

# ZuMult – Zugänge zu Korpora gesprochener Sprache

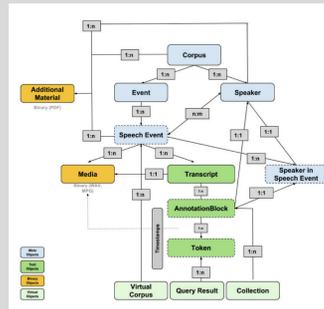
Anne Ferger, Elena Frick, Thomas Schmidt, Matthias Schwendemann, Franziska Wallner

Kooperationspartner:



## Objektorientiertes Korpusdatenmodell

Die wichtigsten Objekte eines Korpus und ihre Beziehungen zueinander:



Mehr zum Projekt **ZuMult** („Zugänge zu multimodalen Korpora gesprochener Sprache: Vernetzung und zielgruppenspezifische Ausdifferenzierung“, Fandrych et al., 2022; 2023, gefördert von DGF im LiS-Programm) unter <https://zumult.org>

## Standards für die Datenrepräsentation

**Medien:** PCM-WAV/MP3 (Audio), MPEG-4 (Video)

**Transkriptionen und Annotationen:** XML-basierter ISO-Standard 24624:2016

- ▶ basiert auf den Richtlinien der Text Encoding Initiative (TEI)
- ▶ interoperabel mit den Formaten der gängigsten Transkriptionseditoren (z. B. EXMARALDA, FOLKER, ELAN, Praat und Transcriber)

**Sprecher- und Sprechereignismetadaten:** XML, CMDI

```
ISO 24624:2016
Language resource management –
Transcription of spoken language

<?xml version="1.0" encoding="UTF-8" ?>
<transcription xmlns="http://www.iso.org/iso/24624" >
  <channel id="1" >
    <media id="1" >
      <transcription id="1" >
        <text id="1" >
          <p>You keep (0.2) interup-ling me.
          Du unterbrichst mich immer.
          I am so sorry.
          Tut mir leid.
          </p>
        </text>
      </transcription>
    </media>
  </channel>
</transcription>
```

## Drei-Ebenen-Architektur

- ▶ **Interne Ebene** (eng. *physical level*) befasst sich mit der physikalischen Speicherung von Daten, z. B. in einer relationalen Datenbank, einem Datenrepositorium oder dem Dateisystem auf einem Server
- ▶ **Konzeptuelle Ebene** (eng. *logical level*) modelliert die Objekte der Anwendung (im Fall von Korpora z. B. Audio- und Videoaufnahmen, Transkripte, Metadaten) in Datenstrukturen und stellt die Operationen zur Verfügung, die auf den Objekten ausgeführt werden (z. B. das Bilden eines Transkriptausschnitts, eine Suche auf Transkripten)
- ▶ **Externe Ebene** (eng. *view level*), umfasst die (typischerweise grafischen) Schnittstellen zum/r Nutzer/in, z. B. Anzeigen eines Transkripts

# ZuMult

## Standardisierung, Vernetzung, Flexibilität

bei der Entwicklung von Webanwendungen für den Zugriff auf

# audiovisuelle Sprachdaten

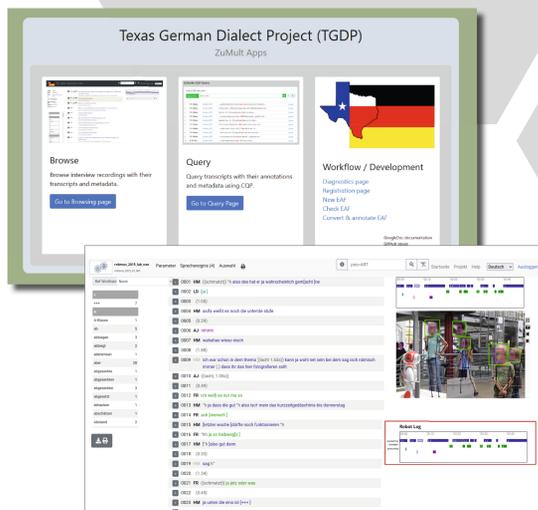
## Open Source Java-Bibliothek

Entwicklung eigener Backend-APIs und Webanwendungen auf Basis der ZuMult.jar-Datei:

<https://github.com/zumult-org/zumultapi> (auf Anfrage)

**Aktueller Einsatz:**

- ▶ am IDS Mannheim für den Zugriff auf AGD-Korpora (u. a. FOLK und GeWiss Korpora)
- ▶ an der University of Texas in Austin für den Zugriff auf Texas German Dialect Corpus (TGDP)
- ▶ an der Universität Duisburg-Essen für den Zugriff auf multimodale und multisensorielle Korpora (MuMoCorp)



## Beispiel-Nutzungsszenarien

- ▶ Arbeiten mit Korpora in DaF/DaZ
- ▶ Korpusgestützte Interaktionsanalyse
- ▶ Korpusbasierte Varietätenforschung

unter <https://zumult.ids-mannheim.de/ProtoZuMult/index.jsp>

## Technische Umsetzung

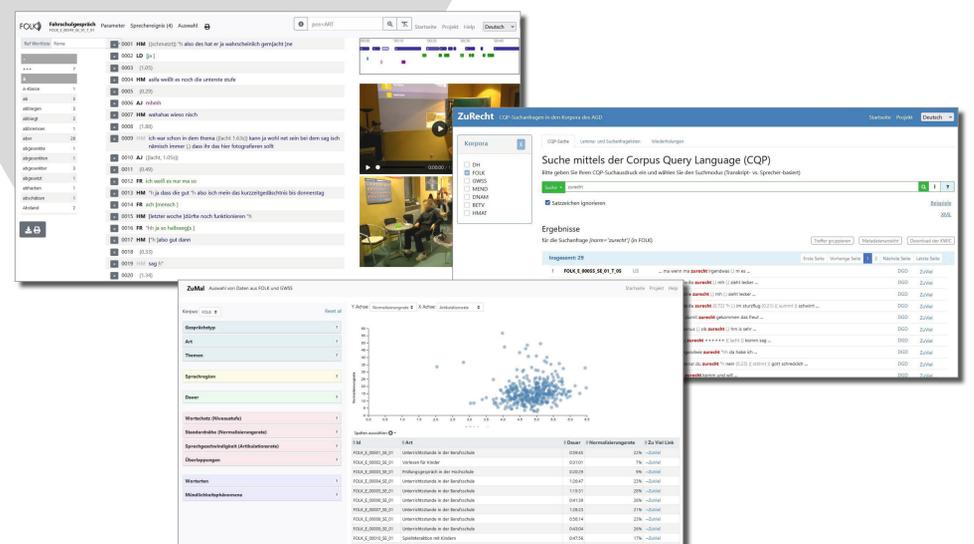
Verwendung möglichst weit verbreiteter und etablierter Technologien

**Backend:**

- ▶ Java EE Framework + REST API
- ▶ Query-Funktionalität, realisiert mit Hilfe von MTAS (*Multi Tier Annotation Search*, Brouwer et al., 2017)

**Client/Webanwendungen:**

- ▶ jQuery, XSLT, HTML5, Bootstrap
- ▶ Datenvisualisierung mit Hilfe von XSL, SVG, WebVT



**Kontakt:**  
Dr. Thomas Schmidt  
thomas@linguisticbits.de  
<https://linguisticbits.de/>  
Elena Frick M.A.  
Leibniz-Institut für Deutsche Sprache  
frick@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/IOA



<https://zumult.org/>

## Literatur:

- Brouwer, M., Brugman, H. und Kemps-Snijders, M. 2017. MTAS: A Solr/Lucene based multi tier annotation search solution. Selected papers from the CLARIN Annual Conference 2016. Linköping Electronic Conference Proceedings 136: 19–37.
- Fandrych, C., Frick, E., Kaiser, J., Meißner, C., Portmann, A., Schmidt, T., Schwendemann, M., Wallner, F., und Wörner, K. (2022). ZuMult: Neue Zugangswege zu Korpora gesprochener Sprache. In Kämper, H. et al. (Hrsg.). Sprache in Politik und Gesellschaft: Perspektiven und Zugänge. Jahrbuch des IDS. Berlin etc.: de Gruyter.
- Fandrych, C., Schmidt, T., Wallner, F. und Wörner, K. (Hrsg.). Zugänge zu mündlichen Korpora für DaF und DaZ: Das ZuMult-Projekt. *Korpora Deutsch als Fremdsprache*. 2023. 3 (1).

