s <b>find</b> ich auch korrech fröschmerstraße da bring ich vorzetn ja
7	FOLK_00013 CJ	gute mal nach ob du die <b>findest</b>
8	FOLK_00015 FN	gibt ich <b>find</b> s sehr schönung

Suchergebnis in KWIC-Ansicht

Für eine weiterführende Analyse am lokalen Rechner erlaubt die DGD2 den **Download** einzelner Belege (Audio- und Transkriptausschnitt).

Aus den Korpora „Deutsche Mundarten“ (Zwirner-Korpus), „Deutsche Umgangssprache“ (Pfeffer-Korpus) sowie FOLK werden zudem ausgewählte komplette Datensätze zum Herunterladen angeboten.

Die Downloads enthalten auch eine Repräsentation der Daten als XML-Dokumente. In dieser Form eignen sie sich besonders gut für eine flexible Weiterverarbeitung.

```

4434 <contribution speaker="00000000000000000000" id="00000000000000000000" file="00000000000000000000" title="00000000000000000000" />
4435 <para level="2" speaker="00000000000000000000" id="00000000000000000000" time="443.714" />
4436 <time timestamp="00000000000000000000" title="00000000000000000000" time="443.714" />
4437 <ch id="00000000000000000000" pos="00000000000000000000" />
4438 <ch id="00000000000000000000" pos="00000000000000000000" />
4439 <ch id="00000000000000000000" pos="00000000000000000000" />
4440 <ch id="00000000000000000000" pos="00000000000000000000" />
4441 <ch id="00000000000000000000" pos="00000000000000000000" />
4442 <ch id="00000000000000000000" pos="00000000000000000000" />
4443 <ch id="00000000000000000000" pos="00000000000000000000" />
4444 <ch id="00000000000000000000" pos="00000000000000000000" />
4445 <time timestamp="00000000000000000000" title="00000000000000000000" time="442.034" />
4446 </contribution />

```

Transkriptausschnitt als XML-Dokument

## Nutzung

Zur Nutzung der DGD2 ist eine einmalige kostenlose Registrierung erforderlich.

Über 2000 NutzerInnen (Stand: März 2014) sind bereits für die DGD2 registriert.

Sie finden die DGD2 unter:

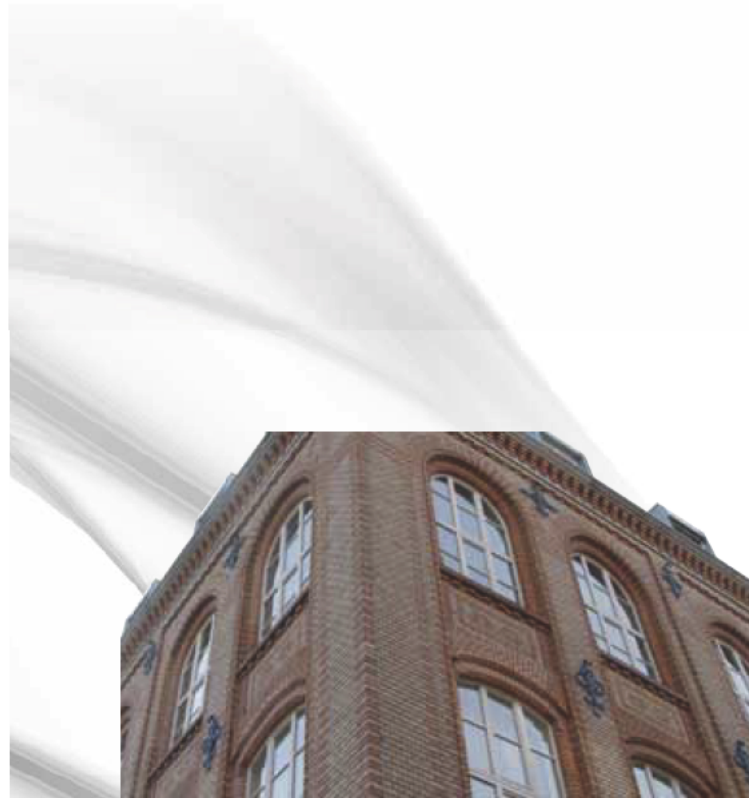
<http://dgd.ids-mannheim.de>

## Forschungsgegenstand

Die Datenbank für gesprochenes Deutsch (DGD2) ist eine Plattform für den Zugriff auf mündliche Korpora im Archiv für Gesprochenes Deutsch (AGD).

Die DGD2 ermöglicht externen Nutzern aus Forschung und Lehre nach einmaliger Registrierung das webbasierte Browsen und Recherchieren in ausgewählten Teilen der AGD-Bestände, darunter im Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK) und im Korpus „Deutsche Mundarten“ (Zwirner-Korpus).

Dr. Thomas Schmidt  
Leiter des Programmbereichs Mündliche Korpora



## Projekt-Kontakt

Dr. Schmidt, Programmbereich Mündliche Korpora  
Institut für Deutsche Sprache, Mannheim  
[agd@ids-mannheim.de](mailto:agd@ids-mannheim.de)

## Beteiligte Projekte

- Archiv für Gesprochenes Deutsch (DGD2)
- Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)

## Beteiligte Mitarbeiterinnen und Mitarbeiter

- Dr. Thomas Schmidt
- Sylvia Dickgießer
- Joachim Gasch



INSTITUT FÜR  
DEUTSCHE SPRACHE

R 5, 6-13 • D-68161 Mannheim  
Tel.: +49 621 1581-0  
Fax: +49 621 1581-200  
[info@ids-mannheim.de](mailto:info@ids-mannheim.de)  
[www.ids-mannheim.de](http://www.ids-mannheim.de)

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 Forschungs- und Serviceeinrichtungen der Leibniz-Gemeinschaft.

Abteilung Pragmatik  
Institut für Deutsche Sprache

DGD  
DATENBANK FÜR  
GESPROCHENES  
DEUTSCH

DATENBANK FÜR  
GESPROCHENES DEUTSCH

Eine Korpusplattform für mündliche Korpora des AGD

Stand 01/2014